

Organizational unit: UNFPA Bosnia and Herzegovina Country Offices **Year of report:** 2019

Title of evaluation report: UNFPA Country Programme Evaluation Bosnia and Herzegovina

Overall quality of report: **Very Good** **Date of assessment:** 2020 MAY

Overall comments: This evaluation is one of four undertaken as part of a cluster evaluation of UNFPA country programmes under the same administrative unit within the Eastern Europe and Central Asia region. All use the same methodological approach and areas of enquiry. This report provides a comprehensive evaluation of the Country Programme for Bosnia and Herzegovina. It is notable for its clear integration of ethical standards and GEEW considerations within the scope, methodology, findings and recommendations. The strengths of the evaluation are also apparent in the reliability of data, analysis and findings, conclusions, recommendations, and integration of gender. The evaluation uses a mixed-methods approach linked to an elaborate and strong evaluation matrix and well-constructed theory of change and logic models. Furthermore, the evaluation consults a diverse sample of stakeholders which is disaggregated by gender. The limitations in data collection are adequately mitigated. The report provides findings that reflect good analysis of the programme's strengths and weaknesses from a human rights and gender equality (GEEW) lens. The conclusions provide a balanced perspective of both the positive and negative processes and outcomes and unintended outcomes. In addition, the conclusions and recommendations are coherently linked and are clearly actionable, prioritized, and target users. However, more attention could be given to strengthening the evaluation design and methodology, most notably by being more explicit about the data analysis processes. There could also be more use of quotes and stories to highlight the voices of different groups.

Assessment Levels

Very Good	strong, above average, best practice	Good	satisfactory, respectable	Fair	with some weaknesses, still acceptable	Unsatisfactory	weak, does not meet minimal quality standards
------------------	--------------------------------------	-------------	---------------------------	-------------	--	-----------------------	---

Quality Assessment Criteria	<i>Insert <u>assessment level</u> followed by main <u>comments</u>. (use 'shading' function to give cells corresponding colour)</i>	
1. Structure and Clarity of Reporting	Yes No Partial	Assessment Level: Good
<i>To ensure the report is comprehensive and user-friendly</i>		
1. Is the report easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors?	Yes	The report is clearly written with minimal errors.
2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations)	Yes	At 72 pages, the report is of reasonable length.
3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The report structure is coherent and includes a clear delineation of sections. The use of evaluation criteria to report key evaluation findings cogently covers the three key components of the evaluation.
4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process?	Yes	Annexes contain all the required information. In addition to the documents expected elements, the annexes include the logic model, a strategic overview of the UNDAF focus areas and the UNFPA Strategic Plan 2018-2021 theory of change, an overview of SDGs, and the targets relevant for UNFPA CP.
<i>Executive summary</i>		
5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Yes	The executive summary is well-written and informative. It presents the main results as a stand-alone section.

6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Partial	The executive summary follows the desired structure and it contains relevant information suitable for people who might not read the main report. However, the intended audience is not specified.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The executive summary is within page limits.
2. Design and Methodology		
	Yes No Partial	Assessment Level: Fair
<i>To ensure that the evaluation is put within its context</i>		
1. Does the evaluation describe the target audience for the evaluation?	Yes	The report briefly describes the primary users of this evaluation as the decision-makers in cluster countries/territory where UNFPA operates, including the UNFPA as a whole, government counterparts, and other development partners.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The evaluation covers both components. The description of the development and institutional contexts is thorough and constraints are well-described.
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Partial	The report provides an overview of UNFPA Strategic Plans in the Annex 8, logic models in Annex 7, and a summary of the theory of change under the current UNFPA Strategic Plan 2018-2021 in Annex 10. Although the ToR notes that the design phase of the evaluation should include an analysis of the intervention logic of the programme, the report does not describe the adequacy of logic models or theory of change.
<i>To ensure a rigorous design and methodology</i>		
4. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?	Yes	The evaluation provides an elaborate evaluation matrix with sections describing the logic chains within the matrix presented in Annex 4. The matrix includes the evaluation questions, assumptions, indicators, data sources and methods for data collection.
5. Are the tools for data collection described and their choice justified? Remember: Please address both aspects of this sub-criteria in the comment: 1) are data collection tools described (i.e. documentary review, interviews, focus group discussions etc.) and 2) is the rationale for their selection detailed	Yes	The mixed methods data collections tools are described and the interview protocols for different stakeholder groups are provided in Annex 6. The justification of the mixed methods approach is explained in the methodology section as necessary for triangulation of information from different sources.
6. Is there a comprehensive stakeholder map? Is the stakeholder consultation process clearly described (in particular, does it include the consultation of key stakeholders on draft recommendations)?	Partial	The report does not provide a stakeholder map (which was requested in the ToR to be included in the design phase), but notes that the sampling framework was based on the list of stakeholders provided by UNFPA. The participation of stakeholders is described in the methodology section of the report. The report acknowledges consultations with the members of the reference group, including taking time to attend the inception and out-briefing and providing comments on key deliverables.
7. Are the methods for analysis clearly described for all types of data?	No	Neither the overall synthesis report or country report are explicit about the data analysis strategy.
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Section 1.3.2 adequately describes the limitations and potential biases introduced in the evaluation and how these limitations and biases were mitigated..
9. Is the sampling strategy described? This criteria is asking whether the methodological approach to determining the sample of stakeholders consulted and the sample of site visits is described. Reviewers should examine whether the evaluation report includes information on how the universe was determined; the sampling approach used (i.e. purposive); the indicators used to develop the sample to be consulted (or visited); the resulting sample; and importantly limitations to the approach (including any potential resulting bias).	Yes	The sampling approach for KIs is described as a purposive and non-random selection of key informants. The justification is described in section 1.3.1. The sampling strategy for the group discussions with training participants is described, and for the client/beneficiary interviews.

10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	This is done. The evaluators note the potential causes of bias and address the threats to reliability of the data by selecting interviewees representing a diverse range of institutional viewpoints on key topics under review. The evaluators provide a good presentation of disaggregated data by gender in the description of the country context.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The evaluators are careful to show how the cross-cutting issues (gender, human rights, and disability responsive) were assessed.
3. Reliability of Data	Yes No Partial	Assessment Level: Very good
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	It is noted that the evaluators triangulated data sources, data collection methods and tools, and validated data and findings through regular exchanges with the UNFPA programme staff and the Evaluation Reference Group. Document sources are regularly footnoted and qualitative findings are regularly attributed to specific stakeholder groups.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluators explicitly state that the sources of qualitative and quantitative data were from documentary review, group and individual interviews, focus groups and field visits to programme sites as appropriate. The reliability of the sources is also discussed.
3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues?	Yes	The potential causes of bias are noted; the threats to reliability of the data are addressed by selecting interviewees representing a diverse range of institutional viewpoints on key topics under review.
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	The evaluators ensured that all participants were treated as autonomous and free to choose whether or not to participate in the evaluation. The ethical considerations were incorporated in the "Ground Rules" in which participants were informed about the confidentiality of the interview data and that their participation was voluntary.
4. Analysis and Findings	Yes No Partial	Assessment Level: Very good
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	The evidence for the findings are indicated in footnotes referencing document reviews or key informant interviews. The evaluation team does a good job providing summaries of key findings for key evaluation questions in textboxes and sentences in bold at the start of relevant paragraphs.
2. Is the basis for interpretations carefully described?	Yes	The report clearly shows how interpretation bias was addressed. The evaluators collected and interpreted qualitative and quantitative data according to the indicators presented in the evaluation matrix in Annex 4, the intervention logic, and principles presented in the theory of change in Annex 10 and Annex 8.
3. Is the analysis presented against the evaluation questions?	Yes	The findings are organized by evaluation questions per evaluation criteria.
4. Is the analysis transparent about the sources and quality of data?	Yes	As noted earlier, the sources of data used are referenced in footnotes although use of direct quotes could further improve the transparency of the analysis. The quality of data was described when necessary.
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The evaluators used the evaluation matrix, the theory of change model, and counterfactual analysis wherever possible to explore the cause-to-effect relationships between the intervention and its end results. The evaluators highlighted unintended positive and negative outcomes of the intervention in response to EQ3.

6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings show evaluation outcomes by programme area for specific groups targeted by the country programme, especially women, adolescent girls and at-risk youth populations, older persons and the migrant population, and sectors.
7. Is the analysis presented against contextual factors?	Yes	In each case, the findings are shown against the relevant contextual factors in which (and with which) the country programme works, including specific population groups, institutions, and the relevant national and international policies.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The evaluators were careful to examine the cross-cutting issues including gender and human rights of targeted participants particularly women and adolescent girls including those at risk of child marriage.
5. Conclusions		
	Yes No Partial	Assessment Level: Very good
<i>To assess the validity of conclusions</i>		
1. Do the conclusions flow clearly from the findings?	Yes	The recommendations logically flow from the conclusions, and each recommendation is linked to specific conclusions.
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated?	Yes	The evaluation team provided five strategic level conclusions, 8 programmatic level conclusions covering all programmatic components, and three conclusions relevant to UNCT Coordination and UNFPA added value. The hierarchy of conclusions clearly underscore the evaluators' understanding of the country, program contexts, and system being evaluated.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no evidence of bias since the conclusions are clearly based on evidence from the findings showing both positive and negative elements.
6. Recommendations		
	Yes No Partial	Assessment Level: Very good
<i>To ensure the usefulness and clarity of recommendations</i>		
1. Do recommendations flow logically from conclusions?	Yes	The recommendations logically flow from the conclusions, and each recommendation notes the conclusions on which it is based.
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	The recommendations specify to whom they are directed, explicitly providing operational and technical implications.
3. Do recommendations appear balanced and impartial?	Yes	The recommendations flow clearly from the findings and clearly specify the relevant conclusions without any evidence of bias.
4. Is a timeframe for implementation proposed?	Yes	The recommendations are classified into high or medium priority for both the strategic and programmatic recommendations. The timeframe for implementation is not proposed but it is clear in the initial sections of the report that recommendation will be used for the next CP planning process.
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The classification into high and medium seems appropriate to facilitate management response, and particularly considering that the evaluators obtained feedback from the CP stakeholders on the recommendations.
7. Gender		
	0 1 2 3 (**)	Assessment Level: Very good

To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)

1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?

a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3)
 The evaluation's purpose includes analysis of Gender Equity along with the other CP focus areas. GEEW was included in responses to all evaluation criteria particularly EQ1 and EQ3 and selected indicators. (Score = 3).

b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3)
 There was no standalone criterion on GEEW, but the assumptions and indicators for EQ1 and EQ3 assesses how GEEW is implemented in line with the priorities set by the international and national policy frameworks and aligned with the UNFPA policies and strategies and the UNDAF. (Score = 3).

c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3)
 As noted earlier, GEEW was integrated in EQ1 (a) To what extent is the UNFPA programme adapted to the needs of women, adolescents and youth, people at risk of HIV infections, disabled, older persons and Roma? and EQ3. To what extent did the outputs contribute to the achievement of these planned outcomes in mainstreaming of provisions to advance gender equality? (Score = 3).

d. 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?(Score: 0-3)
 An assessment of the sufficiency of information collected during the implementation period on specific results on GEEW was addressed within EQ6 (Efficiency). (Score = 3).

2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?

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? (Score: 0-3)
 The evaluation methodology was gender responsive with regards to the methods and tools for data collection. (Score = 3).

b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3)
 The methods and data collection tools included key informant and focus group interviews with primary beneficiaries that included women and youth (interviewed as separate groups). It is noted that "Husbands of women will not be asked to participate in order to ensure women are not inhibited" (Score = 3).

c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3)
 It is clearly stated in the report that triangulation and validation of the data and findings was done, but there is no clear evidence of how analyses ensured that data is disaggregated except having a gender balance in stakeholders consulted (with more women interviewed). The team got feedback from Co and ERG on both the draft and final reports to validate accuracy of findings, thus ensured the credibility of the data. (Score = 2)

d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3)
 Yes, this was ensured with samples representing the primary beneficiaries as noted above. (Score = 3).

e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3)
 The interview and focus groups protocols indicate that data collectors complied with ethical guidelines for informed consent and respect for confidentiality. Although the report is not explicit about the appropriateness of interview and focus groups locations, it is noted in the report that primary beneficiaries included women and youth (interviewed as separate groups). (Score = 3).

<p>3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?</p>	<p>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? (Score: 0-3) Gender-analysis is evident in the background section especially assessment of policies and practice, for example, it is noted that women's political participation is low in BiH despite there being a gender quota of 40% women in elected positions, and that the country has a legal framework for combating gender-based violence. (Score = 3).</p> <p>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? (Score: 0-3) The GEEW key findings summaries in the evaluation matrix in Annex 4 are well-done. Within the main report, while the findings include data analysis that explicitly triangulates the sources of data, there is minimal use of quotes or stories to highlight the voices of different groups, or presentation of the perspectives of the various right-holder groups. (Score = 1).</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) The unanticipated effects of the intervention on human rights and gender equality is reported in the Key Findings section, especially the summaries in the evaluation matrix. (Score = 3).</p> <p>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3) The report provides seven specific conclusions and recommendations that explicitly address gender equality, particularly Conclusion 5 on Relevance, and Conclusion 6 on Effectiveness, Sustainability. (Score = 3).</p>
--	--

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totaling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

(**) Scoring uses a four point scale (0-3).

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

Overall Evaluation Quality Assessment

Quality assessment criteria (scoring points*)	Assessment Levels (*)			
	Very good	Good	Fair	Unsatisfactory
1. Structure and clarity of reporting, including executive summary (7)		7		
2. Design and methodology (13)			13	
3. Reliability of data (11)	11			
4. Analysis and findings (40)	40			
5. Conclusions (11)	11			
6. Recommendations (11)	11			
7. Integration of gender (7)	7			
Total scoring points	80	7	13	
Overall assessment level of evaluation report	Very Good			
	Very good very confident to use	Good confident to use	Fair use with caution	Unsatisfactory not confident to use

- (*) (a) Insert scoring points associated with criteria in corresponding column (e.g. - if 'Analysis and findings' has been assessed as 'Good', enter 40 into 'Good' column).
- (b) Assessment level with highest 'total scoring points' determines 'Overall assessment level of evaluation report'. Write corresponding assessment level in cell (e.g. 'Fair').
- (c) Use 'shading' function to give cells corresponding colour.

If the overall assessment is 'Fair', please explain

- How it can be used?

- What aspects to be cautious about?

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory

Consideration of significant constraints

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes

No

If yes, please explain:

Organizational unit:

Year of report: 2019

Title of evaluation report: **The UNFPA Kosovo (UNSCR 1244)* Programme Evaluation Report**

Overall quality of report: Good

Date of assessment: 2020 MAY

Overall comments: This evaluation is one of four undertaken as part of a cluster evaluation of UNFPA country programmes under the same administrative unit within the Eastern Europe and Central Asia region. The methodological approach and areas of enquiry are consistent for each. This report provides a comprehensive evaluation of UNFPA Country Programme Evaluation for Kosovo. The strengths of the evaluation are in the reliability of data, analysis and findings, and integration of gender. The evaluation uses a mixed-methods approach linked to a strong evaluation matrix and reconstructed theories of change for each program component. However, the methodology section could be more robust, including by describing how data was analyzed and how gender considerations were taken into account, and by including more details on study limitations and mitigation strategies (although limitations are discussed in the overall cluster evaluation report). The report provides findings that reflect good analysis of the programme's strengths and weaknesses. The conclusions thoroughly address some, but not all, of the criteria for each program component but would benefit from overall statements on the achievements, strengths and challenges of the CP. The recommendations generally appear useful for informing the next phase of the programming cycle. Overall, more attention could be given to improving the structure and clarity of some sections, including more use of spacing in the executive summary, highlighting the topic of each conclusion and recommendation, and clearer articulation of the recommendations.

Assessment Levels

	Very Good	strong, above average, best practice	Good	satisfactory, respectable	Fair	with some weaknesses, still acceptable	Unsatisfactory	weak, does not meet minimal quality standards
--	--	--------------------------------------	---	---------------------------	---	--	---	---

Quality Assessment Criteria *Insert assessment level followed by main comments. (use 'shading' function to give cells corresponding colour)*

I. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Fair
--	----------------------	-------------------	---

To ensure the report is comprehensive and user-friendly

1. Is the report easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors?	Yes	The language of the report is accessible; easy to read and understand.
2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations)	Yes	The report is just under the 70 page-limit.
3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The report is structure is coherent and includes a clear delineation of sections. The use of the evaluation criteria to report key evaluation findings cogently covers the three outcome areas of the UNFPA mandate and key components of the evaluation.
4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process?	Yes	Annexes contain all the required information. The evaluation matrix is exceptionally detailed with summaries of key findings for each evaluation questioned organized by programme outcomes. The results chain and results framework are equally detailed and well-organized.

Executive summary

5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Yes	The executive summary is well-written and informative. It presents the main results of the evaluation as a stand-alone section.
---	-----	---

6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Partial	The executive summary follows the desired structure, except that it doesn't have a separate conclusions section (conclusions are woven into Findings). Otherwise, it contains relevant information suitable for people who might not read the main report.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	No	The executive summary is within page limits. However it is text heavy which makes it difficult to read. Clarity would be improved with more spacing and shorter paragraphs. This is particularly an issue for the recommendations section which has 15 recommendations and is presented in one paragraph that is almost two pages in length.
2. Design and Methodology		
	Yes No Partial	Assessment Level: Fair
<i>To ensure that the evaluation is put within its context</i>		
1. Does the evaluation describe the target audience for the evaluation?	Yes	This report does not explicitly state the target audience for the evaluation, however the information can be found in the overall cluster evaluation report.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The description of the development context and its constraints is provided in the Country Context section. The institutional context is discussed as being within the parameters of the UN Kosovo Team CDP and UNDAF.
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Partial	The report presents a reconstructed theory of change for each of the 3 program areas and states these were used to guide the evaluation methodology. The text notes that the team developed 6 different intervention models to cover all possible UNFPA interventions - ones for population dynamics, data availability and data process capabilities were apparently developed but not presented. Figure 1 presents a rough outline of an overall logic model but it is not referenced in the text and its purpose is not clear. The adequacy of logic models and theory of change is not discussed.
<i>To ensure a rigorous design and methodology</i>		
4. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?	Yes	The evaluation provides an elaborate evaluation matrix presented in Annex 2. The matrix includes the evaluation questions, assumptions, indicators, data sources and methods for data collection and summaries of evaluation findings organized by key outcomes.
5. Are the tools for data collection described and their choice justified?	Yes	The data collections tools are described and the interview protocols for different stakeholder groups are provided in Annex 6. The use of mixed methods approach was justified in the methodology section as necessary for triangulation of information from different sources.
6. Is there a comprehensive stakeholder map? Is the stakeholder consultation process clearly described (in particular, does it include the consultation of key stakeholders on draft recommendations)?	Partial	The report does not provide a stakeholder map, however, it is noted that the evaluation team, in cooperation with the UNFPA staff, identified the stakeholders associated with activities, output, and outcomes. The report acknowledges consultations with the members of the reference group, including for taking time to attend the inception and out-briefing, and to provide comments on key deliverables.
7. Are the methods for analysis clearly described for all types of data?	No	Neither the overall synthesis report or country report are explicit about the data analysis strategy.
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Section 1.1.4. only briefly describes the limitations in the evaluation and potential bias introduced in the selection of evaluation participants and how this bias was mitigated. It does not address that training follow up interviews were not carried out as was done in other countries; however, this is noted in the overall synthesis report as being a limitation of the Kosovo study. The overall report also provides further information on limitations and mitigation efforts.

9. Is the sampling strategy described?	Yes	In this report, the sampling approach is described only as purposive and with non-random selection of key informants (noted as not representative of the target population) of both implementation partner organisations and beneficiaries. A more complete description of the sampling approach is provided in the overall report.
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The mixed methods approach is appropriate for collecting and analysing disaggregated data as shown in the list of persons consulted, although evaluators did not present disaggregated data in the findings section since this evaluation was predominantly qualitative.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The design and methodology is appropriate for discussing cross-cutting issues since the evaluators were explicit and careful to show how the cross-cutting issues (gender, human rights, and disability responsiveness) are assessed and presented in the report.
3. Reliability of Data	Yes No Partial	Assessment Level: Very good
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	It is noted that the evaluators triangulated data sources, data collection methods and tools, and validated data and findings through regular exchanges with the UNFPA programme staff and the Evaluation Reference Group. Document sources are regularly footnoted and qualitative findings are regularly attributed to specific stakeholder groups.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluators explicitly state that the sources of qualitative and quantitative data were from documentary review, group and individual interviews, focus groups and field visits to programme sites as appropriate.
3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues?	Yes	The evaluators noted the potential causes of bias and addressed the threats to reliability of the data by selecting interviewees representing a diverse range of institutional viewpoints on key topics under review. It is noted that all the interviews were conducted in private without the presence of any UNFPA staff members.
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	The evaluators ensured that all participants were treated as autonomous and free to choose whether or not to participate in the evaluation. The ethical considerations were incorporated in the interview protocols in which participants were informed about the confidentiality of the interview data and that their participation was voluntary.
4. Analysis and Findings	Yes No Partial	Assessment Level: Good
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	The evidence for the findings are indicated in footnotes referencing document reviews; and in some cases key informant interviews are cited in the text as being the source of findings.
2. Is the basis for interpretations carefully described?	Yes	The report clearly shows how interpretation bias was addressed. The evaluators collected and interpreted qualitative and quantitative data according to the indicators presented in the evaluation matrix in Annex 2, the results chain in Annex 4, and the framework in Annex 5.
3. Is the analysis presented against the evaluation questions?	Partial	The findings are organized by programme area and each area analysed by evaluation criteria, however the evaluation questions are only shown in the evaluation matrix and not in the main report.
4. Is the analysis transparent about the sources and quality of data?	Yes	As noted earlier, the sources of data used are referenced in footnotes although use of direct quotes could improve the transparency of the analysis. The quality of data was described when necessary.

5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The evaluators used the evaluation matrix and the theory of change model to assess the cause-to-effect relationships between the intervention and its end results. The evaluators highlighted impacts where appropriate and the unintended positive and negative outcomes of the intervention.
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings show evaluation outcomes by programme area for specific groups targeted by the country programme, especially women, adolescents and youth and minority communities, especially Kosovo Serbs.
7. Is the analysis presented against contextual factors?	Yes	In each case, the findings are shown against the relevant contextual factors in which (and with which) the country programme works, including specific population groups, institutions, and the relevant national and international policies.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The evaluators were careful to examine the cross-cutting issues including, especially, gender and human rights of women, adolescents and youth and minority communities, especially Kosovo Serbs.
5. Conclusions	Yes No Partial	Assessment Level: Good
<i>To assess the validity of conclusions</i>		
1. Do the conclusions flow clearly from the findings?	Yes	The conclusions are clearly drawn from the findings and are numbered, indicating the relevant evaluation criteria.
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated?	Partial	The conclusions provide an solid understanding of the underlying issues for each programme area in respect to effectiveness and sustainability only. Relevance, efficiency and cross-cutting themes are not explicitly addressed. To make this section more clear, it would be helpful if the topic for each conclusion was highlighted. In addition, overall summary statements would be useful for understanding the main achievements, strengths and challenges of the CP.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no indication of bias since the conclusions are clearly based on evidence from the findings showing both positive and negative elements.
6. Recommendations	Yes No Partial	Assessment Level: Fair
<i>To ensure the usefulness and clarity of recommendations</i>		
1. Do recommendations flow logically from conclusions?	Yes	The recommendations logically flow from the conclusions, and each recommendation is linked to specific conclusions.
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Partial	Some recommendations specify to whom they are directed while others are not targeted. Operational and technical implications are provided, although some explanations lack depth. They could also be more clearly presented by highlighting the subject of each recommendation.
3. Do recommendations appear balanced and impartial?	Yes	The recommendations flow clearly from the findings and clearly specify the relevant conclusions without any evidence of bias.
4. Is a timeframe for implementation proposed?	Yes	The recommendations are classified into high or medium priority for both the strategic and programmatic recommendations. The timeframe for implementation is not proposed but it is clear in the initial sections of the report that recommendation will be used for the next CP planning process.

5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The classification into high and medium seems appropriate to facilitate management response. The MR should also be facilitated by the evaluators having obtained feedback from the CP stakeholders on the recommendations.
7. Gender	0 1 2 3 (**)	Assessment Level: Good
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i>		
1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?		<p>a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3) GEEW was not mentioned as being part of the objectives or scope of the evaluation. GEEW was included in responses to all evaluation criteria and selected indicators. The GEEW key findings summaries in the evaluation matrix in Annex 4 is well-done in addition to standalone section on gender equality in section 4.2.2 on Comprehensive Sexuality Education (Non-Formal Education). (Score = 0).</p> <p>b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3) There is no standalone criterion on GEEW, but the assumptions and indicators for EQ1 and EQ3 assesses how GEEW is implemented in line with the priorities set by the international and national policy frameworks and aligned with the UNFPA policies and strategies. (Score = 3).</p> <p>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) GEEW was integrated in EQ1(a) To what extent is the UNFPA programme adapted to the needs of women, adolescents and youth, people at risk of HIV infections, disabled, older persons and Roma? and EQ3. To what extent did the outputs contribute to the achievement of these planned outcomes in mainstreaming of provisions to advance gender equality? (Score = 3).</p> <p>d. 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?(Score: 0-3) Assessment of the sufficiency of information collected during the implementation period on specific results, including on GEEW, was addressed within EQ6 (Efficiency). However, GEEW was not explicitly mentioned in the question. (Score = 2).</p>
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		<p>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? (Score: 0-3) The evaluation methodology does not mention gender considerations. (Score = 0).</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3) The methods and data collection tools included key informant and focus group interviews with primary beneficiaries that included women and youth (interviewed as separate groups). (Score = 3).</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) It is clearly stated in the report that triangulation and validation of the data and findings was done, but there is no clear evidence how analyses ensured that data is disaggregated except having a gender balance in stakeholders consulted (with more women interviewed). The team got feedback from CO and ERG on both the draft and final reports to validate accuracy of findings, thus ensured the credibility of the data. (Score = 2)</p> <p>d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3) Yes, this was ensured with samples representing the primary beneficiaries as noted above. (Score = 3).</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3) The interview and focus groups protocols indicate that data collectors complied to ethical guidelines for informed consent and respect for confidentiality. The Cluster Evaluation Synthesis report also notes that UNEG ethical guidance was followed. However, neither report is explicit about the appropriateness of interview and focus groups</p>

		locations. (Score = 2).
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		<p>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? (Score: 0-3) Gender-analysis is evident in the background section especially in regards to the assessment of policies and practice; for example, it is noted that women's unemployment rate (36.6%) is higher than men's unemployment rate (28.7%), and there are graphic representations of the pyramids of the Kosovo population 2011 and 2017 by gender and age, and maternal mortality. (Score = 3).</p> <p>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? (Score: 0-3) While the findings include data analysis that explicitly triangulates the sources of data, there is minimal use of quotes or stories to highlight the voices of different groups. There is presentation of gender-disaggregated data. (Score = 2).</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) The unanticipated effects of the intervention on human rights and gender equality is reported in Key Findings section and occasionally in the summaries in the evaluation matrix. (Score = 3).</p> <p>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3) The report provides specific conclusions and recommendations that explicitly addresses gender equality, particularly Recommendations 1 and 8 that are linked with Conclusions 1 and 12 respectively. (Score = 3).</p>

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totaling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

(**) Scoring uses a four point scale (0-3).

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

Overall Evaluation Quality Assessment

Quality assessment criteria (scoring points*)	Assessment Levels (*)			
	Very good	Good	Fair	Unsatisfactory
1. Structure and clarity of reporting, including executive summary (7)			7	
2. Design and methodology (13)			13	
3. Reliability of data (11)	11			
4. Analysis and findings (40)		40		
5. Conclusions (11)		11		
6. Recommendations (11)			11	
7. Integration of gender (7)		7		
Total scoring points	11	58	31	

Organizational unit: **Year of report:** 2019

Title of evaluation report: UNFPA Country Programme Evaluation Report: The Republic of North Macedonia (period covered 2012-2018)

Overall quality of report: Very Good **Date of assessment:**

Overall comments: This evaluation is one of four undertaken as part of a cluster evaluation of UNFPA country programmes under the same administrative unit within the Eastern Europe and Central Asia region. The methodological approach and areas of enquiry are consistent for each. This is a comprehensive evaluation of UNFPA Republic of North Macedonia Country Programme. It shows the CP's accomplishments, including by use of reliable data and a clear presentation of evaluation findings. Data was collected through a range of techniques including document analysis, structured and unstructured interviews, field visits as well as direct observation, and focus group discussions using a purposive sample. There was a thorough consultation process with stakeholders and the data collection, within the limits of what was possible, was comprehensive and undertaken in accordance with good ethical practice. The findings were based on the data and were well-balanced between those that were positive and those that indicated areas for improvement. The conclusions clearly flow from the findings but could go further by providing an overview of the strengths of the programme's design and implementation. The recommendations could be more targeted to other primary intended users of the evaluation especially government agencies and other funders. Additional areas for improvement include the need for data analysis processes to be explained in the methodology section, and the need for a more concisely presented Executive Summary to increase its usability.

Assessment Levels

	Very Good	strong, above average, best practice	Good	satisfactory, respectable	Fair	with some weaknesses, still acceptable	Unsatisfactory	weak, does not meet minimal quality standards
--	--	--------------------------------------	---	---------------------------	---	--	---	---

Quality Assessment Criteria	<i>Insert <u>assessment level</u> followed by main <u>comments</u>. (use 'shading' function to give cells corresponding colour)</i>		
1. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Fair
<i>To ensure the report is comprehensive and user-friendly</i>			
1. Is the report easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors?	Yes	The report is easy to read and understand. It is written in an accessible language appropriate for the intended users.	
2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations)	Yes	The report is at the maximum length.	
3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The report structure is coherent with a clear delineation of sections consistent with UNFPA standards. The key findings are organised by the four focus programme areas and evaluation criteria applied to each area.	
4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process?	Partial	The annexes are comprehensive and contain all of the expected material. In addition, the list of participants interviewed, the evaluation matrix, and the logical framework are exceptionally detailed. However, the report does not conform to UNFPA standards of providing detailed data collection protocols for different stakeholder groups in the annex.	

Executive summary

5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Partial	The executive summary is informative but it is not suitable as a stand-alone section. The paragraphs are long and lack spaces in between. All the recommendations target UNFPA and are not prioritized as in the main report.
6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Yes	The executive summary follows the desired structure and it contains relevant information suitable for people who might not read the main report.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	No	The executive summary exceeds the maximum length by two pages, and would be considerably longer if spacing was used.
2. Design and Methodology	Yes No Partial	Assessment Level: Fair
<i>To ensure that the evaluation is put within its context</i>		
1. Does the evaluation describe the target audience for the evaluation?	Yes	The target audience is well-described. It is noted that intended users are UNFPA national and regional offices, government counterparts, and other development partners.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The development and institutional contexts are described and constraints explained in Chapter 2. The report provides detailed analysis and overview of the population dynamics and the sexual and reproductive health and maternal and child health in North Macedonia.
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes	The reconstruction of the intervention logic is well-explained in the annex. In the main report, the reconstruction is also well-described. The evaluation notes the evolution of the UNFPA theory of change starting with the 2014-2017 and 2018-2021 strategic plans. The reconstructed theory of change is found in Figure 5.
<i>To ensure a rigorous design and methodology</i>		
4. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?	Yes	The evaluation framework is clearly described in both the text, Table 6, and elaborated in Annex 2. The questions, assumptions, indicators, data sources and methods are well-described including summaries of evaluation findings organized by the four programme areas.
5. Are the tools for data collection described and their choice justified?	Yes	The data collections tools are described in the main report and the details provided in the evaluation matrix. The mixed methods approach was justified as necessary for triangulation of information from different sources. It is noted that "All the questions in evaluation matrix were answered and served as a ground for elaboration of the findings, conclusions and recommendations."
6. Is there a comprehensive stakeholder map? Is the stakeholder consultation process clearly described (in particular, does it include the consultation of key stakeholders on draft recommendations)?	Yes	The report provides an overview of the stakeholders consulted by programme area and gender in Table 2, and a detailed description of stakeholders by programme area in Annex 7.
7. Are the methods for analysis clearly described for all types of data?	Partial	The evaluation is not explicit about data analysis strategy per ToR request, but it is noted that "The analysis was built on triangulating information obtained from various stakeholders' views as well as with secondary data and documentation reviewed by the team."
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Section 1.3.3. describes the time and resource limitations in the evaluation and potential bias introduced in the selection of evaluation participants and how this bias was mitigated.
9. Is the sampling strategy described?	Yes	The sampling approach is described as purposive and non-random selection for evaluation sites to visit and selection of the key informants. The criteria for the purposive selection was made with an attempt to achieve a balance of key informants according to region, focus area, and female versus male respondents. Annex 3 provides exemplar details of the site visit planning calendars for each programme.
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The mixed methods approach is appropriate for collecting and analysing disaggregated data as shown in the summaries of the list of persons consulted in Tables 3, 4 and 5.

11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The theory-based design and mixed methods approach are appropriate for discussing cross-cutting issues. As a result, the evaluators were explicit and careful to show how the cross-cutting issues (gender, human rights, and disability responsive) are assessed and presented in the text and in the annexes of the report.
3. Reliability of Data		
	Yes No Partial	Assessment Level: Very good
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	It is noted that the evaluators triangulated data sources; data collection methods and tools and document sources are regularly footnoted and qualitative findings are regularly attributed to specific stakeholder groups.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluators explicitly state that the sources of qualitative and quantitative data were from documentary review, group and individual interviews, focus groups and field visits to programme sites as appropriate. The evaluators question the reliability of population data in the absence of population census considering that the last census was done in 2002.
3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues?	Yes	The evaluators state the potential causes of bias and addressed those threats to reliability of the data by selecting interviewees representing a diverse range of institutional viewpoints on key topics under review independent of UNFPA. It is noted that the team conducted all interviews in private without any UNFPA staff present to avoid the possibility of bias from the presence of UNFPA staff.
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	As noted in the cluster report, the evaluators ensured that all participants were treated as autonomous and participants were informed about the confidentiality of the interview data and that their participation was voluntary.
4. Analysis and Findings		
	Yes No Partial	Assessment Level: Very good
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	The report uses footnotes to reference the evidence for the findings from the document reviews and in-text quotes for key informant interviews as evidence from the qualitative sources. The evaluation team did a good job providing summary findings in textboxes for each evaluation question by criteria.
2. Is the basis for interpretations carefully described?	Yes	The report clearly shows how interpretation bias was addressed. The evaluators used the outcome and input indicators to interpret qualitative and quantitative data throughout the Findings section.
3. Is the analysis presented against the evaluation questions?	Yes	The findings are presented against the evaluation questions further organized by evaluation criteria for each of the four programme areas.
4. Is the analysis transparent about the sources and quality of data?	Yes	As noted earlier, the sources of data used are referenced in footnotes although use of more direct quotes could improve the transparency of the analysis. The quality of data was described when necessary.
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The evaluators used the evaluation matrix and the theory of change model to assess the cause-to-effect relationships between the intervention and its end results. For example, it is noted that the results chains of the Youth and Adolescent programme have unclear and weak cause-effect linkages. As such, "This lack of clarity is reflected in some of the indicators. The formulation of indicators does not clearly capture the outcomes or outputs". The evaluators highlighted some unintended outcomes of the intervention.

6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings show evaluation outcomes by programme area for specific groups targeted by the country programme, especially women, adolescents and youth, young key population at risk, Roma, and people with disabilities. For example, it is noted that the evaluative evidence suggests that the needs of the target group of Roma and people with disabilities have not been sufficiently addressed despite the multitude of poor health outcomes for these groups.
7. Is the analysis presented against contextual factors?	Yes	In each case, the findings are shown against the relevant contextual factors in which (and with which) the country programme works, including specific population groups, institutions, and the relevant national and international policies.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	As noted earlier, the evaluators shows analysis on the cross-cutting issues including gender and human rights of women, adolescents and youth, Roma, and people with disabilities.
5. Conclusions	Yes No Partial	Assessment Level: Very good
<i>To assess the validity of conclusions</i>		
1. Do the conclusions flow clearly from the findings?	Yes	Each conclusion is grounded in key findings that respond to the relevant evaluation criteria. However, the report does not provided a good overview of the design and implementation strengths and weaknesses of the programme.
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated?	Yes	The evaluation team set out 16 conclusions covering both strategic and programmatic components that clearly reflect the underlying issues of the country, program contexts, policies, and system being evaluated. For example, regarding South-South cooperation, it is noted that "UNFPA North Macedonia CO could expand its successes to other countries of the region and "export" its know-how and expertise to other countries."
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions are clearly drawn from the findings and do not reflect any bias as they show both positive and negative outcomes of the evaluation.
6. Recommendations	Yes No Partial	Assessment Level: Fair
<i>To ensure the usefulness and clarity of recommendations</i>		
1. Do recommendations flow logically from conclusions?	Yes	The recommendations logically flow from the conclusions, and each recommendation is linked to specific conclusions. Some recommendations [2, 3, and 4] are linked to more than one conclusion, as such, there are 11 recommendations compared to 16 conclusions.
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Partial	Most of the recommendations are targeted to UNFPA RoNM CO. It should be noted that a partnership mapping reveals several key partners as primary intended users of the report. These partners, especially the government agencies and other funders were not targeted. Operational and technical implications are provided, although some explanations lack depth.
3. Do recommendations appear balanced and impartial?	Yes	The recommendations are impartial and balanced.
4. Is a timeframe for implementation proposed?	Yes	The recommendations are classified into high or medium priority for both the strategic and programmatic recommendations. The timeframe for implementation is not proposed but it is clear in the initial sections of the report that recommendation will be used for the next CP planning process.

5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The recommendations are clearly prioritized and rated high, medium or low to facilitate management response.
7. Gender	0 1 2 3 (**)	Assessment Level: Good
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i>		
1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?	2	<p>a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3) GEEW is not explicitly mentioned in the objectives but was a theme in the evaluation. (Score = 2).</p> <p>b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3) There was no standalone criterion on GEEW, neither are there indicators for GEEW in EQ1 that assesses the extent to which the UNFPA programme adapted to the needs of women. It was considered under other criteria. (Score = 1).</p> <p>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) GEEW was integrated in EQ1(a) To what extent is the UNFPA programme adapted to the needs of women, adolescents and youth, people at risk of HIV infections, disabled, older persons and Roma? and EQ3. To what extent did the outputs contribute to the achievement of these planned outcomes in mainstreaming of provisions to advance gender equality? (Score = 3).</p> <p>d. 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?(Score: 0-3) Assessment of the sufficiency of information collected during the implementation period on specific results on GEEW was addressed within EQ3. (Efficiency). (Score = 3).</p>
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?	2	<p>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? (Score: 0-3) The evaluation methodology was gender responsive with regards to the methods and tools for data collection such as interviews. More women representatives of stakeholders, trainers, and beneficiaries were interviewed. However, there was not an explicit description of how a gender responsive evaluation process was ensured. (Score = 2).</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3) The methods and data collection tools included key informant interviews with primary beneficiaries that included women and youth but the report does not show their voices in terms of quotes. (Score = 2).</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) It is clearly stated in the report that triangulation and validation of the data and findings was done, but there is no clear evidence how analyses ensured that data is disaggregated except having a gender balance in stakeholders consulted (with more women interviewed). (Score = 2)</p> <p>d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3) Yes, this was ensured with samples representing the primary beneficiaries as noted above. (Score = 3).</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3) Although the report is not explicit about the appropriateness of interview and focus groups locations, it is noted in the report that primary beneficiaries included women and youth. Participants gave informed consent and were informed that their participation was voluntary and that their data was confidential. It is also noted that UNEG ethical guidance was followed. (Score = 2).</p>

<p>3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?</p>		<p>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? (Score: 0-3) Gender-analysis is evident in the background section especially assessment of policies and practice. (Score = 3).</p> <p>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? (Score: 0-3) While the findings include data analysis that explicitly triangulates the sources of data and provides clear gender-disaggregated data, there is minimal use of quotes or stories to highlight the voices of different groups. (Score = 2).</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) The unanticipated effects of the intervention on human rights and gender equality are reported in Key Findings. For example, it is noted that that there is no specific focus on very young adolescent girls and that UNFPA has not attached sufficient attention to the situation faced by girls at risk of child marriage as required in the Strategic Plan. (Score = 3).</p> <p>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3) The report provides a few specific conclusions and recommendations that broadly address gender equality, particularly Recommendations 3 and 10 that relate to policies. (Score = 3).</p>

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totaling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

(**) Scoring uses a four point scale (0-3).

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

Overall Evaluation Quality Assessment

	Assessment Levels (*)			
Quality assessment criteria (scoring points*)	Very good	Good	Fair	Unsatisfactory
1. Structure and clarity of reporting, including executive summary (7)			7	
2. Design and methodology (13)			13	
3. Reliability of data (11)	11			
4. Analysis and findings (40)	40			
5. Conclusions (11)	11			
6. Recommendations (11)			11	
7. Integration of gender (7)		7		
Total scoring points	62	7	31	

Overall assessment level of evaluation report	Very Good			
	Very good very confident to use	Good confident to use	Fair use with caution	Unsatisfactory not confident to use

- (*) (a) Insert scoring points associated with criteria in corresponding column (e.g. - if 'Analysis and findings' has been assessed as 'Good', enter 40 into 'Good' column).
- (b) Assessment level with highest 'total scoring points' determines 'Overall assessment level of evaluation report'. Write corresponding assessment level in cell (e.g. 'Fair').
- (c) Use 'shading' function to give cells corresponding colour.

If the overall assessment is 'Fair', please explain

- How it can be used?

- What aspects to be cautious about?

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory

Consideration of significant constraints

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes

No

If yes, please explain:



Organizational unit: Year of report: **2019**

Title of evaluation report: **UNFPA Country Programme Evaluation Report - The Republic of Serbia**

Overall quality of report: **Good** Date of assessment: **2020 MAY**

This evaluation is one of four undertaken as part of a cluster evaluation of UNFPA country programmes under the same administrative unit within the Eastern Europe and Central Asia region. The methodological approach and areas of enquiry are consistent for each. This report provides a reasonably solid evaluation of UNFPA Country Programme Evaluation in the Republic of Serbia. The strengths of the evaluation include the Analysis & Findings and Gender sections. The evaluation uses a mixed-methods approach linked to an evaluation matrix and well-constructed theory of change and logic models. Findings are consistently presented in the context of these models and key performance indicators. Furthermore, the evaluation consults a diverse sample of stakeholders which is disaggregated by gender. The limitations of the evaluation primarily lie in the information omitted: there is no description of data analysis processes, and minimal reference to limitations, constraints and mitigation strategies, affecting the assessment of the reliability of data (although most of this information can be found in the overall cluster evaluation document). This report provides findings that reflect a decent analysis of the programme's strengths and weaknesses contrasted against contextual factors and performance indicators, though a clear gender and vulnerability analysis is not present in the findings (only in the country context). The conclusions provide a balanced perspective, though are few, and do not expand on the findings nor identify key conclusions under the value added and coordination criteria. Recommendations are coherently linked and are clearly actionable and prioritized, though could be more clearly articulated. Overall, more attention could be given to strengthening the evaluation design and methodology, including data reliability, and to improving the structure and clarity of some sections, including the executive summary.

Assessment Levels

	Very Good	strong, above average, best practice	Good	satisfactory, respectable	Fair	with some weaknesses, still acceptable	Unsatisfactory	weak, does not meet minimal quality standards
--	--	--------------------------------------	---	---------------------------	---	--	---	---

Quality Assessment Criteria	<i>Insert <u>assessment level</u> followed by main <u>comments</u>. (use 'shading' function to give cells corresponding colour)</i>		
1. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Fair
<i>To ensure the report is comprehensive and user-friendly</i>			
1. Is the report easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors?	Partial	The report is accessible and easy to understand, though some grammatical and spelling errors are present. In addition, in some cases there are incomplete sentences or text cuts off (p65) or edits from track changes left in the document (p48).	
2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations)	Partial	At 78 pages, the length of the report strays from the maximum 70 page-limit for CPEs by 8 pages, excluding annexes. The country context is over 10 pages, and could be cut shorter to better meet length requirements.	
3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The report structure is coherent and includes a clear delineation of sections. The use of evaluation criteria to report key evaluation findings cogently covers the three key components of the evaluation.	
4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process?	Yes	Annexes contain most of the required information. The ToRs, evaluation matrix, methodological tools, bibliography, and list of persons consulted are all present. There is some information on the stakeholder consultation process in Annex 3 Schedule of Field Work Activities.	
<i>Executive summary</i>			
5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Yes	The executive summary serves as a stand-alone section, summarizing the findings for relevance, efficiency, effectiveness and sustainability by the three program areas.	

6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Partial	The executive summary follows the desired structure, though does not identify the intended audience, and finding and conclusions are presented separately. There is some blending of recommendations in the conclusions.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Partial	The executive summary, at 4 pages, is within the page limit. However it is text heavy; clarity would be improved with more spacing and shorter paragraphs. This is particularly an issue for the conclusions and recommendations sections which have paragraphs that are over 1/2 page, making them difficult to read.
2. Design and Methodology		
	Yes No Partial	Assessment Level: Fair
<i>To ensure that the evaluation is put within its context</i>		
1. Does the evaluation describe the target audience for the evaluation?	Yes	This report does not explicitly state the target audience for the evaluation, but the information can be found in the overall evaluation report for the cluster.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The description of the development context and its constraints is provided in the Country Context section. The institutional context is discussed as being within the parameters of the UN Serbia Team CDP and UNDAF.
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes	The report methodology mentions that the theory of change was reconstructed in the evaluation design process, and a graphic depiction of the UNFPA strategic plan 2018-2021 theory of change is included as well as a logic model for the Serbia CP.
<i>To ensure a rigorous design and methodology</i>		
4. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?	Yes	The evaluation provides an elaborate evaluation matrix with sections describing the logic chains within the matrix presented in Annex 4. The matrix includes the evaluation questions, assumptions, indicators, data sources and methods for data collection.
5. Are the tools for data collection described and their choice justified?	Yes	The tools for data collection are carefully described and justified. The mixed methods data collections tools include a desk review, site visits, interviews with national counterparts, IPs and development partners, interviews with clients/beneficiaries, and follow-up surveys and group discussions with training participants. Justification for the mixed methods approach and types of data collected through each tool is provided.
6. Is there a comprehensive stakeholder map? Is the stakeholder consultation process clearly described (in particular, does it include the consultation of key stakeholders on draft recommendations)?	Yes	The participation of stakeholders is described in the methodology section of the report. The report details the timeline and contributions of the evaluation reference group, including validation and quality control of draft deliverables (as well as findings and recommendations). A stakeholder mapping exercise is mentioned, and is evident in Annex 8.
7. Are the methods for analysis clearly described for all types of data?	No	Neither the overall synthesis report or country report are explicit about the data analysis strategy.
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	This report does not explicitly describe the methodological limitations, although some limitations and mitigation strategies can be found for each question in the evaluation matrix. The overall synthesis report does include a sub-section on limitations that notes the number of planned interviews were not able to be carried out in Serbia, and a sub-section on general strategies used to reduce bias.
9. Is the sampling strategy described?	Yes	The full set of sampling approaches can be found in the overall cluster report. Within this report, the sampling approach is described for select data collection strategies. For example, the sampling approach is described as purposive and non-random for the selection of sites visited to achieve a balanced review of activities; but the sampling methodology for semi-structured interviews is not described. The sampling strategy for the group discussions with training participants is described, though not for the client/beneficiary interviews.
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The evaluators used a mixed-methods approach, allowing for the collection and analysis of disaggregated data. The evaluators also provide tables on the numbers of respondents reached, which present gender disaggregated data.

11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The design is appropriate for assessing cross-cutting issues, which includes gender equality as a theme to be assessed across evaluation criteria. The evaluators also mainstream an assessment of the program's responsiveness to the needs of vulnerable groups within the relevance criterion.
3. Reliability of Data	Yes No Partial	Assessment Level: Fair
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	It is noted that the evaluators triangulated data sources, data collection methods and tools, and there is evidence of this in the findings. Document sources are regularly footnoted and contrasted against qualitative findings, which are regularly attributed to specific stakeholder groups.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Partial	The evaluators note the sources of qualitative and quantitative data in the findings text and through footnotes. Most data is drawn from documentary review and individual interviews, as well as group discussions with training participants. The report mentions the use of brief surveys conducted with training participants - although survey results are often used as a primary source of quantitative data, there is no visible reference to information drawn from these in the report. The reliability of data sources is not discussed.
3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues?	Partial	The evaluation does not explicitly note the potential causes of bias and threats to reliability of the data, though the methodology ensured the selection of interviewees representing a diverse range of institutional viewpoints on key topics under review. The evaluation also notes gaps in outcome data, largely noting that stakeholders felt some outcome-level changes have not yet been realized and therefore it is too early to assess indicators.
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	The evaluators note that all participants were treated as autonomous and free to choose whether or not to participate in the evaluation.
4. Analysis and Findings	Yes No Partial	Assessment Level: Good
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	The evidence for the findings are indicated in the text and footnotes, referencing documents or key informant interviews. The evaluation team did a good job providing summaries of key findings for evaluation questions in textboxes.
2. Is the basis for interpretations carefully described?	Yes	The evaluators collected and interpreted qualitative and quantitative data according to the indicators presented in the evaluation matrix. Contextual factors affecting results achievement are carefully interpreted.
3. Is the analysis presented against the evaluation questions?	Yes	The findings are organized by evaluation questions per evaluation criteria and programmatic focus area.
4. Is the analysis transparent about the sources and quality of data?	Partial	As noted earlier, the sources of data used are referenced in footnotes although use of direct quotes could improve the transparency of the analysis. The quality of data is not systematically analyzed.
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The evaluators used the evaluation matrix, the theory of change model, and counterfactual analysis wherever possible to explore the cause-to-effect relationships between the intervention and its end results. In addition, UNFPA and other UN Agency and government contributions to results are noted. The evaluators highlighted unintended positive and negative outcomes of the intervention in response to EQ3, mostly noting progress against planned output indicators, though also highlighting the unintended effects of trainings.

6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings occasionally show outcomes by program area for specific groups targeted by the country program, especially women, adolescent girls, older persons and the Roma population. However, the findings presented are primarily at the output level, and do not systematically reflect a gender and human rights analysis to demonstrate differences in access and results. As noted above, this may be due to the fact that the evaluation notes gaps in outcome data, largely noting that stakeholders felt some outcome-level changes have not yet been realized and therefore it is too early to assess indicators.
7. Is the analysis presented against contextual factors?	Yes	The findings are shown against the relevant contextual factors, noting implementation timelines, evolving humanitarian situations, and limiting gender norms.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The evaluators include some examination of cross-cutting issues including analysis of equity and vulnerability issues, primarily in the context section.
5. Conclusions		
	Yes No Partial	Assessment Level: Good
<i>To assess the validity of conclusions</i>		
1. Do the conclusions flow clearly from the findings?	Yes	The conclusions flow logically from the findings, which are organized around the 4 UNEG evaluation criteria of relevance, effectiveness, efficiency and sustainability.
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiatives/system being evaluated?	Partial	While the evaluation conclusions demonstrate a clear understanding of the program under evaluation, they do not extend beyond the findings. This section could be improved by having overall summary statements for each criterion, particularly for Effectiveness which has 8 individual conclusions, and on the strength and weaknesses of the CP. Conclusions are very briefly described within 2 pages, and there are no conclusions which clearly draw on UNFPA's value added or coordination, key criteria assessed in the evaluation.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no evidence of bias since the conclusions are clearly based on evidence from the findings showing both positive and negative elements.
6. Recommendations		
	Yes No Partial	Assessment Level: Fair
<i>To ensure the usefulness and clarity of recommendations</i>		
1. Do recommendations flow logically from conclusions?	Yes	The recommendations logically flow from the conclusions, and each recommendation notes the conclusions on which it is based.
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Partial	The recommendations specify to whom they are directed within the text, with most being directed to the UNFPA CO, and provide operational and technical implications. However, most could be more clearly written to increase their usefulness for decision makers. They could also be more clearly presented by highlighting the subject of each recommendation. An example of the need for additional clarity is #2 : "Having in mind an affirmative legislative framework and national orientation toward joining EU, the UNFPA CO should continue their activities in emphasizing their complementarities and providing added value in terms of collaboration and advocacy efforts related to the needs of population groups that are within the UNFPA mandate."
3. Do recommendations appear balanced and impartial?	Yes	The recommendations flow clearly from the findings and clearly specify the relevant conclusions without any evidence of bias.
4. Is a timeframe for implementation proposed?	Yes	The recommendations are classified into high or medium priority. The timeframe for implementation is not proposed but it is clear in the initial sections of the report that recommendation will be used for the next CP planning process.

5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The recommendations are brief, though specific, and their classification into high and medium seems appropriate to facilitate management response.
7. Gender	0 1 2 3 (**)	Assessment Level: Good
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i>		
1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?	3	<p>a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3) GEEW is considered as a cross-cutting theme within the evaluation scope and objectives. (Score = 3).</p> <p>b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3) There is no standalone criterion on GEEW, but as mentioned, it is included as a cross-cutting theme to be addressed, and the assumptions and indicators for EQ1 and EQ3 assesses how GEEW is implemented in line with the priorities set by the international and national policy frameworks and aligned with the UNFPA policies and strategies. (Score = 3).</p> <p>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) GEEW was integrated into questions under the relevance and effectiveness criterion. (Score = 3).</p> <p>d. 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?(Score: 0-3) Assessment of the sufficiency of information collected during the implementation period on specific results on GEEW is not specifically addressed, though there is some analysis provided within the relevance section. (Score = 1).</p>
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?	2	<p>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? (Score: 0-3) The evaluation methodology is gender responsive with regards to the methods and tools for data collection. (Score = 3).</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3) The methods and data collection tools included key informant and focus group interviews with stakeholders and the stakeholders consulted are disaggregated by sex. However, there is minimal information provided on how the sample size was drawn to ensure representation from groups, though more females were interviewed than males based on the disaggregated data provided. (Score = 2).</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) Data sources were diverse. It is clearly stated in the report that triangulation of the data and findings was done, but there is no clear evidence of how analyses ensured that data is disaggregated except having a gender balance in stakeholders consulted (with more women interviewed). There is no information on evaluation reference group participation or other validation processes used. (Score = 2)</p> <p>d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3) Yes, beneficiaries were engaged within the evaluation. (Score = 2).</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3) The evaluators note following UNEG's Ethical Guidelines for Evaluations. The interview and focus groups protocols indicate that data collectors complied to ethical guidelines for informed consent and respect for confidentiality. However, the report is not explicit about the appropriateness of interview and focus groups</p>

		locations. (Score = 3).
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		<p>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? (Score: 0-3) Gender analysis is evident in the background section, clearly disaggregating data by sex and providing an intersectional analysis of how different social groups are affected by policies, practices and social norms. (Score = 3).</p> <p>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? (Score: 0-3) While the findings include data analysis that explicitly triangulates the sources of data, there is minimal use of quotes or stories to highlight the voices of different groups or clear gender-disaggregated data. (Score = 1).</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) The unanticipated effects of the intervention on human rights and gender equality is reported in the Key Findings section, especially the summaries in the evaluation matrix, however the unintended effects could have been described in a separate section or noted more explicitly to improve clarity. (Score = 2).</p> <p>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3) The report provides 2 recommendations that address gender equality (Recommendations 5 and 8). (Score = 3).</p>

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totaling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

(**) Scoring uses a four point scale (0-3).

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

Overall Evaluation Quality Assessment

Quality assessment criteria (scoring points*)	Assessment Levels (**)			
	Very good	Good	Fair	Unsatisfactory
1. Structure and clarity of reporting, including executive summary (7)			7	
2. Design and methodology (13)			13	
3. Reliability of data (11)			11	
4. Analysis and findings (40)		40		
5. Conclusions (11)		11		
6. Recommendations (11)			11	
7. Integration of gender (7)		7		
Total scoring points		58	42	

Overall assessment level of evaluation report		Good		
	Very good very confident to use	Good confident to use	Fair use with caution	Unsatisfactory not confident to use

- (*) (a) Insert scoring points associated with criteria in corresponding column (e.g. - if 'Analysis and findings' has been assessed as 'Good', enter 40 into 'Good' column).
 (b) Assessment level with highest 'total scoring points' determines 'Overall assessment level of evaluation report'. Write corresponding assessment level in cell (e.g. 'Fair').
 (c) Use 'shading' function to give cells corresponding colour.

If the overall assessment is 'Fair', please explain

- How it can be used?

Although they could be more clearly stated, the evaluation appears to provide conclusions and recommendations that are relevant for informing the next Country Program.

- What aspects to be cautious about?

The main concern is the gaps in the description of the methodology used, most notably how the data was analyzed and what the limitations and constraints were. Since the evaluators did triangulate data sources and the findings appear comprehensive, it can be assumed that the report is useful for decision-making. It is suggested that Executive Summary be reformatted to include shorter paragraphs and separate lines for each recommendation if it is being shared as a stand alone section with other stakeholders.

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory

Consideration of significant constraints

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes No

If yes, please explain: