

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

Title of evaluation report: **JOINT EVALUATION OF THE UNFPA UNICEF JOINT PROGRAMME ON THE ABANDONMENT OF FEMALE GENITAL MUTILATION: ACCELERATING CHANGE**

Overall quality of report: **Very Good** Date of assessment: **11 June 2019**

Overall comments: The programme is now in its third phase and the evaluation, taking into account a previous evaluation of the first phase, has presented a clear picture of what works and what needs to be improved in order to accelerate efforts to abandon female genital mutilation (FGM). This is a joint programme and, as such, there has been a careful focus on planning, monitoring and evaluation of the combined effort. The evaluation uses a sound mixture of data collection and analysis techniques. Of particular strength, the evaluators undertook an innovative analysis of "jointness" and had a focus on the longer-term elements that need to be taken into account by the programme. The evaluation also clearly indicates limitations, including a limitation on data on results at the sub-national level, leading to conclusions and recommendations on how this problem can be addressed in the current third phase.

Assessment Levels

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

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

1. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Good
--	----------------------	-------------------	---

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 evaluation is well-written with no errors and considerable use of illustrations.
2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations)	Partial	The report itself is 90 pages including the executive summary, but given its importance and complexity, this is reasonable.
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 clearly structured and includes a clear delineation of 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?	Yes	The annexes are thorough and complete. In addition to the documents expected, the annexes also include notes from meetings of the evaluation reference group as well as detailed presentations on the case studies as country case study evidence tables.

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 presents the main results and is a stand-alone section.
6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Yes	The executive summary follows the desired structure.
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Partial	The executive summary is 6 pages long, but this is not unreasonable for an evaluation of this complexity.

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 is the managers and the steering committee of the Joint Programme, as well as other staff members at UNFPA and UNICEF (at the global, regional and country level), partner countries, and civil society (including non-governmental organizations, feminists and women's rights activists, and gender equality advocates).
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The evaluation covers two phases (and the beginning of a third phase) of a complex programme. The description of the development and institutional is complete and constraints are well-described. Table 2 captures the background of the joint programme and activities at country, regional and global levels while figure 3 details the timeline/evolution of the JP.
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes	Considerable effort was made to define a clear and useable theory of change (a reconstructed theory of change was developed for the Phase I evaluation). Starting with a theory of change in the terms of reference, based on initial work, the evaluation matrix shown in Annex 1.1 reflects the evaluators' definition of the theory of change, Annex 1.6 describes the evolution of the theory and Annex 7 presents a programme logic model.
<i>To ensure a rigorous design and methodology</i>		
4. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?	Yes	The evaluation matrix is described/discussed in the text of the report and is presented in a thorough and detailed form in Annex 1.1.
5. Are the tools for data collection described and their choice justified?	Yes	The data collection tools used are listed and the rationale described carefully in the text and in Annexes. The rationale for the selection of tools, in particular, is described in section 2.3.
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 (by type of stakeholder) is shown in Annex 5. The consultation process – detailing how each stakeholder is expected to be reached – is detailed in the text. The process of consultation on the recommendations in particular was described, having been via a participatory workshop held with Joint Programme colleagues.
7. Are the methods for analysis clearly described for all types of data?	Yes	The evaluators thoroughly presented the approach to data analysis on page 10, including how the United Nations System Wide Action Plan (SWAP) on gender equality criteria within the evaluation were addressed. The evaluators used contribution analysis to assess causal links and reduce uncertainty about the contribution the intervention was making to the observed results. This was described in detail in Annex 17.
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	Section 2.4 describes the limitations and constraints of the evaluation and how these were mitigated.
9. Is the sampling strategy described?	Yes	The sampling strategy for selection of case-studies was described in detail. At the country-level, purposive samples were used and detailed. The sample for the survey was described, and its limitations were indicated, including the uneven geographic representation of responses.
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The evaluators were careful to use disaggregated (by location and gender) data, making an effort to collect data on this basis (i.e. on its ability to be disaggregated).
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	The evaluators were careful to show how the evaluation methodology would allow for the assessment of gender and human rights.
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 were careful to triangulate data, particularly when analyzing support at the local level.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluators used quantitative data from reports and from the survey, noting limitations of the latter, and extensive use of qualitative data from interviews and focus groups. The evaluators noted the extent of reliability in presenting their findings.
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 had a clear section on limitations and what it had done to mitigate them. An example is the survey, where there was a different procedure followed in Nigeria than elsewhere and the evaluation was careful to keep that limitation in mind in the analysis.
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 ensure the anonymity of persons who were interviewed, especially at the local level, and focus groups were informed about ethical issues and how they were addressed. There was an effort to ensure that all stakeholders were represented in interviews and in focus groups.

4. Analysis and Findings	Yes No Partial	Assessment Level:	Very good
<i>To ensure sound analysis and credible findings</i>			
1. Are the findings substantiated by evidence?	Yes	The evaluation reported 48 findings. In each case, evidence substantiating the finding was provided.	
2. Is the basis for interpretations carefully described?	Yes	The basis for interpretations was, in each finding, described carefully.	
3. Is the analysis presented against the evaluation questions?	Yes	The findings were organized by evaluation question which were also connected with the evaluation criteria.	
4. Is the analysis transparent about the sources and quality of data?	Yes	In each case, the sources of data used were shown, often as footnotes, and their quality was described when necessary.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The evaluation made an effort, starting with the evaluation matrix, to show causal connections. In each case, what happened as a result of the intervention was shown, often indicating only a partial outcome. The evaluators also showed, in a number of cases, where unintended outcomes were found or the expected outcome could not be measured as was the case with community declarations.	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	There are a number of target groups, ranging from global to national to sub-national, and the different outcomes for these groups are shown.	
7. Is the analysis presented against contextual factors?	Yes	In each case, the findings are shown against the rather complex contextual factors in which (and with which) the joint programme works.	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The evaluators were careful to examine the cross-cutting issues including, especially, gender and human rights.	
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 ten conclusions reflect (and summarize) the 48 findings. They are structured, like the findings, to the evaluation questions, with the connection between the findings and the conclusions very clear.	
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 consider/compare/analyze the findings holistically in order to demonstrate the larger picture and an understanding of the broader context.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no evidence of bias in the conclusions, which are always balanced by showing both positive and negative elements.	
6. Recommendations			
	Yes No Partial	Assessment Level:	Very good
<i>To ensure the usefulness and clarity of recommendations</i>			
1. Do recommendations flow logically from conclusions?	Yes	Each recommendation notes the conclusions on which it is based. All conclusions are connected with a 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 specify to whom they are directed, by when they should be implemented and, in detail, what would need to be done to implement them.	
3. Do recommendations appear balanced and impartial?	Yes	The recommendations clearly specify their connection with the conclusions and are balanced; there is no evidence to suggest that they are not impartial.	
4. Is a timeframe for implementation proposed?	Yes	In each recommendation, the timeframe for implementation is given ranging from immediately, to the mid-point of Phase III or to the end of Phase III.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The recommendations are all given a priority, based on urgency, expected impact and difficulty.	
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?		<p>The evaluation mainstreams gender and human rights objectives throughout, as reflected in the methodology. In addition, while not explicitly listed as a standalone objective, gender and human rights are assessed vis a vis a stated objective of the evaluation "To assess the extent to which UNFPA and UNICEF, through the Joint Programme, have effectively positioned themselves as key players in contributing to the broader 2030 development agenda, in particular, Goal 5...[...]", as Goal 5 focuses on the achievement of gender equality and women's and girls empowerment.</p> <p>Gender quality is mainstreamed across evaluation criteria (as seen in Table 3 of the report), whereby the definition of the relevance, effectiveness, sustainability and coordination include a gender and human rights dimension, and includes EQs with gender equality integrated (e.g. EQ 2 asks "To what extent has the Joint Programme contributed to supporting governments, communities and the girls and women concerned towards the abandonment of FGM through the establishment of conducive legal and policy environments, support for the provision of FGM health services, and the shifting of social norms?"). In addition, the evaluation report provides considerable information on issues of data collection vis a vis gender and human rights results.</p>
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		<p>In the design of the evaluation, gender-responsiveness is one of the main elements, and data collection and analysis are disaggregated by gender/reflect a gender analysis. The evaluation employs a mixed-methods approach, appropriate to evaluating GEEW considerations; a diverse range of data sources were used, including documents, interviews, focus groups and a survey and data was triangulated in each finding. As noted, the evaluation worked to ensure that disaggregated data was used and that there was gender balance in stakeholders consulted (with more women interviewed). There is no evidence in the evaluation report to suggest that ethical standards were not upheld or that stakeholder groups were not treated with integrity and respect.</p>
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		<p>Yes, the findings, conclusions and recommendations reflect a gender analysis. The background section of the evaluation report clearly delineates the main groups impacted/involved by the issue, including both young women who are threatened by FGM and those in their communities, including health personnel, who may practice FGM. The findings reflect an analysis that explicitly and transparently triangulates the voices of different social role group involved, at global, regional, national and subnational levels, including by disaggregating by age and gender.</p> <p>Further, the findings note that, in some cases (particularly at the community level), interventions may only partially contribute to outcomes and a lack of data at that level makes it difficult to assess effectiveness (i.e. convincing communities to make declarations against FGM do not always lead to a reduction of FGM and an absence of good data at that level makes determining effectiveness difficult). Finally, the recommendations are specific about GEEW issues to be taken into account in Phase III and beyond, including both the need to use a gender equality approach and acquire more data on effectiveness.</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).

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

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

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

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

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

Overall Evaluation Quality Assessment

Quality assessment criteria (scoring points*)	Assessment Levels (*)			
	Very good	Good	Fair	Unsatisfactory
1. Structure and clarity of reporting, including executive summary (7)		7		
2. Design and methodology (13)	13			
3. Reliability of data (11)	11			
4. Analysis and findings (40)	40			
5. Conclusions (11)	11			
6. Recommendations (11)	11			
7. Integration of gender (7)	7			
Total scoring points	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 evaluation of a complex programme was executed with careful attention to all of the standards that should apply.

Consideration of significant constraints

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

If yes, please explain: