

Organizational unit:

Year of report:

Title of evaluation report:

Overall quality of report:

Date of assessment:

Overall comments:

This is a country programme evaluation of the 4th country programme for Uzbekistan for the period between 2016 and 2020 . The evaluation is well done, and uses a mixed method approach including document review, direct observation, informal and semi-structured face-to-face interviews and focus group discussions. The sampling framework was quite clear, comprehensive and was presented in a table, disaggregated by stakeholder group, location and gender. It incorporated the voices of a diverse group of stakeholders, including the most vulnerable, through several focus groups. However, the mitigation strategy for the one limitation mentioned, access to beneficiaries due to time and sensitivity of the topics discussed, could have been made more clear through the inclusion of a more extensive limitations and mitigation strategies section as well as a section on ethics within the methodology. The findings are well validated, with associated conclusions and recommendations which are forward-looking to the next programme cycle.

Assessment Levels



strong, above average, best practice



satisfactory, respectable



with some weaknesses, still acceptable



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)*

1. Structure and Clarity of Reporting

Yes  
No  
Partial

Assessment Level:

*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 report is mostly well-written, though there are some grammatical, spelling and punctuation errors in the executive summary and main report.

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 has 81 pages, excluding the annexes. Although it is slightly over the maximum page length for a country programme evaluation, it does not detract from the report quality.

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

A logical structure is followed.

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

Annexes do not include the ToRs, which are essential to understanding the evaluation expectations.

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 comprehensive. It serves as a stand-alone section that covers the evaluation purpose and objectives, a brief methodology, findings and main recommendations. It provides sufficient information for the end users of the summary.
6. Is there a clear structure of the executive summary, (i.e. i) Purpose; ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Yes	The section has a clear structure.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The Executive Summary is 5 pages in length and concisely presented.
<b>2. Design and Methodology</b>	Yes No Partial	Assessment Level: <b>Good</b>
<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 this evaluation are decision-makers within UNFPA and the UNFPA Executive Board, as well as government counterparts, donors and interested partners.
<b>2. Is the development and institutional context of the evaluation clearly described and constraints explained?</b>	Yes	The development and institutional context of the evaluation is clearly described. According to the report, no major limitations were encountered.
<b>3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?</b>	Yes	The evaluation report describes how the evaluation team reconstructed the theory of change (intervention logic) in detail and includes a graphic depiction for each one of the four programme components.
<i>To ensure a rigorous design and methodology</i>		
<b>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?</b>	Yes	The evaluation framework is addressed in the text. The annexed matrix includes assumptions, indicators, sources of information and data collection methods for each evaluation question.
5. Are the tools for data collection described and their choice justified?	Yes	The evaluation used the typical data collection tools: document review, informal and semi-structured interviews and focus groups discussions. They are briefly described in the report and are noted in the evaluation framework as being designed around the assumptions and indicators.
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	A comprehensive stakeholder map is included in Annex 3. The consultation process with stakeholders is described. In addition to the map developed by UNFPA country office, the evaluators supplemented the stakeholder consultation by an overview or training courses/sessions. Table 2 shows the number of persons interviewed disaggregated by stakeholder group, location, and gender. A total of 178 stakeholders were interviewed in addition to 13 beneficiary groups. An Evaluation Reference Group was formed and provided feedback on the draft and final evaluation report.
7. Are the methods for analysis clearly described for all types of data?	Partial	There is a section on data validation and analysis, however, it primarily only discusses methods for triangulation of data, which is a validation technique.
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	The reports explicitly acknowledges that no major limitations were encountered.

9. Is the sampling strategy described?	Yes	The report mentions that "sampling was purposive and non-random" accordingly to UNFPA Country Programme Evaluation Handbook to ensure an illustrative sample. The criteria and process for selecting stakeholders and locations covered is clearly outlined under "Section 1.3.2 Site and stakeholder sampling."
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The methodology is sufficient for capturing disaggregated data. Table 2 and 3 show the number of persons interviewed, disaggregated by stakeholder group, location and gender; in addition, the evaluators organized focus group discussions and met end beneficiaries on-site, which are also disaggregated by location and gender.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The evaluation effectively covers cross-cutting issues of gender, equity and vulnerability in the evaluation matrix and sampling framework, which includes focused discussions with youth union leaders and volunteers, Mahalla specialists, teachers, students, homeless mothers, vulnerable households and girls and young women in rehabilitation centres.
<b>3. Reliability of Data</b>		
	Yes No Partial	Assessment Level: <b>Good</b>
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluators triangulated data sources and data collection methods from an expansive group of stakeholders. Pictures are also integrated as evidence, as well as findings from the document review and interviews.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluation makes use of monitoring data as well as data gathered through the evaluation process. There is an extensive and comprehensive list of documents and stakeholders consulted.
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	Meeting with end beneficiaries had certain constraints because of time restrictions and sensitivities around the topics. However, the description of the process for stakeholder mapping describes a process which ensured the reliability of data collected from a diverse group of stakeholders, including beneficiaries.
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Partial	There is some evidence that sensitivity to issues of discrimination and other ethical considerations were applied, including referencing GBV victims as survivors and receiving consent for photographs. While the evaluation consults a wide range of stakeholders, including vulnerable groups, it was mentioned that end beneficiaries weren't fully reached because of time restrictions and sensitivities about topics, though description of a mitigation strategy for this was not explicitly mentioned. Perhaps excluding some groups or discussion of some subjects was a mitigation strategy, but as mentioned, this was not explicit. As such, it is seen that ethical considerations could have been described in more detail.
<b>4. Analysis and Findings</b>		
	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	The evaluators consistently substantiate findings with evidence from diverse data sources.
2. Is the basis for interpretations carefully described?	Yes	The basis for interpretations are carefully described.

3. Is the analysis presented against the evaluation questions?	Yes	The analysis is presented by each criterion, and analyzed according to the assumptions linked to every evaluation question in the evaluation matrix.
4. Is the analysis transparent about the sources and quality of data?	Yes	Data sources are clear and either referenced in the text or in footnotes. Interviews are mentioned throughout the findings, though the general source (e.g. the stakeholder group it was drawn from), was not always mentioned.
<b>5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?</b>	Yes	Causal effects are shown. This is particularly clear under effectiveness where tables are used to show performance assessments against CPD indicators. Tables show 'Indicator, Baseline, Target' compared to Evaluator Assessment of Achievement.
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings present different outcomes for different target groups. For instance Finding 4 mentions that the country office has not (yet) responded to priorities in the areas of youth empowerment and participation and support for a national policy on people with disabilities.
<b>7. Is the analysis presented against contextual factors?</b>	Yes	Findings present a clear analysis of contextual factors, especially political contexts.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The analysis elaborates on cross-cutting issues, including noting where the needs of specific groups are not sufficiently addressed.
<b>5. Conclusions</b>		
	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To assess the validity of conclusions</i>		
<b>1. Do the conclusions flow clearly from the findings?</b>	Yes	The conclusions are clearly drawn from the findings. The evaluators have specified the respective evaluation questions, evaluation criteria and associated recommendation linked to each conclusion.
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 reflect the findings and are organized by two categories: strategic level and programmatic level. A rationale for each conclusion is included which assists in making the linkages clear.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions are clearly based on evidence from the findings and therefore there is no evidence of bias.
<b>6. Recommendations</b>		
	Yes No Partial	Assessment Level: <b>Good</b>
<i>To ensure the usefulness and clarity of recommendations</i>		
<b>1. Do recommendations flow logically from conclusions?</b>	Yes	The conclusions are clearly drawn from the findings. The evaluators have linked each conclusion to a respective evaluation question(s), evaluation criterion and also an associated recommendation.
<b>2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?</b>	Yes	The recommendations are clearly targeted and action-oriented, explicitly defining one or more users (for instance, UNFPA country office or UNFPA regional office). Each recommendation includes reference to operational implications.
3. Do recommendations appear balanced and impartial?	Yes	The recommendations are derived from the findings and conclusions. There is no evidence of bias.

4. Is a timeframe for implementation proposed?	Partial	In the introduction (p 20) is mentioned that recommendations are provided for the next programming cycle in light of UNFPA's strategic goal. It can also be inferred by the prioritization. However, there is no explicit mention of the timeframe in either the conclusions or recommendations.
5. Are the recommendations prioritized?	Yes	Yes, recommendations are classified between 'very high' and 'high priority'.
7. Gender	0 1 2 3 (**)	<p style="text-align: right;">Assessment Level:</p> <p style="text-align: center;"><b>Good</b></p>
<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><b>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)</b> The evaluation does not define a specific objective related to human rights and gender equality. However gender equality and women's empowerment is covered as part of the evaluation scope. (2)</p> <p><b>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)</b> Although gender and/or human rights was not included as a standalone criteria, it is mainstreamed through the indicators in EQ1. (2)</p> <p><b>2 c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3)</b> There is not specifically any evaluation question or sub question regarding the incorporation of GEEW.(1)</p> <p><b>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).</b> Yes, the evaluation teams assess whether sufficient data gathering has taken place to measure progress on human rights and gender equality results. (3)</p>
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		<p><b>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)</b> There is not an explicit discussion on how the methodology was gender responsive. However the total number of evaluation participants were gender disaggregated. (2)</p> <p><b>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)</b> Overall, the methods used appear appropriate - a mixed methods approach was followed, with a total of 178 stakeholders (of which 67 male and 111 female) were interviewed and 13 beneficiary focus groups were held. (3)</p> <p><b>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3)</b> Diverse sources were used and the description of the site and stakeholder sampling selection is clear. There is also evidence of triangulation. (3)</p> <p><b>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)</b> There was representation from a diverse range of stakeholders, including beneficiaries, explicitly the most vulnerable population. However, while vulnerable groups were mentioned in the sampling frame, the specific numbers were not and strategies for overcoming limitations to their accessibility were also not mentioned. (2)</p> <p><b>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3)</b> Attention to ethical issues such as informed consent and confidentiality only mentioned as adherence to the UNEG Ethical Guidelines for Evaluation and the UN Code of Conduct for Evaluations, but additional steps taken to ensure the comfort of beneficiaries participating in particular are not discussed. For instance when some topics were sensitive and end beneficiaries would not meet evaluators. (2)</p>

3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		<p><b>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)</b> The evaluation includes a section on 'development challenges and priorities' that includes an intersectional analysis on how women are affected by normative instruments and policies. (3)</p> <p><b>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)</b> It is not clear in the findings whose voices (from which stakeholders) are being represented. It would have been useful to use quotes or stories to highlight the voices of different groups, particularly participants and rights holders. Especially with the number of stakeholders and focus groups involved. (2)</p> <p><b>3 c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3)</b> The evaluation team includes unanticipated effects of the intervention into the findings. However a more thorough analysis on it would have been useful. (2)</p> <p><b>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)</b> Recommendation 4 addresses how the UNFPA country office should strategize on its advocacy and communications programme by changing attitudes and behaviours in order to increase demand for SRHR and gender equality. There are other recommendations oriented towards human rights considerations as well. (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		

<b>Total scoring points</b>	<b>51</b>	<b>49</b>		
<b>Overall assessment level of evaluation report</b>	<b>Very Good</b>			
	<b>Very good</b> very confident to use	<b>Good</b> confident to use	<b>Fair</b> use with caution	<b>Unsatisfactory</b> 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**

The presentation of findings was quite detailed and well presented and validated. The stakeholder groups met and sampling strategy were also comprehensive, making this a very good report with minimal limitations. The primary limitations related to a more clear incorporation (or explicit mention of) ethical procedures and strategies for mainstreaming GEEV.

**Consideration of significant constraints**

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:  Yes  No

If yes, please explain: