

**Organizational unit:** \_\_\_\_\_ **Year of report:** 2018

**Title of evaluation report:** Evaluation of UNFPA support to the prevention of, response to and elimination of gender-based violence and harmful practices (2012-2017)

**Overall quality of report:** **Very Good** **Date of assessment:** 1 February 2019

**Overall comments:** The GBV thematic evaluation report is of high quality and can be used as a model for other evaluation reports in terms of detailed coverage of the criteria from the EQA grid. This includes a detailed description of the intended users, the evaluation methodology, the integration of human rights and gender equality, and stakeholder involvement.

The report also demonstrates innovative approaches to design of the evaluation: notably on the reconstruction of the theory of change (clear structure and formulations), data collection (utilisation of several different methods, in particular the case studies which were used to collect the data in a way that could show the causal connections), design of the recommendations (included useful components such as urgency, impact, difficulty, intended users - "directed to", and operational implications - "achieve this by").

The report proposes practical means of dealing with the increasing incorporation of GBV work in UNFPA programming, at global, regional and country levels as well as in coordination with other organizations. It has a clear focus on humanitarian activities, touching on strategic issues such as dealing with the reduction in core resources to the organization by finding ways to utilize non-core resources more effectively.

The only weaknesses of the evaluation report are the length and time specification for some recommendations.

**Assessment Levels**

- Very Good** (blue): strong, above average, best practice
- Good** (green): satisfactory, respectable
- Fair** (yellow): with some weaknesses, still acceptable
- Unsatisfactory** (red): 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)*

<b>1. Structure and Clarity of Reporting</b>	Yes No Partial	Assessment Level: <b>Good</b>
----------------------------------------------	----------------------	-------------------------------

*To ensure the report is comprehensive and user-friendly*

<b>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?</b>	Yes	The report is easy to read and understand. It is written in an accessible language appropriate for the managers of the UNFPA.
<b>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)</b>	Partial	The report is 105 pages long, although there are at least five pages of pictures. It is a complex evaluation and it could be argued that a longer length than the 80 pages for a thematic evaluation can be justified.
<b>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)?</b>	Yes	The report structure is consistent with UNFPA standards. The report could be a model in terms of detailed coverage of the criteria from the EQA grid.
<b>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?</b>	Yes	The annexes are comprehensive and contain all of the expected material. In addition to the expected minimum annexes, the GBV report also includes other useful information such as Annex 2 (the minutes of the Evaluation Reference Group).

*Executive summary, as well as in the text and in Annex 6 on the evaluation methodological approach.*

<b>5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?</b>	Yes	The summary is written as an excellent stand-alone summary. An innovation is the graphic presentation of the data sources for the evaluation, which provides a basis for appraising the methodology.
<b>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)?</b>	Yes	The structure matches the requirements of the criteria.
<b>7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?</b>	Yes	The executive summary is exactly five pages long.

<b>2. Design and Methodology</b>	Yes No Partial	Assessment Level: <b>Fair</b>
----------------------------------	----------------------	-------------------------------

*To ensure that the evaluation is put within its context*

1. Does the evaluation describe the target audience for the evaluation?	Yes	The target audience is well-described. Table 2 shows both the intended users and the extent to which they are expected to be able to use it by broad category (accountability, decision-making, learning and evaluation approach. The main users are UNFPA global, but, especially for learning, also regional and country offices and donors.
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The development and institutional context of the evaluation clearly is described and constraints are explained (Chapter 2 "GLOBAL CONTEXT AND THE UNFPA RESPONSE). The description is very clear. The report provides extensive analysis of the concept of gender-based violence and the difficulties of translating this into what can be measured in a UNFPA context.
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes	In both an annex and in the main report, the reconstruction is well-described. The evaluation notes that there was a process of revising the theory of change as the evaluation proceeded to take into account what was being learned. The reconstructed theory of change is found in Figure 4. It should be noted that the terminology used is somewhat inconsistent with RBM norms in that what are called outcomes are really specific objectives and what are called output are really outcomes that are expected to be made happen by the modes of intervention.
<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 framework is clearly described in both the text and in more detail in annex 5. The questions, assumptions, indicators, data sources and methods are well-described.
5. Are the tools for data collection described and their choice justified?	Yes	The tools for data collection are described and their choice justified in the section 1.3.4 Methods for data collection, specifically in the Table 5, including document review, group interviews, semi-structured interviews, observation and an internet survey. The numbers of each are shown in a figure in the executive summary. The approach was innovative because they used several different methods (especially the case studies) to collect the data in a way that could show the causal connections.
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	Table 4 describes the stakeholders by type, gender and number. The consultants do a detailed stakeholder analysis presented in the table 4 "Identification of stakeholders using Critical Systems Heuristics (CSH)" and the table 5 "Distribution of interviewed people by stakeholder type and by level of analysis." The consultation methods including, especially, the Evaluation Reference Group, are well described. For instance, the evaluators clarify that "... the following areas for recommendations have been developed and validated with the evaluation reference group" (p. XIV).
7. Are the methods for analysis clearly described for all types of data?	Yes	The methods for analysis are clearly described in the section 1.3.5 "Methods for data analysis." Table 6 shows the methods for analysis by type of data.
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	In a number of places, including both conceptual and methodological issues, limitations are well-described. How they have been addressed was also shown.
9. Is the sampling strategy described?	Partial	There were two areas involving sampling. The first was the selection of countries for case studies. For this a purposive sample, using indicators, was applied. The second was the Electronic survey of UNFPA staff, United Nations and Implementation Partners. There was only a 21% return on country offices and 66 percent of regional offices, so the evaluators note that the sample is not representative. A sampling strategy was not described for the survey nor were the reasons of low response rate (p. 10).
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The methodology enables the collection and analysis of disaggregated data, for instance, the consultants explain that "...a systems-based approach (critical system heuristics) was used to map the key categories of stakeholders in UNFPA's interventions, disaggregated by human rights roles and an intersectional gender analysis where relevant (p. 146). As a result, in data collected through interviews as well as in the electronic survey, data can be disaggregated for analysis.

<p>11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?</p>	<p>Yes</p>	<p>Both the design and methodology have been developed to deal with gender equality and human rights issues, for instance, the consultants explain that "...a systems-based approach (critical system heuristics) was used to map the key categories of stakeholders in UNFPA's interventions, disaggregated by human rights roles and an intersectional gender analysis where relevant" (p. 146).</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------	------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>3. Reliability of Data</b></p>	<p>Yes No Partial</p>	<p>Assessment Level:</p>	<p><b>Very good</b></p>
--------------------------------------	-------------------------------	--------------------------	-------------------------

To ensure quality of data and robust data collection processes

<p><b>1. Did the evaluation triangulate data collected as appropriate?</b></p>	<p>Yes</p>	<p>Triangulation was used throughout the analysis. The evaluators paid careful attention to the data triangulation: "cross-comparing the information obtained across various data collection methods (e.g. comparing data from interviews with data from desk review/survey) and within a method from different sources (e.g. compare results obtained through interviews with government staff with those of rights holders)" (p. 6). They also explain in the text the limitations of triangulation, for instance, "...existing gaps in structures for consulting with constituencies to triangulate programming priorities... These gaps may partly explain the somewhat mixed results from the global survey and desk review: population-based data is often lagging the actual situation on the ground by the time it is made available, and so relying solely on it for programming can lead to poor decisions" (p. 60).</p>
--------------------------------------------------------------------------------	------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?</b></p>	<p>Yes</p>	<p>There is no specific discussion on reliability of data in the methodology except some cases, for instance, on the results of the global survey. But the evaluators explain how they assess the reliability of the data in the analysis of the findings, for instance: "The contribution of UNFPA to global public goods in terms of knowledge production was largely triangulated by the evaluation case studies. In Central African Republic, the lack of reliable data has undermined planning in any sector..." (p. 82). There is a Finding which discusses the issues of Data and Evidence "...but reliable and timely data on service quality or well-being outcomes is not often available" (p. 75).</p>
--------------------------------------------------------------------------------------------------------------------------	------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>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?</b></p>	<p>Yes</p>	<p>The report highlights possible limitations (bias, data gaps etc.) in primary and secondary data sources, including mitigation measures (section "1.3.7 Limitations and mitigating actions"). The minor issue is that some methodological limitations are not explained in detail. The main limitations included:</p> <ol style="list-style-type: none"> <li>1) "no assessment of attribution to impacts using statistical techniques;"</li> <li>2) "the reductionist nature of all theory-based approaches that cannot be fully overcome, but can be mitigated through full transparency about evaluative reasoning and judgements;"</li> <li>3) "constrained involvement of large numbers of rights holders and marginalized people in the commissioning and design of the evaluation, or as data collectors and interpreters;"</li> <li>4) "and the potential for bias in the data collection, which was mitigated through triangulating data, critical analysis by the evaluation team, and validation by the evaluation reference group, national reference groups and participants of summit workshops" (p. 12).</li> </ol>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?</b></p>	<p>Yes</p>	<p>There is evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations: the data collection process was designed in the way to ensure that the rights of respondents are protected. This was achieved by ensuring confidentiality, informed verbal consent, same-sex facilitators, comparable power and status, use of translators to local languages. Another evidence is that "the evaluators did not work directly with any stakeholder below 15 years of age. The perspective of children was gained through interviews with representatives" (p. 11).</p>
--------------------------------------------------------------------------------------------------------------------------------------------------	------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>4. Analysis and Findings</b></p>	<p>Yes No Partial</p>	<p>Assessment Level:</p>	<p><b>Very good</b></p>
----------------------------------------	-------------------------------	--------------------------	-------------------------

To ensure sound analysis and credible findings

1. Are the findings substantiated by evidence?	Yes	Each of the 48 findings is clearly based on the evidence that was collected (country case findings, global and regional level interviews, and others). The findings clearly reference the sources of information.
2. Is the basis for interpretations carefully described?	Yes	The basis for interpretation is shown in the explanation of the findings, for instance: "country case findings made clear that UNFPA technical expertise ...provide the qualifications and critical entry points to efficiently address both policy on and the experience of GBV and harmful practices" (p. 30). Another example: "...GBV quantitative data is perpetually difficult – and ethically complicated – to collect and therefore it is always a struggle to present GBV within Humanitarian Needs Overview processes to the same level as other clusters. Therefore, the qualitative aspect of 'Voices' highlights both GBV needs and the importance of mainstreaming GBV mitigation and prevention within other sectors" (p. 28).
3. Is the analysis presented against the evaluation questions?	Yes	All of the findings were grouped according to the evaluation questions.
4. Is the analysis transparent about the sources and quality of data?	Yes	The main sources of data are group interviews and semi-structured interviews, desk review including text coding of documented sources, and site visits. The evaluators discuss data sources and quality of data in the methodology and the analysis. As an illustrative example, the Finding 37 discusses DATA AND EVIDENCE, for instance: "... secondary monitoring data was generally found to be insufficient to capture outcome level changes..." (p. 75).
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	In each finding, there was a clear effort to show causal connections. More than in many other evaluations, unintended outcomes were noted. For example, a finding on programs directed to boys and men included some unanticipated results.
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation report provides a disaggregated analysis that differentiates outcomes for different target groups, in particular, women and girls (section 3.3.2 "Contribution to outcomes for women and girls"), including women refugees (p. 74).
7. Is the analysis presented against contextual factors?	Yes	Contextual factors were utilized in the analysis, particularly those that were discussed in case studies, which illustrated how the national context can strongly affect results by showing how the details of the work process led to different results..
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The analysis elaborates effectively on cross-cutting issues. For instance, the evaluators explain that "...for example, UNFPA Guatemala addresses, at a fundamental level, the unequal power relations that regulate gender social order and enable violence to become 'natural' and acceptable. This addresses the most vulnerable and marginalized people, particularly indigenous women and adolescents..." (p. 27).

<b>5. Conclusions</b>	Yes No Partial	Assessment Level:	<b>Very good</b>
<i>To assess the validity of conclusions</i>			
1. Do the conclusions flow clearly from the findings?	Yes	Each conclusion is grounded in key findings that respond to the 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 go beyond the findings, for instance, the evaluators discuss that "...the evaluation found that some lacked gender-awareness or did not apply the same level of human rights principles (participation, inclusion, transparency, universality, accountability) ... This is of particular concern in regard to work on social and gender norms, and institutional transformation. While the configurational case analysis found relatively linear explanations of causality for policy and humanitarian outcomes, achieving social and institutional outcomes produced very complex solutions" (p. 92).	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions are clearly drawn from the findings and do not reflect any 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 11 recommendations are structured by area of work (development, humanitarian, corporate); the basis for the recommendations generally flow from the conclusions.	
<b>2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?</b>	Yes	The recommendations are clear in terms of urgency, impact, difficulty, and intended users. There is information in each case about what needs to be done to implement the recommendation.	
<b>3. Do recommendations appear balanced and impartial?</b>	Yes	The recommendations are impartial and balanced.	
<b>4. Is a timeframe for implementation proposed?</b>	Partial	In most cases, the time horizon is specified, usually within the current strategy. In a few cases, the time horizon for achievement is not clear.	
<b>5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?</b>	Yes	The recommendations are clearly prioritized and presented in view of their urgency, impact and difficulty, and in each criteria are rated high, medium or low to facilitate an appropriate response to the recommendations.	

Fine.

<b>7. Gender</b>	0 1 2 3	Assessment Level:	<b>Very good</b>
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i>			
<b>1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?</b>		GEEW is integrated in the evaluation scope of analysis through in terms of "gender-based violence (GBV) and three harmful practices – female genital mutilation, child marriage, and son preference" (p. x). There is a Table 3 "Integration of human rights and gender equality" which addresses GEEW issues explicitly. Sufficient information was collected on GEEW indicators during the evaluation. 3 We can find gender-responsive indicators in the evaluation matrix, for instance: "Alignment of UNFPA's work (in both process and substance) with the guidance of international human rights conventions, instruments and reports; and National Plans of Action, and national gender-equality strategies" (p. 131). Findings include information on this specific indicator (section 3.1.1.) and others.	
<b>2. Do evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved?</b>		Evaluation criteria includes specific additional criteria which addresses how GBV is considered in programme design, planning and implementation (p. 143): 1. Coverage of population groups facing GBV and harmful practices wherever they are. 2. Connectedness between short-term emergency response and longer-term prevention of GBV and harmful practices. 3 Evaluation questions 1, 4, 5, and 7 incorporate GEEW dimensions (pp. 142 - 143), for instance, EQ 7: "To what extent have UNFPA's interventions and approaches contributed ... to strengthening the sustainability of international, regional, national and local efforts to prevent and eradicate GBV ... ?" (p. 143).	
<b>3. Have gender-responsive evaluation methodology, methods and tools, and data analysis techniques been selected?</b>		Annex 6 "Evaluation methodological approach" describes design principles applied in the evaluation that are relevant to the gender-responsive methodology, for instance: a mixed-methods design, methods of data collection and analysis that apply human rights principles (participation, non-discrimination, accountability), and feminist evaluation (Collaborative Outcomes Reporting Technique – CORT -, contribution analysis). It also 3 included a table (Table 1) that outlined UN SWAP criteria and how it was integrated into the evaluation as well as limitations to approach. The sampling for the selection of persons interviewed was gender responsive as it is evident from the Table 5 "Distribution of interviewed people by stakeholder type and by level of analysis" p. 146). Minor issue is that the annex 3 "List of people interviewed" does not disaggregate the participants in terms of gender.	

4. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		<p>The evaluators triangulate the voices of different social groups, and disaggregate quantitative data throughout the analysis, for instance: “women are up to 10 times more likely to experience sexual violence” (p. 27). The table 7 “Perceptions of the reach of UNFPA GBV and harmful practices programming to selected people...” divide groups into adolescent and young girls, poor women and girls, and older women.</p> <p>Conclusions reflect a gender analysis, for instance, the conclusion 1 states that “The evaluation found multiple lines and levels of evidence that indicate UNFPA has made direct contributions to national capacity and institutions across the health sector and national gender equality mechanisms... (p. 88).</p> <p>Recommendations reflect a gender analysis, for instance, the recommendation 6 reveals that “the evaluation found that UNFPA work on primary prevention (including through communication, adolescents and youth empowerment, engagement with non-traditional partners, and gender transformative programming with men and boys), has a more uneven library of evidence to draw on...” (p. 99).</p>
-------------------------------------------------------------------------------------------	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

(\*) This assessment criteria is fully based on the UN-SWAP Scoring Tool, see Annex 7. 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).

#### 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			
<b>Total scoring points</b>	<b>69</b>	<b>18</b>	<b>13</b>	
<b>Overall assessment level of evaluation report</b>	<b>Very Good</b>			
	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 methodology and its execution has led to an almost model corporate evaluation.

**Consideration of significant constraints**

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes  No

If yes, please explain: