

Organizational unit: Year of report:

Title of evaluation report: **Joint Evaluation of the UNFPA-UNICEF - Joint Programme on the Abandonment of Female Genital Mutilation: Accelerating Change Phase III (2018-2021)**

Overall quality of report: **Very good** Date of assessment: **18 November 2021**

Overall comments: This is a strong evaluation of a complex programme and subject. The scope was also complex as it covered three-levels of programming - global, regional and national. A solid mixed-methods approach was used, notable aspects being the range of data collection methods and the multiple types of well-described analysis processes. The analysis extended beyond gender-responsiveness by looking at the gender transformative aspects. A gender framework (Gender Equality Continuum) was used to analyze higher-level results. The findings were clearly presented and led to a solid set of conclusions and recommendations. In the context of the pandemic, evaluators did very well in designing and carrying out a remotely implemented evaluation process that involved extensive data collection from a wide range of stakeholders. Innovation was evident in the use of U-Report to engage youth and in incorporating analysis of social media posts. Main shortcomings are that the evaluation was not disability inclusive, the involvement of the Evaluation Reference Group is not discussed, and the report is lengthier than requested.

Assessment Levels

Very Good (blue box)	strong, above average, best practice	Good (green box)	satisfactory, respectable	Fair (yellow box)	with some weaknesses, still acceptable	Unsatisfactory (red box)	weak, does not meet minimal quality standards
-----------------------------	--------------------------------------	-------------------------	---------------------------	--------------------------	--	---------------------------------	---

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

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 structured in a logical way? 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? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The report is clearly and logically structured. The language is appropriate (although right-based terminology is only used for duty bearers and not rights holders). There are no noticeable errors. There is good use of infographics.
---	-----	---

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)	No	The main volume of the report is 99 pages, including the Executive Summary and 5 photo pages. The ToR notes the report should be 70-80 pages without annexes.
--	----	---

3. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological and data collection tools used (e.g. interview guides; focus group notes, outline of surveys)?	Yes	The annexes are extensive and are provided in a separate volume. All of the required elements are included. The data collection instruments are each presented along with an analysis of results.
--	-----	---

Executive summary

4. Is an executive summary written as a stand-alone section, presenting the i) Purpose; ii) Objectives, scope and brief description of interventions; iii) intended audience; iv) Methodology; v) Main results; Vi) Conclusions and Recommendations?	Partial	The Executive Summary is well constructed but does not include an overview of the programme (or intended results) which would help to provide important context when used on a stand-alone basis.
---	---------	---

5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	At 4 pages, it is concisely presented with a useful infographic of data collection processes.
--	-----	---

2. Design and Methodology	Yes No Partial	Assessment Level: Very good
----------------------------------	----------------------	------------------------------------

To ensure that the evaluation is put within its context

1. Is the development and institutional context of the evaluation clearly described and constraints explained?	Partial	There is a succinct discussion on FGM, relevant international frameworks and conventions set up to address FGM, and the strategic approaches of UNFPA and UNICEF, donor contributions and geographic coverage. There is an overview of the three phases of the JP in the main report with more details provided in the discussion on the ToC in Annex 2. However, the institutional context would be clearer if there was an explanation of how UNFPA and UNICEF work together, and the ways in which the JP interacts with governments and with other institutions.
---	---------	--

2. Does the evaluation report discuss and assess the intervention logic and/or theory of change?	Yes	The evaluators note using primarily a theory-based qualitative approach to assess the programme's change model. The ToC, being one of the foundations of the evaluation, is captured in figure 4. The assumptions and causal connections underlying the ToC are assessed throughout in the findings.
<i>To ensure a rigorous design and methodology</i>		
3. 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 evaluators provide a clear framework in the main document and a complete evaluation matrix is annexed.
4. Are the tools for data collection described and their choice justified?	Yes	Tools are described and justified. Their use in assessing different components/criteria is well depicted diagrammatically in figure 4. Further information is provided in a supplementary discussion on methodology in annex 3.
5. 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	Although a stakeholder map is not apparent (one that sets out the various types of actors, their roles and relationships, and from which a sample would be drawn), the annexed list of evaluation participants shows that a diverse range of stakeholders were consulted during the evaluation. It is understandable that a comprehensive stakeholder mapping process would be beyond the scope of a global evaluation, however a high-level overview - that for example included a description of how a) the range of duty bearers such as INGOs, civil society, and academic institutions are involved, and b) how UNFPA and UNICEF work with other duty bearers and rightsholders - would have been useful to help understand the extent to which the sample was representative or reflective of all these different stakeholder groups. A stakeholder map would also have been useful for capturing the extent to which highly marginalized groups such as PWD are considered in the programme and the evaluation. The report could also be more explicit about stakeholder engagement in data validation and the development of recommendations. Although it is clear that an Evaluation Reference Group was involved (from the Acknowledgements and ToR), their role is not discussed in the report.
6. Are the methods for analysis clearly described for all types of data?	Yes	It is clear the evaluation relied primarily on contribution analysis and this is explained in the main report. The annexes include a solid explanation of the other types used - qualitative content analysis, gender results effectiveness scale analysis, descriptive statistics, process tracing, realist evaluation, and social media analysis - and for what purpose and level of synthesis (global, HQ, regional and thematic). The evaluation is exemplary in using and describing multiple types of analysis processes.
7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?	Yes	Limitations and mitigation efforts are clear.
8. Is the sampling strategy described?	Yes	The sampling approach (purposive) is described for selecting countries for 'deep dives'. The criteria for selecting interviewees for defining the sample for surveys of staff and implementing partners is also covered. More details for each selected country are provided in an annexed table.
9. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Data is noted as being disaggregated by different target populations groups as part of the realist evaluation approach used to look at differential effects.
10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)?	Partial	The design was appropriate for addressing cross-cutting issues. U-report was used to collect data from JP participants. However, there did not appear to be attention given to PWD.
3. Reliability of Data	Yes No Partial	Assessment Level: Very good
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	The use of different data for triangulation is well explained. Triangulation is enabled by the multiple types of data collection methods - document review, KIIs, country case studies, indepth cases, 2 surveys, U-reports and social media analysis - and extensive amount of data obtained from these.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	Both types are used and reliability is discussed, including in Limitations for document-based quantitative data collected at the national level and for data collected via mobile devices.

3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	Ethical considerations are discussed in a specific subsection under Data Collection Methods and appear appropriate for the remote methods used. Reference is made to the UNEG Ethical Guidelines and the Code of Conduct for Evaluation in the UN System and an explanation given of how the following issues were addressed - involvement of minors; rights to self-determination, fair representation, protection and redress; and confidentiality.
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	Sources of evidence are provided, including through extensive footnotes. Icons are also used to denote the types of data sources drawn on for each question.
2. Is the basis for interpretations carefully described?	Yes	This is well done with evaluators bringing in examples from different countries to back up key points from other data sources.
3. Is the analysis presented against the evaluation questions?	Yes	The findings are organized by evaluation question and assumptions. A brief summary of findings is provided under each question and brings additional clarity to this section.
4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	Links are frequently made between JP outputs and outcomes using indicators from the results framework (for example, the discussion on p 59 - Output 1.2.3 that focuses on increased engagement of men and boys on changing social and gender norms which includes baseline, as well as planned and achieved targets). Annex 5 shows the results framework with performance by year against baseline data where available. Although unintended outcomes were not explored in the findings, in Volume 3 Thematic Notes there is a discussion in the Problem Statement on the unintended results of the health-first approach that drove the medicalization of FGM.
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	This is adequately covered given the global nature of the programme. There are examples of different output- and outcome-level results shown for different countries, regions and contexts. There is also some discussion of the shortcomings of interventions in reaching those in more remote communities.
6. Is the analysis presented against contextual factors?	Yes	The political, legal and social contexts are consistently presented, through country-level examples. Context is also provided in a discussion on cultural relativism versus radical universalism.
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Partial	Cross-cutting issues are a focus, particularly vulnerability and HRGE. However, disability is not addressed beyond stating that FGM can result in serious injury, disability or death.
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 well structured and clearly reflective of the findings. The relevant evaluation questions that each conclusion is based upon are shown.
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated and reflect as appropriate cross-cutting issues such as equality and vulnerability, disability inclusion, gender equality and human rights?	Yes	The conclusion are sufficiently high-level and forward looking. HRGE issues are fore fronted, as would be expected from the nature of the JP. There is a specific conclusion on gender responsiveness and gender-transformative approaches.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no indication 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	Recommendations are clearly formulated and the conclusions on which they are based are shown.
2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	It is clear what bodies the recommendations are directed towards. There is a useful level of detail on how each can be operationalized. The first recommendation addresses the need for a clear resource mobilization strategy.
3. Do recommendations appear balanced and impartial and address, as relevant, key cross cutting issues such as equity and vulnerability, disability-inclusion, gender equality and human rights?	Yes	They appear balanced and impartial. Cross-cutting issues taken up in the report are well covered.
4. Are the recommendations prioritized?	Yes	Each of the eight main recommendations is prioritized in terms of urgency and impact. A scale of high, medium and low is used for both categories. The formatting of this section is very effective for the way in which it presents and highlights the main recommendation, its level of urgency and impact, to whom targeted and the relevant conclusions - it is an example of good practice in this regard.
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?		<p>a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3) Assessing gender responsiveness is highlighted in objectives. GEEW is also part of thematic scope: "the extent to which the Joint Programme has integrated a gender-responsive and/ or a gender-transformative approach". = 3</p> <p>b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3) Gender was mainstreamed. = 3</p> <p>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) Yes = 3</p> <p>d. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?(Score: 0-3) UNICEF's Gender Equality Continuum was used to frame analysis of the gender impact of programme components. Findings explore the extent progress has and should be measured, including the new ACT framework = 3</p>
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		<p>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex? (Score: 0-3) Although it is clear how data was analyzed using a gender lens, the specific steps taken to make sure the evaluation process was gender responsive is less clear. It would have been helpful to have a subsection that addressed this. Evaluation participants are disaggregated by gender and stakeholder group. = 2</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3) An appropriate mixed-methods approach was used = 3</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) This was well done. = 3</p> <p>d. Do the evaluation methods and sampling frame address the diversity of</p>

3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?	<p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality? (Score: 0-3) This is well covered in the context section. = 3</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable? (Score: 0-3) There is a solid section on the extent to which the programme is gender responsive and transformative. Different experiences and voices are brought out to some extent in the main report and further in Volume 3 Thematic Notes. = 3</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) As discussed above there is not a discussion on unintended effects of the JP but there is of past approaches to address FGM. = 2</p> <p>d. Does the evaluation report provide specific recommendations addressing</p>
---	---

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted.
(**) Scoring uses a four point scale (0-3).

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		