Organizational unit:

Year of report: 2019

Title of evaluation report: Evaluation of the 3rd and 4th UNFPA Country Programme for Kazakhstan

Overall quality of report: | Very Good

Date of assessment: 10 December 2019

Overall comments: This is a well-designed and executed evaluation. The evaluation clearly shows those aspects of interventions under the country programmes that were successful, as well as areas for improvement. It covers two country programmes (the 3rd and 4th) which added to the complexity of the evaluation as data from the 3rd country programme required extensive document review. There are some issues with the level of clarity of the Executive Summary, but the overall report is well structured and effectively presented. Care was taken in citing data sources and in showing causal connections including through detailed documentation of findings within the annexed evaluation matrix. The conclusions are clearly stated and linked to the relevant criteria, evaluation questions, findings and 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 <u>assessment level</u> followed by main <u>comments</u>.* (use 'shading' function to give cells corresponding colour)	
1. 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?	Partial	Readability is somewhat limited by grammatical mistakes, mostly in the Executive Summary, but the report is otherwise well written. Clarity is enhanced by frequent use of tables and bulleting to highlight key points. There is some use of infographics but not all are labelled or have large enough typeface to be easily read (p 21 and 22).	
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 length of the main report (not including the executive summary) is 60 pages.	
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 evaluation is structured accordingly (but does not include lessons learned).	
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 methodological tools are missing from the annex, but all other elements are included, as is a 1.5 page abstract of the evaluation report which appears to be a useful communication tool.	
5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Partial	This would better serve as a stand-alone section if the recommendations were more fully explained. Although the findings section captures the main results, the conclusions take up 3 pages and present more detail than is normally included in an executive summary. The recommendations are, in contrast, more briefly stated and have minimal explanatory text.	
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 structure is clear. The intended audience is not specified in the executive summary, but can reasonably be assumed based on this being a country programme evaluation.	
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Partial	At 6.5 pages, the executive summary is longer than expected.	

2. Design and Methodology	Yes No Partial		Assessment Level:	**Very good**
To ensure that the evaluation is put within its context				
1. Does the evaluation describe the target audience for the evaluation?	Yes		The audience and primary users of the evaluation are broadly identified as the decision-makers within the UNFPA country offices and organization as a whole, government counterparts in the country, the UNFPA Executive Board, and other development partners. It is noted that the UNFPA Regional Office for Eastern Europe and Central Asia and UNFPA Headquarters divisions, branches and offices will also use the evaluation as evidence for a review of programme performance and for decision-making.	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes		There is a thorough description of the context, including particularly the changes that have occurred since UNFPA began working in the country.	
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 evaluators discuss reconstructing the theory of change (ToC) as part of the evaluation design process to better account for advocacy work of the country programme (CP) and to improve measurability of results. The existing CP theory of change is included as the 5th annex of the evaluation terms of reference (although the text in some of the boxes is cut off) and the logic model for each program component is presented in Annex 7.	
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 framework and evaluation questions are expressed clearly and concisely in the text, and more extensively in a full evaluation matrix.	
5. Are the tools for data collection described and their choice justified?	Yes		The concise methodological section describes the tools, which are fairly standard, but also shows why, given the evaluation objectives and practical considerations, they were chosen. For example, the purpose of on-site observations was described as being to "observe operations in real circumstances (church, hospital and crisis center) and/or [to meet participants in programme activities] to talk about UNFPA activities in the city and results achieved to date".	
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 map is described in the methodological section, and there is a more complete presentation in Annex 6. The consultation process in the form of briefings and interactions with the Evaluation Reference Group, including their input into the draft recommendations, is described.	
7. Are the methods for analysis clearly described for all types of data?	Yes		The methods of analysis are described, although fairly concisely. It is noted that the evaluation framework was used to structure the analysis and findings. Stated analysis methods include document analysis, counterfactual analysis, theories of change and contribution analysis, as well as the review of data from a gender perspective. There is brief reference to the types of data used for each method.	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes		The evaluators noted the specific limitations including the timing of the evaluation (during holidays), time constraints regarding visits, access to some sources (especially documentary, but also interviewees since the government changed just before the evaluation), and recall bias. The evaluators discuss mitigation strategies, including tracking down retired civil servants to fill in data gaps.	

9. Is the sampling strategy described?	Yes	The sampling of stakeholders was purposive, and relied on the stakeholder mapping exercise to identify who should be chosen for interviews. When the selected interviewee was not available, snowball sampling was used
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Disaggregated data was available for quantitative data from statistics and from a survey that had been completed with UNFPA support in 2018. The interviews were also disaggregated by sex and institution.
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	Adequate attention appears to have been given to assessing cross-cutting issues. A mixed methods approach was used with qualitative methods including informal and semi-structured interviews and focus group discussions. In addition, the evaluation questions for multiple criteria assess the extent to which gender-specific outcomes have been achieved and needs have been met, including for the most vulnerable and marginalized groups. Representatives from beneficiary groups were interviewed - the numbers were quite low (a total of 8), but did include 2 female youth - and women made up approximately 2/3rds of total interviewees.

3. Reliability of Data	Yes No Partial	Assessment Level:	Very good
To ensure quality of data and robust data collection processes are used in the findings			
1. Did the evaluation triangulate data collected as appropriate?	Yes	Yes. In most of the findings, both documentary reviews and interviews were triangulated.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	Quantitative data was obtained from government sources as well as surveys. The source was clearly indicated in each finding, including, where relevant, how the absence of data - mostly quantitative - made findings difficult. The qualitative data, mostly in the form of interviews, was triangulated with other sources of data, improving reliability, and sources interviewed were familiar with/clearly had knowledge of UNFPA's work and the country programme. Again, where data were not available, this was noted.	
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	Limitations in data sources are noted and included the timing of the evaluation (during holidays) being when many stakeholders were out of reach, access to some government data sources being affected by government restructuring, and possible recall bias in cases where respondents were asked about earlier stages of the country programme. Mitigation strategies are provided.	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	A clear explanation is provided of the ethical considerations undertaken in collecting data and that this was done in accordance with the UN Evaluation Group Code of Conduct and Ethical Guidelines for Evaluations. The evaluators noted how they used ethical considerations when they collected data, for example by conducting interviews outside of government offices to make public officials feel more relaxed, ensuring that venues were accessible for people with disabilities, and informing respondents about confidentiality and anonymity, and acquiring informed consent.	

4. Analysis and Findings	Yes No Partial	Assessment Level:	**Very good**
To ensure sound analysis and credible findings			
1. Are the findings substantiated by evidence?	Yes	In each case, the evidence is clearly described/findings are clearly substantiated by evidence. Of importance is when the absence of data, such as when data for indicators of expected results had not been collected, was noted and explained in the findings.	
2. Is the basis for interpretations carefully described?	Yes	In each case, the evaluators explain the factual basis for their interpretations. Again, when data were not available, the evaluators noted that a credible interpretation could not be made, except about why the data had not be collected by the programme.	
3. Is the analysis presented against the evaluation questions?	Yes	The findings are structured according to the evaluation questions.	
4. Is the analysis transparent about the sources and quality of data?	Yes	The sources of data are always shown, with specific references to sources (whether documents or interviews) in footnotes.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The evaluators were very clear about the causal connections. This was done in great detail in the Annex showing the results matrix, but in the main text, the anticipated causal relationships between UNFPA interventions and expected results (called outputs and outcomes) were a starting point in the analysis. Then the causal connections shown by the data were given. Often, when there was not sufficient data either because the indicator was too general, or data on it had not been collected, this was noted. Unintended outcomes were also noted (in a few of the findings).	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	There was a focus on outcomes by different groups, including especially women, youth, and government officials concerned with population data. (Differentiated) outcomes for each group were shown.	
7. Is the analysis presented against contextual factors?	Yes	The context was provided in the introductory sections, but in explaining the findings, the evaluators were careful to explain the context in which the result did or did not occur.	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	While cross-cutting issues were addressed in all of the finding sections, there were also specific sections on gender equality and human rights.	

5. Conclusions	Yes No Partial	Assessment Level:	Very good

1. Do the conclusions flow clearly from the findings?	Yes	The conclusions are clearly linked to / flow from the findings.
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 summarize a variety of findings and, organized by strategic and then programmatic conclusions, clearly demonstrated an understanding of both the context (and its impact on the programme) and key underlying issues, including when data are not available.
3. Do the conclusions appear to convey the evaluators' unbiased judgment?	Yes	The conclusions are very straightforward and are based on findings, suggesting no bias in the judgments.

6. Recommendations	Yes No Partial	Assessment Level:	Very good

To ensure the usefulness and clarity of recommendations

1. Do recommendations flow logically from conclusions?	Yes	In each case, the recommendations flow logically from the conclusions.
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 stated as a sentence, but are followed by additional details and their implications for implementation including human, financial and technical issues.
3. Do recommendations appear balanced and impartial?	Yes	Flowing from the conclusions and taking into account the institutional and political contexts, they are clearly balanced and impartial.
4. Is a timeframe for implementation proposed?	Yes	They are all to be implemented in the final stages of the current country programme and in 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	They are either high priority (and this includes all of the strategic recommendations) or medium priority (which includes all of the specific programmatic 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? | | 3 | While the evaluation did not explicitly include GEEW in the scope of analysis (i.e. there was not a specific objective to assess human rights and gender equality, nor was there an explicit mention of HRGE being mainstreamed across objectives), given how evaluation criteria were interpreted/applied and the inclusion of both GEEW in evaluations questions, it can be deduced that GEEW was integrated (implicitly) into the evaluation's scope of analysis. While the evaluation did not feature a standalone evaluation criteria on HRGE, it was mainstreamed into other criteria, including primarily under relevance and effectiveness (though, in the annexed evaluation matrix, it is evident that gender was assessed under each criteria), and there were dedicated evaluation questions on gender equality (it was also mainstreamed across other evaluation questions). The evaluation had a particular focus on the extent to which monitoring and results information was available, and found shortcomings in how results were defined and in UNFPA's monitoring system that limited a thorough assessment of progress on human rights and gender equality issues. | |

| 2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques? | | 3 | While the evaluation uses a gender-responsive methodology, the evaluation describes only briefly how GEEW is integrated into the methodology (noting how documents reviewed and interviews held would be sensitive to gender equality). The methodology could have provided more detail about how the evaluation was gender responsive (for example, the evaluation states that "The Evaluation Team ensured gender considerations while conducting on-site observation" (p 11) but there was no description of how this was done). Mixed-methods was used including document review with a gender lens, group and individual interviews, and focus group discussions. The evaluators noted, for example, that "the collected data was analyzed with gender dimension (sic) in mind to reflect a gender perspective so as to exclude gender bias and to enable the team to map all possible consequences for women and men." (p 11). A diverse range of sources were consulted and triangulated, through the evaluations notes that time and resources limited the diversity of stakeholders consulted. There is no indication that ethical standards were not followed (the evaluation specifically notes that it maintained confidentiality). | |

3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?	2	The evaluation includes a background section that includes the normative instruments and policies related to human rights and gender equality. While the findings include an analysis that assesses the extent to which the CP has contributed to gender equality and women empowerment results - such as contributing to the adoption of gender equality and domestic violence laws, and engaging men and boys in issues of gender equality and reproductive rights – when references are made to stakeholder perspectives, these are mostly grouped as "respondents" or the source cited as "key informant interviews. As such, the voices of different social/stakeholder groups are not, on the whole, distinguished (for example, p. 40, the discussion on SRH). An exception to this is featured on p 44 of the report which notes the perspective of an earlier UNFPA survey of vulnerable groups. Unintended or unanticipated effects of the CP on gender equality and/or human rights are not discussed in the report. There is not specific reference to unanticipated effects on HRGE. There are five conclusions related to gender equality, including how the country programme could further gender equality results, and two recommendations that are specific to developing related policies.

(*) 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

Quality assessment criteria (scoring points*)	Very good	Good	Fair	Unsatisfactory
			Assessment Levels (*)	
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	93		7	
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

The sound methodology and careful connection of data have produced a very good evaluation with sound conclusions and recommendations.

Consideration of significant constraints

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

If yes, please explain: