

Organizational unit:		Year of report:	2019	
Title of evaluation report:	UNFPA Country Programme Evaluation Report - Azerbaijan			
Overall quality of report:	Good	Date of assessment:	25 November 2019	
Overall comments:	<p>The evaluation is well-designed – and provides a thorough picture of the results of the Country Programme. That said, there were a few challenges with the methodological approach, including a lack of information on the data analysis techniques used, and the findings that kept the report from being assessed at the level of "very good". However, the evaluation methodology allowed the evaluators to show that UNFPA had been an effective partner with the government, as well as with NGOs in this middle-income country and identified several areas where improvements could be made, particularly in addressing issues of contraceptive use, reproductive health care in rural areas and in improving youth access to information. The evaluation report includes well grounded conclusions and recommendations.</p>			
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>	
I. 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	While the report was, on the whole, well-written, there were a number of spelling errors. For example, a chart title said: "Chart 2: Bilateral ODA by Secor for Azerbaijan, 2016-2017 average".
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	There is a total of 74 pages, reasonable given that this also includes acknowledgements, key facts and structure of the report.
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	Yes, it is structured logically, and includes the sections delineated in this sub-criteria.
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	The 10 annexes contain all of the required material. They include a stakeholder map, and the consultation process is described in the ToR.
<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 summary is thorough and complete.
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 summary includes all of the required sections, and also includes the findings of the evaluation.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Partial	The executive summary is 7 pages long. The presentation of findings could have been made shorter.

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 described, although somewhat generally since it includes almost all groups related to the UNFPA programme.	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	There is a good description of the evaluation, which is part of a cluster evaluation (together with the evaluation of Georgia and Turkey's country programmes) in the sub-region.	
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 evaluation clearly describes what the intervention logic is, illustrating it through graphs as well in the evaluation matrix. It also shows how the approach/intervention logic evolved from the previous country programme.	
<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	There is a complete matrix in the annexes, and the framework is also described in the text including particularly the evaluation questions.	
5. Are the tools for data collection described and their choice justified?	Yes	The standard mixed-method tools are used and described: document review, in-depth interviews, focus groups and observation. Their selection is based on time and access to information during the evaluation.	
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	There is a complete stakeholder map in annex 6 and the stakeholder process is generally described in the text of the report (with UNFPA staff, UN staff, central- and local-level government counterparts, donors, international and national NGOs, CSOs, service providers and end beneficiaries included as stakeholder categories/consulted). The mapping formed the basis for sampling stakeholders and beneficiaries to be met and programme sites to be visited during the in-country data collection missions. Validation of data focused primarily on exchanges with CO staff and the Evaluation Reference Group which was involved in the process, including being consulted on the recommendations (from the report: "Besides a systematic triangulation of data sources and data collection methods, the validation of data was sought through regular exchanges with concerned UNFPA Country Office staff and a debriefing with the Evaluation Reference Group.").	
7. Are the methods for analysis clearly described for all types of data?	Partial	While the report notes "the populated evaluation matrix was the starting point for analysis, responding to the evaluation questions and arriving at evidence-based findings," the report does not detail the specific methods of data analysis used (it does discuss data validation techniques - such as triangulation).	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	There is a short description of limitations, focused primarily on limitations vis a vis physical access to sites; mitigation measures are described (interviews with direct beneficiaries, experts, and trainers were proposed as mitigating measures).	
9. Is the sampling strategy described?	Partial	The sampling approach to site visits is detailed, resulting in an illustrative sample. Evaluators include the criteria for selection as follows: "1. Existence of stakeholders and beneficiaries targeted by all UNFPA programme components, 2. Existence of stakeholders and beneficiaries related to interventions implemented in the capital and at sub-national level and 3. Existence of stakeholders and beneficiaries related to on-going and completed activities. However, while the evaluation reports lists the specific criteria used - and indicated who was interviewed (in Annex 2) - the details of how	

10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Yes, the methodology enabled the collection of disaggregated data. Quantitative data from government and UNFPA sources is used that were disaggregated in the analysis. In presenting evidence from documents containing quantitative data, there was disaggregation by group and by area. Similarly, qualitative data from interviews and observations were disaggregated by group and area.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	One of the three programmatic foci in data collection was gender and women's empowerment. This required a gender-sensitive methodology which was used. This included reviewing documents to see the extent to which expected outcomes were reported in statistics. There were also in-depth interviews with a sample of providers and beneficiaries, as well as focus groups with beneficiaries. The evaluation questions were specific about the gender issues in all of the programmes, but especially related to gender-based violence.

3. Reliability of Data	Yes No Partial	Assessment Level:	Very good
<i>To ensure quality of data and robust data collection processes are used in the findings</i>			
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluators clearly describe how they've triangulated across sources and data collection methods (with document review and interviews most frequently triangulated).	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluation clearly showed how the quantitative data were reliable (most were, but they suggested a few areas where they were not) and were careful in using interview data to ensure reliability in qualitative data.	
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	While the documentary data were very thorough, the evaluation notes a few limitations (with mitigating measures presented), including in some of the interview data because the interviewers could not go to all of the areas. For these, an effort was made to do distance (e.g. Skype) interviews.	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	The evaluators were careful to note that interviews were anonymous and this was reflected in how the data from interviews was presented in the report. As the report noted: "All interviewees were assured of confidentiality. National evaluation team members closely adhered to the UNEG Ethical Guidelines for Evaluation and the UN Code of Conduct for Evaluations in the UN System."	

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	There were 31 findings and in each case the evidence underpinning the findings was described in detail and by source.	
2. Is the basis for interpretations carefully described?	Yes	For each finding, the evaluators were clear in how they approached the analysis and reached the finding.	
3. Is the analysis presented against the evaluation questions?	Yes	This was done consistently throughout.	
4. Is the analysis transparent about the sources and quality of data?	Yes	The data on which the findings were based is carefully described in each case. Limitations, if any, were shown and the evaluators were careful to not go beyond what the data suggested.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Partial	The evaluators were careful to show the causal connections between the support provided by UNFPA and the observed result, both at output and outcome levels. However, while the extent to which targets were achieved is shown consistently, there was no reference to unanticipated outcomes, although this was included in the ToR.	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	There was reference to intended target groups in all cases. These were varied and included the Ministry of Health and other government personnel, youth, women and vulnerable groups. The analysis highlighted the different outcomes for different groups, which was expected given their varied composition and location.	
7. Is the analysis presented against contextual factors?	Yes	Contextual factors were a key ingredient in considering the findings. A factor that was seen in different findings was the political opposition to some aspects of SRHR based on cultural or religious factors.	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	These issues were central to the UNFPA country programme and therefore to the evaluation. Where applicable, they were used in specific analysis and findings. For example, the relevance question led to the following finding: "With regard to the SRH, GEWE and PD programme components, the CP is in line with the principles of the ICPD Programme of Action that sets goals in the context of sustainable development, gender equity and equality; infant, child and maternal mortality reduction; and the provision of universal access to reproductive health services, including family	

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	There are eight conclusions and in each case the link to the appropriate findings is made.	
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	Yes, the conclusions go beyond the findings and reflect an understanding of the country programme and underlying dynamics/issues.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no evidence of bias.	
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	In each case, the conclusion(s) upon which a recommendation is based is cited.	
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 norms were followed. All recommendations were targeted at the UNFPA Country Office and in each case the activities necessary to implement the recommendation was specified, including, where appropriate, financial issues.	
3. Do recommendations appear balanced and impartial?	Yes	The recommendations appear balanced and impartial, and speak to the breadth of work under the country programme.	
4. Is a timeframe for implementation proposed?	Yes	The timeframe for implementation is the next country programme.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The recommendations are presented in a practical way that would enable a management response. All eight, however, were given high priority.	

7. Gender	0 1 2 3 (**)	Assessment Level:	Very 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	While the evaluation does not include a specific objective on the assessment of human rights and gender equality, it mainstreamed gender equality across the evaluation criteria and questions. EQ3 is dedicated to examining the extent to which UNFPA support strengthened national institutional capacity to design and implement policies to advance gender equality and reproductive rights, for example. The evaluation notes assesses the extent to which information on progress on gender equality and human rights has been collected, and notes that sufficient information was collected (e.g. underscoring, for example, influence on CEDAW reporting).
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		3	While the report does not explicitly specify how the evaluation's methodology will be gender responsive, the evaluation uses a mixed methods approach, with an emphasis on document review, statistics, and interviews and focus group discussion, the latter of which focused gender equality (with discussions on gender equality policies and programmes). A range of data sources were consulted, and triangulation (and other validation techniques) was used. Diverse stakeholders were consulted, including vulnerable communities (in this case, rural communities), though this could have been improved (with additional consultations). The evaluation noted that it followed UNEG ethical guidelines on evaluation, and there is no evidence to suggest these were breached.
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		3	The background section includes normative policies related to human rights and gender equality. For example, the context section states: "Gender inequality continues to be one of the key challenges to realizing SRHR. Although legal guarantees for the promotion of women's human rights are in place, including the laws on gender equality (2006) and prevention of domestic violence (2010) ²⁴ , gender-based discrimination and the lack of effective implementation mechanisms on GBV leave hundreds of women vulnerable to abuse". The analysis triangulates information from different groups, including those vulnerable and marginalized communities, reflected in documents and statistics and in interviews and focus groups. While the evaluation did not discuss any unintended or unanticipated impact on GE and HR, it did include gender responsive recommendations. Three recommendations address GEWE, one of which (Recommendation 7) states "In GEWE UNFPA should continue efforts for monitoring the implementation status of
<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 totalling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).</p> <p>(**) Scoring uses a four point scale (0-3).</p> <p>0 = Not at all integrated. Applies when none of the elements under a criterion are met.</p> <p>1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.</p> <p>2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.</p> <p>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.</p>			

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	40	47	13	
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?

- 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:

[Redacted]

Year of report:

2019

Title of evaluation report:

UNFPA Cluster Evaluation Report: Azerbaijan, Georgia and Turkey Country Programmes - Georgia

Overall quality of report:

Good

Date of assessment:

26 November 2019

Overall comments:

This evaluation is one of three undertaken as part of a cluster evaluation of middle-income country UNFPA country programmes within the Eastern Europe and Central Asia region, all using the same methodological approach and areas of enquiry. As the evaluation process was comprehensive and the report generally well written, this evaluation serves as a valuable resource for the development of the next CPs in this cluster. The evaluation provides a thorough picture of UNFPA Georgia's work, under each focus areas of the Country Programme, as well as a useful set of conclusions and recommendations. In particular, the evaluation framework is well detailed and the report presents a well developed evaluation matrix, which includes a comprehensive listing of sources of information of the evaluation.

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	Although mostly well written, there are a few editing issues, including a number of typos, several instances where references appear as "Error! Bookmark not defined", changes in font size and spacing, and text in footnotes and figures that is too small to easily read (e.g. Table 11).	
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 main body of the report is 73 pages long.	
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 logically structured, with clear distinctions among sections.	
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 do not include methodological tools or information on stakeholder consultation.	
<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 presented as a stand-alone document.	
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	All of the elements are presented as per the suggested structure in this sub-criteria.	
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	While the summary is exactly 5 pages in length, as it is presented, certain sections are difficult to read. Overall, the summary includes more detail than is needed or customary (for an Executive Summary), and the recommendations section in particular is cramped, resulting in difficulty distinguishing among individual recommendations.	

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 primary intended users of the country programme evaluation are identified.	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	Both the development and institutional contexts are clearly presented within the section on country context section, as well as under the description of the country programme.	
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 methodology section includes a table of the evaluation questions and references the programme areas (SRH, GEWE, PD) to which each question is linked. A further explanation of the expected results and activities is provided in the section on country context (Section 3). The intervention logic is evident in the evaluation matrix in the annex. While the evaluation mentions that two different theories of change exist, the evaluation appears not to assess the adequacy of either, or other aspects of the programme logic.	
<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	An overview of the framework is included in the text of the main report and a detailed evaluation matrix is included in the annex. The latter includes all required elements, and is, usefully, quite detailed/specific about sources of information.	
5. Are the tools for data collection described and their choice justified?	Yes	The data collection methods identified were document review, observation, individual and group interviews and focus groups. The report included some explanation of the methods and the way they were used.	
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	It is noted that a stakeholder map was developed by the UNFPA CO and included as Annex 3 (the overall cluster synthesis report also notes that individual stakeholder maps for each CP were developed), however it appears that annex 3 is in fact a list of documents consulted, with the stakeholder map missing. There is a brief description of programme stakeholders and a table that identifies the types of stakeholders consulted for each programme area (p. 18-19). The members of the evaluation reference group are listed at the beginning of the report, where it is also mentioned that they participated in providing feedback to findings and recommendations.	
7. Are the methods for analysis clearly described for all types of data?	Partial	There is a paragraph on data analysis, however it is mainly a discussion on the use of triangulation, a validation technique, and how it was undertaken.	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Methodological limitations noted include an inability to access some regions of the country, the lack of reliable research-based data, and the evaluation team's inability to conduct site visits to all of the large number of implementing partners whose work contributes to country programme results. Triangulation was indicated as a way of addressing the limitations, including the need to reach an illustrative sample of stakeholders and sites.	

9. Is the sampling strategy described?	Yes	The report includes a section on Site and Stakeholder Sampling. It notes that the sampling was purposive and non-random, and that key informants were selected in consultation with UNFPA staff. General criteria are stated and the criteria for participant selection of the one FGD held is quite specific to ensure that beneficiaries and service providers from the range of interventions were involved.
10. Does the methodology enable the collection and analysis of disaggregated data?	No	The data was disaggregated in respect to the number of evaluation participants from each stakeholder group, but gender-disaggregated data was not provided in the main report or for the list of persons interviewed/consulted (in annex 2).
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	Cross-cutting issues includes the extent to which the CP protects and promotes human rights and applies gender-responsive approaches. The evaluators were purposeful in engaging all partners, at least in regions they were able to access. The focus group discussions were designed to capture perspectives of a diverse range of rights holders/UNFPA beneficiaries, although only one representative of each "beneficiary group" was included.

3. Reliability of Data	Yes No Partial	Assessment Level:	Very good
<i>To ensure quality of data and robust data collection processes are used in the findings</i>			
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluators are explicit about their use of triangulation, including through checking and validating findings with the Reference Group.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The methodology relied primarily on qualitative data from multiple sources, and reliability of sources is addressed. The evaluators noted that the unavailability of some quantitative research-based data (MICS at the time) was a limitation, particularly for assessing the quantitative indicators in the programme framework. Quantitative data sources were referenced in the findings, mainly statistical data from programme documents and government sources.	
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 limitations are discussed, and mitigation measures for four of the limitations were listed. A key limitation identified was the "Limited access to beneficiaries" in Abkhazia, Georgia. It is noted that "meetings with reasonable number of participants, including from different regions, would have required more resources in terms of time and finances". Drawing on external programme evaluation and stakeholder interviews was used as a mitigating measure.	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	Ethical guidelines were considered when engaging with vulnerable groups. It is noted that all beneficiary interviewees were assured of confidentiality and that UNEG Ethical Guidelines and the UN Code of Conduct for Evaluations were followed.	

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	This section is clearly presented with detailed evidence systematically provided for each finding.	
2. Is the basis for interpretations carefully described?	Yes	This is well done, with additional context provided as needed to explain the findings.	
3. Is the analysis presented against the evaluation questions?	Yes	The analysis and findings are presented according to each evaluation question.	
4. Is the analysis transparent about the sources and quality of data?	Yes	The sources are routinely identified and different perspectives are provided (for example, that government officials and civil society representatives have different views of the government's commitment to free access to family planning).	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Partial	Causal connections are addressed; tables in the Effectiveness section show the links between activities, indicators, targets, CP outputs and outcomes. However, unintended outcomes/effects (and whether evaluators considered these/whether there were any) is not mentioned in the evaluation report, even though their consideration was part of the ToR.	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	Yes, the analysis discusses whether outcomes differed among groups. This was, for example, done for the National Cancer Screening Programme; the evaluators looked at differences in the quality assurance processes in Tbilisi compared to Gurjaani, a more rural region. The evaluators also noted where improvements that are starting to be seen in government data collection systems should soon enable data to be disaggregated by target group.	
7. Is the analysis presented against contextual factors?	Yes	Contextual factors are routinely provided throughout this section.	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The analysis explicitly addresses these issues. It also notes that the language of 'no one left behind' was introduced after the CPD was approved but that the programme had a focus on vulnerable groups, nonetheless.	

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	Yes, conclusions clearly flow from the findings, with the conclusions linked to the corresponding finding number.	
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 conclusions provide a good understanding of the successes and challenges of the CP.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no apparent bias.	
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 number of the corresponding conclusion (and finding(s)) are provided for each recommendation.	
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 are very clearly written, action-oriented and targeted to the intended user. Note that the section is longer than is normal - 7 pages with most of the space given to the rationale and operational considerations for each of the 11 recommendations - which may deter some readers.	
3. Do recommendations appear balanced and impartial?	Yes	There is no evidence of bias.	
4. Is a timeframe for implementation proposed?	Yes	The section is not explicit about the timeframe, however it is clear that an objective of the evaluation is to inform the next programme cycle.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	They are prioritized as either high or medium, and enable a management response.	

7. Gender	0 1 2 3 (**)	Assessment Level:	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?		Human rights and gender equality appear to have been mainstreamed but the evaluation objectives/scope is not explicit in this regard. HRGE is addressed under both the Relevance and Effectiveness criteria, and there are several sub-questions addressing the country programme's work in advancing gender equality and in addressing the various needs of vulnerable populations. While the evaluation notes that government IT government data systems do not yet provide data to measure progress on gender equality/human rights results, the evaluation does not reference UNFPA's internal monitoring systems and the extent to which its able to capture progress on gender equality.
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		The methodology does not explicitly mention how it is gender responsive, including how the evaluation would ensure that data collected is disaggregated, however a diverse range of sources were consulted, and triangulation was used through. Stakeholder diversity was taken into account, and participants included representatives from vulnerable groups though the sample size for beneficiaries was quite low. Ethical guidelines are noted as guiding the evaluation process and steps were taken to ensure confidentiality.
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		The evaluation includes a background section that references various policies and guidelines and related to gender equality and human rights including, for example, Georgia's National Action Plan on Human Rights (2018-20). The evaluation's findings reflect the voices of different groups in certain cases - for example, in Finding 13 on Healthy Lifestyle Education, the views of stakeholders from state institutions and CSOs are provided – though this could be further strengthened (and made more explicit). Unintended effects of the country programme on gender equality or human rights are not described, though it's unclear whether this is because none were found or the evaluators did not query/assess this. The evaluation includes a specific recommendation on the need to broaden gender transformative programming.

(*) 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 totalling 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	33	54	13	
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?

- 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:

Title of evaluation report:

UNFPA Country Program Evaluation Report (Period Covered 2016-2019): Turkey

Overall quality of report:

Good

Date of assessment:

Overall comments:

The evaluation report is fairly easy to read, using succinct and accessible language, though spelling and grammatical errors and typos are frequent enough to be noticeable and affect report quality. The data collection process consulted a wide variety of stakeholders, although the basis for selecting data collection methods/approaches was not made clear and sources were not disaggregated by gender, age or other demographics to understand how diverse the stakeholder group was. In addition, methods of analysis, beyond triangulation of sources (as a validation technique), and how ethical principles were applied, were not described in detail within the report. Therefore, the potential strength of the methodology was limited.

The evaluation did well to take into account contextual factors that could have limited results, and detailed constraints to the evaluation process, primarily time and resource constraints as well as a lack of outcome-level data. The evaluation featured a large number of findings (48) that responded to the evaluation questions and reflected a gender and vulnerability analysis, though findings were not always substantiated by clear evidence. The conclusions and recommendations were well-drafted and were based on the findings. Gender and human rights were reasonably well-covered.

Assessment Levels



strong, above average, best practice



satisfactory, respectable



with some weaknesses, still acceptable



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>		
I. 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 fairly easy to read, using succinct and accessible language, however spelling and grammatical errors and typos are frequent enough to be quite noticeable and affect the report quality. There are numerous grammatical and spelling issues, like "The interventions supported by UNFPA has contributed to a stronger legal and policy framework"	
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 main report is 69 pages, not including the executive summary, acronyms and acknowledgements.	
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 structured to contain all of the desired sections.	
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 appendices do not include the methodological tools (i.e. interview guides, etc.). There is a reference in the text on UNFPA activities that refers to an Annex 5 (list of consulted persons), that is not there.	
<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?	Partial	The executive summary presents the main results of the evaluation, however, it is not written well and as a stand-alone section. Primarily, this is because findings were first presented by evaluation criteria (effectiveness, efficiency, sustainability, coordination and added value), but then a section called 'conclusions' detailed findings related to 'coherence' as well as additional findings related to the previously mentioned evaluation criteria. As such, it was neither clear nor user-friendly, including for high-level decision-makers.	
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 a structure, including a description of the purpose, objectives, methodology, main findings and conclusions and recommendations.	
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	No	At 7 pages, the executive summary is too long, and does not present the main findings in a concise manner. Rather, there is a findings and conclusions section, which could have been combined and better integrated to shorten the section and improve clarity.	

2. Design and Methodology	Yes No Partial	Assessment Level:	Fair
To ensure that the evaluation is put within its context			
1. Does the evaluation describe the target audience for the evaluation?	Yes	The evaluation describes the target audience for the evaluation. With the overall purpose of accountability and evidence-based decision-making, the primary users are mostly donors and decision-making bodies, including the UNFPA Executive Board, government counterparts, and UNFPA donors.	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	There were two independent sections describing the country and institutional context. For example, the following was described: As a 'pink quadrant' country within the UNFPA business model, interventions are expected to focus on capacity building and policy dialogue, which was further represented within the reconstructed Theory of Change. However, Turkey is unique in that it is also classified as a humanitarian setting due to the growing refugee population as a result of the Syrian conflict. Programmatically, the extensive work in the country has increased the number of indicators to be tracked and stakeholders engaged, presenting a challenge to the evaluation. The 2016 coup also closed down program activities .	
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 evaluation report provides a visual representation of the theory of change developed by the country office. The report very briefly highlights a limitation, or context, surrounding the theory of change: primarily that UNFPA contributes to outcome-level change (defined as state and non-state behavior change) in partnership with other stakeholders, which is not reflected in the theory but perhaps is an assumption that should have been included. More detail could have been provided to more fully assess the adequacy of the Theory of Change.	
To ensure a rigorous design and methodology			
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 described, and the process of developing and assessing the country programme against the framework is detailed within the text (p19) and in the appendices. The evaluation matrix is included and includes the expected information.	
5. Are the tools for data collection described and their choice justified?	Yes	Data collection methods were described and included collection from both quantitative and qualitative sources. The methods were primarily qualitative in nature, with quantitative data largely collected through document review (e.g. analysis of financial data and monitoring data). The rationale for their selection, however, was not fully described, though the limitations mentioned suggest that the choice of primarily qualitative methodologies was due to limited time and resources. The cluster evaluation approach also contributed to the time and resource constraints ultimately influencing the more limited scope/data collection tools.	
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	A stakeholder mapping exercise was conducted and information on their consultation is provided in different sections of the report, including a full list of implementing partners provided as an annex (annex 2). The report mentions the inclusion of a 'list of consulted persons' as an annex, but the annex is not included (annex 5 was described as a list of persons consulted as well as a list of key results, but was not attached in either case). It also states that a stakeholder map is included as annex 2 (p30), but this is only the list of implementing partners (as mentioned). As such, there is not a full stakeholder map, but it can be deduced/pieced together from the methodology, UNFPA programme response and annexes. A timeline is included, which infers that several stakeholder debriefing exercises were conducted throughout the evaluation process (p24).	
7. Are the methods for analysis clearly described for all types of data?	No	The evaluation primarily describes triangulation within the section 'data validation and analysis' and does not provide any reference to varying modes of analysis by data type.	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Limitations of the methodology are provided and include time and resource constraints, stakeholder availability, and a high number of indicators to be assessed. "Inclusion of the Humanitarian Programme into the 6th CP Evaluation" was mentioned as a limitation, but it is not clear why this is considered a limitation of the methodology. Presumably it is because the evaluation was already facing time and resource constraints, and the consideration of the humanitarian programme added additional work for the evaluators for Turkey CPE when compared to Georgia and Azerbaijan, other countries considered in the cluster evaluation. The evaluators overcame this limitation by referring	
9. Is the sampling strategy described?	Yes	The sampling approach was described as 'purposive' and directed by the UNFPA Country Programme Evaluation Handbook, which states, "the evaluators should not aim at obtaining a statistically-representative sample, but rather an illustrative sample." Criteria for the selection of areas to interview stakeholders (which were identified through a mapping exercise) were identified based on location and program type, including areas with a high number of activities covering GEWE and SRH and areas with densely located government and private sector partners, amongst other criteria. Limitations of the sampling approach were mentioned within the section on 'limitations' and included the small sample size (78) and potential respondent bias, which were overcome with the application of triangulation. Time and resources constrained the evaluators ability to reach more stakeholders.	
10. Does the methodology enable the collection and analysis of disaggregated data?	Partial	The methodology selected should facilitate the collection of disaggregated data, however the stakeholder table (Table 4, p22) does not disaggregate stakeholders by sex and the data collection tools were not included as annexes in order to assess whether this data was collected. Data can be disaggregated by stakeholder type.	
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Partial	Primarily qualitative methods are conducive for assessing cross-cutting issues, especially when interviewing diverse groups of beneficiaries. In addition, the evaluation questions for effectiveness and relevance assess the extent to which gender-specific outcomes have been achieved and needs have been met, even for the most vulnerable and marginalised groups. However, it is not clear the extent to which women were included as the consulted stakeholders are not gender disaggregated; in addition, youth were mentioned as a cross-cutting issue of focus, but	

3. Reliability of Data	Yes No Partial	Assessment Level:	Very good
<i>To ensure quality of data and robust data collection processes are used in the findings</i>			
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluation triangulated data, as appropriate, referencing both written documents and reports as well as interviews and focus groups.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	Both qualitative data, mostly in the form of interviews with the selection of stakeholders, and quantitative data in the form of government census and other data, were used. In all cases, they were reliable.	
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 evaluation report made note of limitations in monitoring data, presenting a constraint in accessing and using quantitative data. This constraint was mitigated through the review of documents.	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	There is evidence to suggest that confidentiality was respected within the evaluation process. However, it is not clear whether the evaluation considered the needs of different groups interviewed, and any potential issues of discrimination.	

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 48 findings are usually substantiated by evidence, though the flow of the discussion could be improved to draw out the evidence more clearly. For example, a key finding presented under relevance (finding 4) was that the political environment has not been enabling to take fundamental actions on transforming gender norms and roles, however no data is provided to support this finding. In fact, most evidence provided within this section references a supportive institutional framework for gender equality. Frequently the report also mentions that there are "supporting" or "verified" examples from FGDs and In-depth interviews, but does not provide an example (see page 44). However, this is not always the case, for example directly following the lapse in substantiated evidence, the next	
2. Is the basis for interpretations carefully described?	Yes	The basis of interpretations is described. For example, the evaluation report highlights the importance of UNFPA's work in terms of policy influence and achievements, going on then to note the ratification of the Istanbul Convention as an important step towards preventing violence, protecting women and prosecuting criminals, providing a strong legal tool for women to protect themselves, while simultaneously increasing access to quality services envisaged within this convention through standard operating procedures. However, while the basis for interpretation is well	
3. Is the analysis presented against the evaluation questions?	Yes	The analysis is presented against the evaluation questions.	
4. Is the analysis transparent about the sources and quality of data?	Yes	There is not much analysis on the quality of data, though reference to sources are made broadly (i.e. UNFPA, interviewed stakeholders, focus group participants, etc). Sources of quantitative data from the desk review are also referenced, when used, including the full names of reports, policies, DHS surveys, etc.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Partial	While cause-and-effect links are sometimes provided, the evaluation report is primarily descriptive with minimal quantitative data or direct quotes and insights drawn from focus group discussions and interviews to strengthen drawn links. For example, page 49-50 describes the number of beneficiaries reached in the peer education programme, stating that 925 young people acquired skills and knowledge on SRH issues in 2017 and 1720 in 2018. Triangulated evidence from interviewed participants to support knowledge application and acquisition was: "interviewed participants were grateful for having been given the opportunity for using new approaches, like theatre-based knowledge change, and there are indications that the new method has been effective in the universities." While the sentence continues to state that it was perceived to be cost-effective, there is no statement drawn from interviews that demonstrate how it has changed behaviors and knowledge of young people, the primary outcome in this case. In fact, later in this section it states that the workshops and trainings have not changed anything in adolescent life. Additional quotes from conducted interviews and groups could have been used to strengthen the analysis, whether making positive or negative claims about programme outcomes. Another example includes reference to a hotline for HIV-positive persons, which is now operating 7 days per week. The effectiveness of this intervention could have been measured through asking relevant stakeholders about the frequency of use of the hotline. A good example of a drawn cause-and-effect link is the discussion of the guidebook on mainstreaming BADV	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	While cause-and-effect links drawn are sometimes weak, the outputs and limited discussion of outcomes are presented across different target groups, referencing training and knowledge acquisition by lawyers, youth, HIV-positive persons, and sex workers.	
7. Is the analysis presented against contextual factors?	Yes	Contextual factors were referenced, where relevant. For example, the report highlights how availability and accessibility of SRH services is more limited for refugee women as a result of language, financial and legal barriers and therefore notes UNFPA's focus on decreasing these barriers through programme activities, like increasing access to interpreters or providing dignity kits to women. Furthermore, the evaluation notes that tensions between host and refugee communities sometimes interfered with the distribution of kits	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The analysis does well to discuss cross-cutting issues. For example, under relevance, it highlights vulnerable groups, young people, women and refugees involvement in the development of the country programme. Reflection on this cross-cutting issue (gender equality) is quite robust, included in 12 associated findings. Other cross-cutting issues are addressed in other areas.	

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 conclusions are directly linked to findings both in content as well as in the effective use of bold and underlined text and explicit links to evaluation questions.	
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 conclusions do well to link and bring together different findings, especially conclusions 1-3 which provide a more detailed discussion.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no bias present in the conclusions.	
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 conclusions and recommendations were included within one section, demonstrating a clear link between them.	
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 are targeted, well-written and action-oriented, though could have increased specificity and targeting with additional reference to their human and financial implications. The technical implications were <u>adequately covered</u> .	
3. Do recommendations appear balanced and impartial?	Yes	Recommendations appear balanced and impartial.	
4. Is a timeframe for implementation proposed?	Yes	Most recommendations are directed towards the next cycle, and then prioritized as 'high', 'medium' or 'low' priority.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The recommendations are either high or medium. There are fewer high priority than medium, and they are concentrated in the strategic as contrasted with the programmatic recommendations. The recommendations are therefore clearly prioritized and presented to facilitate management response, defining practical implications of implementation.	

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	There is not a specific objective assessing human rights and gender equality, but it can be inferred within the first objective on compliance with national and international frameworks and institutional policies. Gender and human rights were mainstreamed into evaluation criteria, primarily on relevance and effectiveness, through the evaluation questions, and there are several evaluation questions, sub-criteria (assumptions) and associated indicators, which seek to measure the integration of GEEW into the country programme. The evaluation does note limitations in monitoring data collected during programme implementation, but does not make specific reference to indicators measuring human rights and gender equality.	
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?	2	There is little reference to how gender was considered in the evaluation's methodology, including in the collection of data. The evaluation uses a mixed methods approach, with a focus on qualitative data. However, very minimal data was disaggregated by sex (unless focused on women's health). The sampling frame is quite diverse, but the lack of gender-disaggregated or age-disaggregated data on the sample limits the understanding on the extent of diversity of the sample. Validation is applied and sources referenced, particularly if a source from the document review. The evaluation noted adherence to UNEG ethical standards, and confidentiality appears to be upheld.	
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?	2	The evaluation includes a background section in which an analysis of the context and a gender analysis, including references to varying access to rights as a result of vulnerability status, is presented. The findings, while based on triangulated data, do not always transparently (or explicitly) triangulate the voices of different social groups, at times referencing "interview respondents" and, at other times, mentioning the specific respondent category, like the police. None of the evaluation questions specifically queried about unintended outcomes, despite being a requirement within the ToR. In addition, no unanticipated effects on human rights or gender equality were specifically noted or made clear. The evaluation report has specific and clearly designed recommendations about gender.	
<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 totalling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).</p> <p>(**) Scoring uses a four point scale (0-3).</p> <p>0 = Not at all integrated. Applies when none of the elements under a criterion are met.</p> <p>1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.</p> <p>2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.</p> <p>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.</p>			

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	33	47	20	
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?

The conclusions and recommendations are clear, although some of the underlying findings have issues.

• What aspects to be cautious about?

The extent to which some of the data can be used as a basis for future analysis.

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: