

UNFPA Evaluation Quality Assessment Grid

Version: May 2024

REPORT RATING SUMMARY		
Overall Rating	73%	Satisfactory
● ● ● ● ●	Excellent	5
● ● ● ●	Highly Satisfactory	4
● ● ● ●	Satisfactory	3
● ● ● ●	Fair	2
● ● ● ●	Unsatisfactory	1

REPORT DETAILS	
Title of the evaluation report	Formative Evaluation of the UNFPA East and Southern Africa Regional Programme 2022-2025
Region	ESA
Country	n/a
Year of report	2024
Business Unit/programme country (managing evaluation)	ESARO
Date of assessment review (dd/mm/yyyy)	January 7, 2025
Name of assessment review firm	IODPARC
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)	Regional 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.	

This is a reasonably good evaluation, which adopted a non-standard approach to the research and analysis, in line with its nature as an evaluation of a regional strategy, rather than that of a discrete initiative with specific targets. While the analysis and findings of the report are strong, and mostly (but not completely) well-grounded in evidence, the background/context and description of the approach and methods could have been better, particularly given the non-standard nature of the work. Some of the other key strengths and weaknesses are as follows:

Strengths

- The executive summary is concise, well-organized, and reflects the main report effectively.
- Clear description of the evaluation purpose, objectives, and scope, aligned with UNFPA's strategic goals, including integration of cross-cutting issues like gender equality, human rights, and LNOB.
- The evaluation questions align with the OECD DAC criteria and include focus areas relevant to ESARO's strategy.
- An innovative and adaptive evaluation approach.
- Findings are systematically organized under evaluation questions and subheadings.
- Conclusions provide a well-balanced synthesis of findings, addressing positive and negative aspects.
- Recommendations are clearly linked to findings and conclusions, with actionable sub-recommendations.
- A good suite of annexes that provide additional insights (e.g. the deep dives and country case studies).

Weaknesses

- A major weakness is that the conclusions align with the evaluation objectives rather than directly addressing the evaluation questions.❏
- The roles of stakeholders are inconsistently described across sections.
- The report uses outdated context data in places, missing recent statistics for improved analysis.
- The methods section is overly brief and does not fully explain data handling, analysis, or sampling.
- The report does not fully explain the innovative, adaptive evaluation methods applied.
- Some findings rely on insufficiently detailed or unsupported evidence.
- Lessons learned are presented in annexes in various places and not summarised in the main report.
- A variety of formatting/editing issues as well as inconsistent or incorrect cross-referencing annex content to the main report.

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.

The following are the main areas for consideration for improvement in this (and other) evaluations:

- ❏ more comprehensive stakeholder map with clear roles and contributions, consolidating information scattered across sections and annexes.
- ❏Integration of the most recent statistics and data sources, replacing outdated information (e.g., pre-2022 figures).
- ❏Clearer description of the criteria, strategy, and process for selecting participants and sites, including disaggregation by stakeholder type and geography.
- ❏Additional detail in the methods section to explain data handling, coding, storage, and analysis, including how adaptive techniques were applied.
- ❏Additional focus on the level of detail and comprehensiveness of evidence for findings, avoiding reliance on general or anecdotal statements.
- ❏Avoidance of subjective language and evaluator opinions in findings, focusing on analysis of the evidence-base only.
- ❏If explicitly noted as an evaluation objective, provide a specific lessons learned section in the main report – or at the least, more clearly signpost where they can be found in the annexes.
- ❏Much greater attention to addressing formatting and editing errors and inconsistencies.
- ❏Conclusions should have been aligned with evaluation questions.

SECTION RATINGS		
SECTION A:	EXECUTIVE SUMMARY (weight 5%)	83%
Question 1.	Can the executive summary inform decision-making?	
	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 five page exactly, which is as per guidelines. It is clear, represents the content of the main report well, and should be useful to inform decision-making (with the exception of the level of detail in the recommendations - described in 1.iii below).

	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 of the specified components have been included, with a useful summary of the findings, and the conclusions and headline recommendation text presented as per the main report.
	iii	Includes all significant information in a concise yet clear manner to understand the theme, intervention, programme, project and the evaluation.	Partially	The main sections of the report have been summarised well to reflect their key contents in a clear manner. Therefore, the overall nature of the evaluand, and the findings and conclusions of evaluation itself, can be clearly understood from the text. The only issue is with respect to the recommendations, of which the headline text (the recommendation statement) only is included. As discussed under 16.ii below, this headline text is somewhat general for many of the recommendations, but the inclusion of the action points under each provides needed specificity and granularity. For many readers, the executive summary may be the only part of the report that they read, and indeed may well focus on the recommendations, to seek those with relevance to their business unit/sector. Therefore, the summary would have been improved by expanding the recommendations to include these action points and the target for each. To keep within page limits, the conclusions could have been summarised further (they are simply replicated from the conclusions section).
SECTION B:		BACKGROUND (weight 5%)	60%	Comments on Rating
Question 2.		Is the evaluand (i.e. intervention/policy/thematic area etc. that is to be evaluated) and context of the evaluation clearly described?		
	i	Clear description of the evaluand (e.g. intervention), including: geographic coverage, implementation period, main partners, cost/budget, and implementation status.	Partially	The evaluation report section 2.3 provides a high-level summary of the ESA RPAP for 2022-2025, noting the main strategic areas and the specific interventions, including a useful summary table of the programmatic logic (described as the RPAP Theory of Change in table 2, although it is more correctly a results framework). This summary is useful, and, importantly, links to the more detailed Regional Program [sic] Change Story in Annex B, which has more detail on some elements of the regional programme. The implementation period and geographical dimensions of the regional programme are clear, but the main partners and cost/budget are not fully explained. There are inconsistencies here that should have been addressed for clarity (discussed further under 3ii below).
	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	Section 2 describes the regional context of UNFPA's work, drawing on a variety of sources to describe the main trends and drivers in the areas related to the UNFPA mandate. It has some useful headline demographic, health and gender information, but has clear areas for improvement, e.g. and absence of information on some of the "megatrends" that are significant impactors on the region and a key feature of UNFPA's context, specifically humanitarian crises, climate change impacts, and digitalization. Furthermore, some of the data is old - references to 2020 or even earlier data when the evaluation covers 2022-2024 (if more recent data is unavailable, this should be noted). Subsection 2.2 covers the UNFPA global strategic plan, and summarises its main priorities (transformative results, outputs, accelerators, strategic shifts). The context section is quite duplicative of (some) of the information that is presented in the Change Story in Annex B (which also covers the global and regional strategies). The evaluators could have shortened the context section and referred the readers to that annex. This said, some of the context description in the Change Story could be updated and improved - while ostensibly it is a description of the context underlying the 2022-2025 strategy, in actuality it is essentially a slightly summarized version of the 2022-2025 strategy itself and should be updated for 2024.
	iii	Linkages drawn between the evaluand and the ICPD benchmarks and SDGs relevant targets and indicators.	Yes	The report clearly notes the linkages between the strategic plan and RPAP and the ICPD plan of action and specific SDGs. While this type of evaluation does not require a detailed reference or overview of progress against SDG indicators, given its focus on a regional strategy, these are itemized in Section 2.3. Although specific targets or indicators under each SDG are not detailed, Annex B provides frequent contextualized references to various SDGs and their targets, demonstrating clear linkages to the programme.
Question 3.		Are key stakeholders clearly identified and analysed?		
	i	Clear identification of key stakeholders which should include implementing partner(s), development partners, rights holders, and duty bearers among others; and of linkages between them (e.g., stakeholder map).	Partially	With respect to regional stakeholders, the report provides descriptions at varying levels of detail in different places - some of which are inconsistent. For example, subsection 2.3 notes support to government and "regional institutions" of the 23 ESAR countries, while the annex names various partners under each of the programme outputs. The final sentence of the subsection then describes another (general) list of stakeholders at all levels and refers the reader to Annex D, which is a list of stakeholders consulted - not necessarily a comprehensive list of all stakeholders (and indeed, omits any UNFPA stakeholders). Annex B provides more granular detail on stakeholders - naming specific institutions proposed as partners ('proposed' insofar as the annex is a summary of the original regional strategy which was prospective rather than retrospective in nature) under individual outputs, and thus not in one single place. Finally, Annex K provides some more information on partners under results/activities undertaken as part of the RPAP, but this is not referenced in the relevant section in the main report.
	ii	Stakeholders are analysed to understand their specific rights, duties, needs, interests, concerns, and potential impact on the evaluand.	Partially	Stakeholders are not clearly itemised in the background and context sections, resulting in a lack of analysis of their roles within the regional programme in the evaluation report. While this is partly understandable given the high-level and strategic nature of the regional programme action plan, an analysis of stakeholder interests and concerns at this level would have been valuable. However, the inception report (Annex 4) includes a stakeholder map outlining the roles of different stakeholders and their engagement in the regional programme.
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 first subsections (1.1, 1.2) of the evaluation report describe the overall purpose, context and need for the evaluation, situating it well within the context of the overall UNFPA strategic plan (existing and planned) and the regional strategy that is derived from and supports this. The 'triple purpose' of the evaluation is well represented, both narratively and visually, making it clear that such an evaluation is required at this time. The primary and secondary audiences are also clearly noted under subsection 1.2.
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 objectives are fully itemised in section 1.2, with no significant changes from those in the TOR to the evaluation. There are two minor issues worth noting: 1. There is one 'headline' objective and six more specific objectives (as well as an overall purpose) for the evaluation. While this is per the TOR, and thus not an issue with the evaluation report per se, it is somewhat confusing to have a hierarchy of objectives like this. Some revision at TOR development stage might have clarified this. 2. The headline objective and the six objectives are presented twice - in the narrative on pages 13 & 14 and then replicated word-for-word in the figure (fig 2). Once would suffice.
	ii Clear and relevant description of the scope (e.g. thematic, geographic, and temporal) of the evaluation, covering what will and will not be covered, as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political, humanitarian or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).	Yes	The scope of the evaluation is presented in section 1.2 (pp 13/14), with the specific temporal, geographical and programmatic elements noted. The programmatic scope is somewhat more complex than the temporal and geographical aspects (both of which are very clear and concise), insofar as there are a range of "key challenges and priority areas" noted, as well as "accelerators", "strategic shifts" (from the 2022-25 SP) and, finally, "enablers" derived from an unspecified "strategic plan evaluation". The programmatic scope was aligned with the Strategic Plan Evaluation. However, it would have been beneficial to better explain the linkages between the RPE and the Strategic Plan Evaluation or to include a reference to it in the footnotes for clarity.
SECTION D:	EVALUATION DESIGN AND METHODOLOGY (weight 20%)	70%	Comments on Rating
Question 6.	Are the selected evaluation questions and evaluation criteria appropriate for the purpose of the evaluation and is there clear justification for their use? <i>Note: UNFPA evaluation standards refer to the OECD/DAC criteria such as: relevance, coherence, effectiveness, efficiency and sustainability (not necessarily applicable to all evaluations) and, for country programmes that include circumscribed and limited humanitarian and/or emergency interventions, the criteria of coverage and connectedness.</i>		
	i Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Yes	The evaluation questions, as noted in the relevant section describing them (1.4) were developed at HQ level as an overarching set and then 'adapted' for all of the six regional evaluations. The questions for ESARO were stated to have been adapted to include a focus on humanitarian programming and the regional focus on HIV. The questions are presented in tabular format and cross-referenced with four OECD DAC evaluation criteria.
	ii Evaluation matrix clearly presents the evaluation criteria used as well as the corresponding evaluation questions, indicators, lines of inquiry, benchmarks, assumptions, source of data, methods for data collection and analysis, and/or other processes from which the analysis can be based, and conclusions drawn.	Yes	The evaluation matrix is presented in Annex E. It lists the evaluation questions, the corresponding criteria, and a number of assumptions to be tested, indicators and the sources/methods/tools to be used to answer these. The desk review is quite general, and the matrix could specify the types of documents reviewed to test each assumption or address each evaluation question. While this level of detail is provided for KIIs, identifying relevant stakeholders, it is not included for the document review.
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	As a formative evaluation, the results chain of the regional programme was not directly subject to testing by the evaluators. However, the approach did note that the evaluation would "where appropriate, use a theory-based assessment of change processes."
	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	The theory of change underlying the regional programme was not subject to revision or testing by the evaluators.
	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	The theory of change underlying the regional programme was not subject to revision or testing by the evaluators.
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	Section 1.6 describes the data collection methods in a very summary format. While all of the essential details are there, the process and description is not very clear. For example, the country selection process is not clearly described (though the relevant characteristics of each of the participating countries are noted, so clearly some criteria were applied). The methods appear to have been transcribed directly from the Inception Note, which is similarly brief. Although there is mention of the use of artificial intelligence in the TOR and the Inception Note (for unspecified data collection and analysis), there is no corresponding reference in the final report, so it is unclear if it has actually been used as planned. Notwithstanding the lack of details, the tools to be used (desk review, interviews, FGDs/group interviews, site visits/observations) are generally good practice for an evaluation of this nature so, if applied correctly, would be relevant and robust for the objectives and scope.

	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	<p>The <u>sources</u> of the data are relatively clearly described - qualitative documentary sources (listed in the annexes, though the presentation of the list of documents could have been improved - see 17ii below) and primary stakeholders (listed in annex d). However, the description is not comprehensive and full of clear errors. Most significantly, the list of stakeholders consulted in the annex has only one UNFPA staff in the AU Liaison Office. Further, it makes no reference to any stakeholders in Botswana (the evaluators accompanied the (presumably) global SPE team on their visit - one would assume that some primary data emerged from this visit). Finally, the formatting of the list could have been improved, as many position names have been omitted and at least one personal name included in error. The Stakeholder Consulted annex does not list any of the trainers, educators, or crisis center workers who were reportedly consulted, as stated in Section 1.7.</p> <p>The methods section notes "a combination of qualitative strategies" only, and there is no reference to quantitative data. There is a contradiction between the reference to qualitative strategies only and the stated mixed-methods design in the approach. Additionally, while Figure 3 notes that administrative and financial data were included in the desk review, the reference to quantitative data is unclear and too limited. Given the formative/forward-looking approach described in the overall approach, this is not inappropriate, though normal evaluation practice would generally include at least some secondary quantitative data (even on administrative/financial analysis, such as under the efficiency criterion). Finally, the methods note that site visits were to be conducted in each of the four countries, but the data outputs of these are not explained. A site "observation protocol" is included in Annex F, but is not referenced in the main report.</p>
	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 report does note preparation of a stakeholder map for the four countries to be included in the primary data collection (again, Botswana is noted as being visited, but no mention of data collection therein is described). Using the stakeholder map, the evaluators "developed the interview sample", but do not further describe the sampling strategy (e.g. purposive, opportunity, random, stratified etc.). The actual numbers of individuals sampled are indeed provided in table 2 (appropriately disaggregated).
	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	Although some light review was undertaken, the results chain of the regional programme was not subject to rigorous testing by the evaluators, nor were the results chain/result framework part of the analysis. As this is not a theory-based evaluation, this criterion is not rated.
	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>The analysis process is quite cursorily described in subsection 1.7. The report does provide some general detail on the "adaptive evaluation" approach, but it simply lists some of the, quite specialised, techniques intended to be applied ("creative tensions, causal loop diagrams, iceberg models, contribution analysis, and outcome harvesting") without providing any further explanation of what these are and how they will be applied to the collected data. A footnote does link to the UNFPA guidance document on adaptive evaluation (and the text justifies the approach with reference to UNFPA's A-Compass adaptive management strategy), but readers are unlikely to internalise a 100+ page manual for the purposes of understanding the evaluator's approach in this case. There is no mention of the use of AI in the evaluation process.</p> <p>What is missing from the section is basic information on how the collected data is collated, coded/treated, cleaned, stored and analysed, cross-referenced etc. There is some reference to this in the data collection tools annex (Annex F), but this would be better placed in the main report.</p> <p>An issue with the data collection methods section is the mischaracterisation of the "deep dives" as data collection methods - they are more correctly analyses of the data collected via primary and secondary means, and are (correctly) described as such in this section on analysis. The previous section should have characterised them correctly.</p>
	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 brief challenges/mitigations subsection provided (1.8) which notes two challenges and explains (narratively) how they were mitigated. While the two challenges noted have merit, and the mitigation strategies are appropriate, the subsection omits to note the extent to which (if any) the challenges materialised or affected the evaluation. Further some limitations of the evaluation were omitted. Specifically, the limited inclusion of rights-holders in the evaluation is a limitation that should have been noted, as the grounding of the work supported by the RO in the lived experience of rights-holders could have been an important dimension, particularly given that four field visits (five including Botswana) were conducted, with associated site visits (also not explained in the methods). However, the Evaluation Team engaged with beneficiaries in other contexts, such as educators, trainers, healthcare providers, and shelter/crisis center workers, and implemented safety and confidentiality safeguards as a result.
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	The report makes clear reference to the compliance of the evaluators with UNEG and UNFPA evaluation guidelines and itemises the UNEG ethical principles in subsection 1.5. While the itemised principles are not contextualised specifically, the description of the protocols <u>subsequently describes well how they are applied</u> .
	ii	Clear description of ethical issues and considerations (e.g. respect for dignity and diversity, fair representation, confidentiality, and avoidance of harm) that may arise in the evaluation, safeguard mechanisms for respondents (e.g. parental consent forms for adolescents, compliance with codes for vulnerable groups; WHO standards of safe data collection on GBV) and ethical considerations in the use of AI as applicable (e.g., transparency of use, explainability, privacy, data protection, accuracy, human rights). If AI is used in the evaluation, there should be transparency and disclosure on the ethical and responsible use of AI in the report.	Yes	The subsection clearly describes how the evaluation team applied ethical principles in the design of the evaluation and tools and their application in practice. The evaluation did not specifically seek out rights-holders as an evaluation target group, although did meet with "beneficiaries" (rights-holders is not used as a term) in the context of institutional interviews, and notes appropriate safeguards in this context. The data collection tools in the annex F notes that appropriate consent and confidentiality processes should be followed.
Question 10.		Does the evaluation incorporate innovative practice that adds value to the evaluation process?		

	i Innovation practice is used to improve the quality of evaluation process. This could include efforts to optimize the evaluation process (e.g., use of AI or new technology for data gathering, content analysis, outcome harvesting among others), or components introduced to enhance inclusion and participation in the evaluation processes (e.g. a youth steering committee), or ways of sharing of evaluation results.	Partially	The adaptive evaluation process itself is quite innovative. Although it is not comprehensively explained or contextualised for the purposes of this evaluation, the techniques noted (in brief) and the linked methodological guidance manual from the UNFPA evaluation office clearly incorporate a variety of innovative methods. The evaluation results sharing process (iterative and participative) is also innovative in this regard.
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	From a systematic perspective, the report findings section is organized around the three evaluation questions presented in the evaluation matrix. There are no further subheadings (e.g. according to subquestions) - the findings are presented directly under the EQ, which is acceptable practice. There is also a useful summary of findings/analysis before each EQ subsection. From an evidential perspective, there are positives and negatives. From a positive perspective, the evaluators frequently cite the sources for their analysis, referring to various key informants directly in the text, with frequent contextualisation of findings with examples and secondary data (including other evaluations, which are appropriately referenced via footnotes, e.g. the Hacklabs evaluation from 2024 under Finding 1, pg. 30). Further, the evaluators link specific analysis under the findings to more comprehensive analyses or examples presented across various annexes - this is a good practice to maintain the report at a reasonable length and avoid excessive detail. On the negative side, the source of some of the evidence cited is not clearly noted. For example, under finding 4 (EQ1), the evaluators note the challenges in working with LGBTQ+ populations but claim that the RO staff are "skilled at navigating these waters" with no explanation or evidence supporting this claim, or even what this (colloquial) phrase means. This is a very important area that could have benefitted from further analysis/unpacking, particularly given the challenges faced by individual country offices in this regard (noted in the text). The selection of deep dives could have been reconsidered. Humanitarian preparedness and response were not included, while HIV transmission prevention was. Additionally, the shift from funding to financing, already one of the 12 strategic shifts of the SP 2022-2025, should have been covered under EQ2, making its inclusion as a deep dive unclear. Similarly, partnerships, South-South cooperation, and financing (one of the six SP accelerators) should have been addressed under EQ1, yet it was also chosen for a deep dive without a clear rationale. There are some minor issues relating to finding/analysis writing - this is explained under 19i below.
	ii Explicit use of the evaluand's theory of change, results chain, logical framework in the formulation of the findings.	Not Rated	As this is not a theory-based evaluation, this has not been rated.
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	Overall, the report relies on a reasonably robust mix of data from both primary (interview/FGD) and secondary (documentary) sources. Qualitative data is utilized only in the evaluation. While the sources are credible (internal and external sources, some representatives of rights-holder groups such as youth network representatives in Zimbabwe cited under Finding 7), there are a few (but not many) instances where the evidence presented against outcomes (rather than outputs) could be more detailed. For example, the analysis of partnerships (finding 2) presents a table that categorises the "current" (what date this refers to is not defined) partnerships according to strength and lists potential opportunities they present - this is a useful analysis, but it quite summary and the underlying evidence linking the outputs to partnership outcomes is unclear. The reader is referred to Annex J ("Deep Dives" - one of which is partnerships), but this supposedly more granular analysis merely presents the same table with limited additional evidence. Another example can be found under finding 3, which presents (in summary, with additional details in Annex J) achievements of the RO with HR/gender transformative approaches - the analysis is a good start, but additional evidence (e.g. from primary stakeholders) to triangulate the finding and confidently attribute the results of the reported work to outcomes is lacking.
	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	As discussed above, there are examples of good and poor approaches to the presentation of a robust evidence base under the findings. From a positive perspective, there are many instances where the evidence (and its source) is directly cited, e.g. Finding 13, where the strong performance of UNFPA within the interagency system is linked to the evidence from other UN agencies and individual stakeholders (such as Resident Coordinators) as well as the documentary evidence of UNFPA participation/leadership in joint initiatives. This example is illustrative of many of the findings and indeed is supported by the annexes which have additional evidence. While it is appreciated that the evaluators are bound by a need for concise analysis (to remain within overall page limits), there are examples where the evidence underlying the analysis is not presented. In parts, the text reads more like a positively-skewed progress report rather than an objective analysis of evidence and data. An example of this is the initial text under Finding 15, which reads " <i>There should be no doubt that the personnel at the Regional Office are dedicated, skilled and respected by stakeholders as advocates for beneficiaries and UNFPA representatives.</i> " This is a judgement statement that does not refer to independent evidence as robust analysis should. The text goes on to note that the evaluators were " <i>impressed by their strategic approach to problem solving...[etc]</i> ". The findings/analysis should be based solely on a review of impartial and independent evidence, not the opinions of the evaluators.
	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	Notwithstanding some of the evidence shortfalls noted above, the evaluation, where appropriate, does a good job of attempting to analyse the causal factors of achievements or lack thereof. A good example is with respect to Finding 8, which explores the alignment of the regional strategy with the 3TRs - the evaluators find that it has been an "imperfect fit", and analyse the reasons why this is the case. Similarly, the subsequent finding (9) looks at the use of data on megatrends and notes shortcomings in this regard - these are usefully presented in table 5 which includes important areas for improvement (i.e. the causal factors). These factors are further explored via the "Deep Dives" of five key themes presented in Annex J - the analysis herein, and particular the Lessons Learned and Conclusions elements of each, are a useful analysis of the ESARO results in these areas.

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	This was not requested of the evaluators and hence is not rated.
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.	Partially	Firstly, the six evaluation conclusions are explicitly formulated to respond to the evaluation objectives rather than the evaluation questions, and this is a major weakness of the report. The conclusions themselves are reasonably summaries of the analysis that precedes, with positive and negative dimensions represented as appropriate. One small issue is in relation to Conclusion 3, which notes that "Decreased resource mobilization is an existential threat to the RO" - this is somewhat extreme insofar as the existence of the RO or UNFPA is not threatened - the threat relates to programming and operations. The language (and that of the related finding (#12) might be moderated to be somewhat more temperate. In addition, the conclusions are overly concise and lack sufficient nuance.
	ii Conclusions are well substantiated and derived from findings and add deeper insight and analysis beyond the findings.	Yes	The conclusions statements generally reflect the analysis that precedes them, representing a reasonable synthesis of many of the findings. The conclusions do reference the specific findings that underpin them (all of the findings are referenced in the various conclusions), which is a positive practice. There was one instance noted where some of the analysis was not fully represented in the conclusions - Conclusion 1 notes the adaptation to "increasing complexity" in the region, but omits in the list of findings #9 related to "megatrends" that are a significant driver of this complexity. Beyond this, the conclusions appear well substantiated. As per the UNFPA Evaluation Handbook, the conclusions should also indicate the recommendations to which they are linked.
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.	Yes	The TOR for the evaluation does not directly specify a lessons learned section, but notes the drawing of "clear lessons" as part of objective 6. There is no specific lessons learned section in the main report, but the four "Deep Dives" in Annex J incorporate lessons learned subsections under each of the short analyses. These lessons (eight in total - between 1 and 3 per Deep Dive) are derived from the findings in the Dives, and are, by their nature, related to practical aspects of ESARO work. There are also some examples of lessons in the Country Success Stories in Annex I, although these are less explicitly noted.
	ii Lessons learned are clearly presented and provide actionable insights on the positive aspects of the evaluand as well as any areas of improvement.	Partially	The lessons themselves are presented in the Annex as a subsection to each Deep Dive, so, for those that take the time to read the annexes, they can be accessed. However, few readers will do so, and the main report might have benefited from a summary of these after the conclusions section, or at least a clear signpost to indicate that the lessons could be sourced in the relevant annex. The report is not long, so 1-2 pages of the key lessons would not be excessive. The content of the lessons themselves is good - they clearly note important insights that determined initiative success (or otherwise) and note what should be learned and/or acted upon as an outcome of the lesson. For example, the lesson under Deep Dive 1 notes the importance of elevating HIV transmission work to the regional level - thus unlocking additional funding.
SECTION G:	EVALUATION RECOMMENDATIONS (weight 15%)	100%	Comments on Rating
Question 16.	Are recommendations well-grounded and articulated?		
	i Recommendations are clearly formulated and logically derived from the findings and/or conclusions.	Yes	The seven recommendations are presented in a table format, with a headline recommendation statement and sub-recommendations/action points. The layout and formulation is very clear and comprehensible. The recommendations section clearly notes that each recommendation is 'tied' to the relevant conclusions and findings in the preceding section, and the formulation of the recommendations reflects this, with the specific conclusions and findings itemised for each. This is a strong positive feature of the section. As per the UNFPA Evaluation Handbook, recommendations should be structured according to programmatic and strategic recommendations.
	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	While the overall headline text of the recommendation is, in some cases, somewhat general, the specific action points under each are considerably more specific and highly actionable for the relevant stakeholders (which are also noted for each recommendation). No specific deadlines are given under each recommendation, as the introductory text notes that they are focused on the development of the upcoming RPAP in "fall" 2024 (note that the use of an Americanised term for autumn is not per UNFPA guidelines and UN style guidelines recommend not to refer to seasons in any case).
	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 preamble to the recommendations section (5) notes the process of development. This included consultation with Country Offices in the region, stakeholders from the ESARO and the members of the ERG. As this is a high-level regional evaluation, these are the appropriate stakeholders that are most affected by the recommendations.
	iv Recommendations are clearly articulated and prioritized based on their importance, urgency, and potential impact.	Yes	The recommendations are very clearly articulated - the table format is useful and clear, with the key roles, linkages to findings/conclusions and the action points all well-presented. The evaluators note that the focus of the recommendations on the planned development of the next RPAP means that differentiated timelines and priority levels were deliberately not assigned. This is presumably with the agreement of the stakeholders participating in the co-creation of the recommendations and is this acceptable. As per the UNFPA Evaluation Handbook, recommendations should be prioritized as follows: high priority, medium priority, low priority.
SECTION H:	REPORT STRUCTURE AND PRESENTATION (weight 5%)	83%	Comments on Rating
Question 17.	Does the evaluation report include all required information?		

	i Opening pages include: Name of evaluation and/title of evaluation, timeframe of the evaluation, date of report, location of evaluand, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents (including, as relevant, tables, graphs, figures, annexes)-; list of acronyms/abbreviations.	Yes	The cover page includes all of the relevant details specified in the criterion. The following pages include a full table of contents, list of tables/figures (the table of contents breaks across two pages and could easily be reformatted to be on one), a list of annexes and a list of acronyms/abbreviations. These (while potentially benefitting from some formatting improvements - see further below) are per the requirements. One issue is that the membership of the ERG is not listed in any of the evaluation documents - the general membership categories and the role/responsibilities are explained in the TOR, but the membership is not listed in the opening pages per the UNFPA Evaluation Handbook (although not specified here as a criterion).
	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 an extensive suite of annexes provided with the report (in a separate volume), which include all of the mandated elements and a variety of additional analyses that provide depth and detail to support the main report. This is useful, although some of the cross-referencing in the main report has errors (noted below). One specific issue is that the list of documents reviewed in the annexes, while extensive, is a poorly-formatted list of basic document titles (e.g. lacking detail, such as "annual report 2022" without noting the organisation/location etc.), not organised alphabetically nor with the authors noted - this limits the utility of this annex considerably.
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 structure follows the appropriate UNFPA evaluation guidance, and is therefore straightforward to navigate, with sections numbered and titles clear. The formatting could have benefited from a more careful editing for readability - many sections or tables either break across pages (e.g. the acronyms/abbreviations table, the table of contents, annex B) or (for sections) do not start on a fresh page. Some adjustment of line spacing and use of page breaks would have addressed this and given the report a cleaner look. There are also some other formatting issues with the figure placement on pages (see Fig 2, pg. 14); the use of capitalization (e.g. Regional Office/Region in places, regional office/Region in others); missing/incorrect footnote references (e.g. Annex B, footnote 13, main report footnote 33); incorrect page orientation (e.g. Annex B, pg. 43).
	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	As noted above, the structure follows the guidance from UNFPA in the Evaluation Handbook. Excluding the cover pages and executive summary, the report is 53 pages, which is within the mandated page length (60 pages for institutional evaluations).
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	Overall, the writing style is acceptable - the report is accessible and comprehensible. There are a variety of issues that still remain, however. One minor issue related to how the findings are articulated/written is the use of "canned" text at the beginning of findings 1-6 (relating to the six UNFPA accelerators) - each finding begins with the same sentence "[accelerator title] is one of six accelerators of the UNFPA Strategic Plan 2022-2025, which the Regional Office has adopted for the Regional Programme Action Plan (RPAP) 2022-2025." This is repetitive and unnecessary. Another issue is linking findings to additional detail in annexes - examples from individual countries have incorrect linkages in the Annex I (the relevant findings in the annex are misnumbered). There are also various formatting issues. Examples include: - The line spacing should be reduced slightly to bring the table of contents onto a single page. - correct indenting of page numbers in table of contents for annexes. - Per the UNFPA style guidelines, numbers 0-9 should generally be written in full ("three") and those above in numerals. - Page numbering for the opening pages should have been in roman numerals (i, ii, etc.) and from the executive summary (pg. 8) onwards restarted with Arabic numerals (1, 2, etc.). - Some acronyms/abbreviations not explained (e.g. ERG). - Figures are formatted poorly, with text wrapping incorrectly and some pushed to the margins. Editing should have followed the UNFPA style guidelines and relevant guidance in the UNFPA Evaluation Handbook (2024 edition).
	ii Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labeled, and referenced in text.	Partially	There is a good range of tables and figures presented with the report. All of these provide good illustration of key analysis/findings. However, not all are labelled - for example, the map at the start of section 2 has no label or source (it is copied from the original TOR). Further, the list of figures and tables on page 4 includes the captions/titles of the figures & tables (which is appropriate) but also the sources (which is not). Some of the formatting of the figures is out (e.g. figure 10, pg. 53 is pushed to the page margins).
SECTION I: CROSS-CUTTING ISSUES (weight 10%)		77%	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.	Yes	The evaluation was focused on the assessment of regional work, which does not reach down to the rights-holder level directly, so inclusion of the perspectives of these groups might not have been expected. However, as noted above under question 8(vi), their absence should have been noted as a limitation, particularly as they may have been indirectly included during site visits and/or consultations with rights-holder organizations which were included.
	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	Insofar as the evaluation assessed (under EQ1) the effectiveness of the ESARO achievements in relation to the three TRs, and the six UNFPA accelerators - which include LNOB, gender and human rights transformative approaches and megatrends such as climate change, these issues were well-integrated into the evaluation.
	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).	Not Rated	The evaluation did not collect or analyse quantitative data, so no disaggregation according to these groups was undertaken, hence this criterion is not rated. However, gender disaggregation is provided for the stakeholders consulted, outlined in Table 2.

	iv	Intersectional lens is applied in the data analysis, looking at various and multiple forms of exclusion and discrimination (and how they overlap with each other) and how this may impact the performance or results of the evaluand.	Yes	The qualitative analysis and findings did integrate a good focus on the various factors that drive exclusion and marginalisation for vulnerable groups, including documenting work on multiple (i.e. intersectional) vulnerabilities. This is well-reflected, for example, in findings 4 (which discusses LNOB as an accelerator), 5 (which looks at resilience and humanitarian actions) and recommendation 1, which has a specific action point related to ensuring an intersection lens is applied to social and gender norms change across all UNFPA work in the region.
	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	Issues of marginalisation are well reflected in findings, conclusions and recommendations (as noted above). Issues of equality and vulnerability are core to the realisation of the 3TRs and the six accelerators, achievement of which were comprehensively integrated in the analysis. For example, disability inclusion is discussed (with references to examples in the annexes) under Finding 4. This is, however, somewhat general - the area is considerably more nuanced across the region than the analysis covers, although space/size limitations necessarily constrain the evaluators. A more positive example under the same finding is around the challenges in working with LGBTQ+ groups - which is a clear area where UNFPA falls short, and is well-noted by the evaluators. However, these issues are absent from conclusions, pointing to lack of systematic integration.
	vi	Inclusion of young people in the evaluation team and/or Reference Group [N/A if not requested in ToR]	No	The TOR for the evaluation does mandate the inclusion of a "young evaluator", with specific details of their role and profile provided. However, there is no mention of such an individual being part of the evaluation team in either the main report, annexes or the inception report.
Question 21.		Does the evaluation meet UN SWAP evaluation performance indicators? <i>Note: this question will be rated according to UN SWAP standards with detail provided below</i>	8	Comments on Rating
	i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Fully integrated	The evaluation was designed in line with the UNFPA mandate and the related overall evaluation of the UNFPA strategy, and therefore integrates GEEW considerations well into its design. This is reflected across the evaluation purpose, objectives, scope and questions, via the focus on the 3TRs and six accelerators.
	ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Fully integrated	The evaluation scope specifically notes the inclusion of the UNFPA strategic priorities and accelerators, that include GEEW considerations. Further, the overall evaluation approach highlights adherence to "the principles of gender-responsive evaluations" and notes that gender and human rights were a cross-cutting criterion of the evaluation analytical approach. The data collection tools (specifically the interview questions in Annex F) include questions specifically probing the effectiveness of programming related to GEEW and analysis disaggregates respondents by gender as well as specific analysis relating to the GEEW performance of UNFPA in the findings.
	iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	As per the design, the evaluation findings analyse the effectiveness of UNFPA with respect to gender-transformative approaches as one of the six UNFPA accelerators (notably under findings 3 & 7, though cross-cutting other findings too). This is also reflected in the recommendations, with both recommendations 1 and 2 highlighting (in the action points) the need for more focus on gender transformative approaches to social norms change. However, the conclusions did not incorporate gender equality considerations, which is why this has been rated as "Satisfactorily integrated."

SWAP Rating Guidance

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.

- 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?
- Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?
- Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?
- 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.

- 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?
- Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?
- Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?
- Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?
- 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.

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

- No Poverty
- Zero Hunger
- Good Health and Well-being
- Quality Education
- Gender Equality
- Clean Water and Sanitation
- Affordable and Clean Energy
- Decent Work and Economic Growth
- Industry, Innovation and Infrastructure
- Reduced Inequality
- Sustainable Cities and Communities
- Responsible Consumption and Production
- Climate Action
- Life Below Water
- Life on Land
- Peace, Justice and Strong Institutions
- Partnerships for the Goals

Three transformative results

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

Six outputs

- Policy and accountability
- Quality of care and services
- Gender and social norms
- Population change and data
- Humanitarian action
- Adolescents and youth

Six accelerators

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

