

Organizational unit: **Year of report:**

Title of evaluation report:

Overall quality of report: **Date of assessment:**

Overall comments: This is an formative thematic evaluation to assess UNFPA support to gender equality and women's empowerment (GEVE) across development and humanitarian settings, from 2012 to 2020. The purpose of the evaluation is to provide learning inputs to inform the strategic positioning and inform the design of the next UNFPA Strategic Plan. The evaluation objectives are achieved through a theory-driven mixed-methods approach. Due to COVID-19 restrictions, the evaluation is based on 2 in-country case studies, 2 virtual case studies, plus 10 desk-based case studies and 1 regional case study. Findings are comprehensive and based on evidence from different data sources and ensuring that the voices of rights holders, including the most vulnerable and with disabilities, when possible, were reflected. Conclusions and recommendations derive appropriately from findings and are useful for the intended purposes, although the language, in some cases, could be more accessible. The entire evaluation adequately integrates human rights and particularly gender equality principles given its objective.

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

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

To ensure the report is comprehensive and user-friendly

I. 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 written and highly accessible and readable. It is logically set out, and well balanced in content and coverage. The report follows a logical structure and effectively uses headings and text boxes to guide the reader. There are minimal spelling, grammar and punctuation errors and some places where figures are not accurately labelled (e.g. Volume 2, Annexes, p. 78 or the entry of one person multiple times in the interview list), though they do not detract from the clarity of reporting.
---	-----	--

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 evaluation is 83 pages, and therefore extends slightly beyond the maximum pages desired for a thematic evaluation.
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 comprehensive and contain all necessary elements.
<i>Executive summary</i>		
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 includes the evaluation purpose, objectives, scope, methodology as well as main findings, conclusions and recommendations. The intended audience is inferred through its aim to inform the design of the next UNFPA Strategic Plan, though it is not made explicit.
5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The executive summary is concise and within the 5-page limit.
2. Design and Methodology		
	Yes No Partial	Assessment Level: Very good
<i>To ensure that the evaluation is put within its context</i>		
1. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	The context of the evaluation is clearly described and constraints explained, most notably the implications of COVID-19. The context describes succinctly through narrative and graphics the global normative and operational context, UNFPA global COVID-19 response priorities, the strategic frameworks shaping UNFPA support over the period under evaluation, and dedicated financial investments to gender outcomes and mainstreamed activities.
2. Does the evaluation report discuss and assess the intervention logic and/or theory of change?	Yes	The evaluation approach was described as 'theory-driven' and based on a reconstructed theory of change which is provided within Annex 3, and draws significantly from the UN-SWAP Theory of Change for Reporting on System-wide Strategic Gender-related Results and several iterations of UNFPA's Strategic Plans.
<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 final evaluation matrix, including evaluation questions and assumptions, indicators, data sources and data collection methods, is included in Annex 5. The assumptions being tested are integrated throughout, and guide, the findings section.
4. Are the tools for data collection described and their choice justified?	Yes	Data collection tools are extensive and clearly described and justified within the report and Annex 6, including individual and group interviews, focus group discussions, survey and document review.

<p>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)?</p>	<p>Yes</p>	<p>The stakeholder map is referenced in Annex 6, which provides the detailed methodological approach, though the actual map is not provided, only the disaggregation of actual interviewees undertaken by stakeholder type and sex. There were debriefs in each country undertaking a case study and separate sessions with the Evaluation Reference Group (ERG) which were mentioned to focus on data, results and conclusions. Recommendations were presented to the Evaluation Reference Group for validation. According to the report, rights holders, duty bearers, and rights defenders were included in stakeholder mapping and analysis, with particular emphasis on rights holders who constitute the second largest stakeholder group consulted for in-country case studies. In addition, while it is not made explicit whether persons with disability were also included in the stakeholder mapping, rights holders with disabilities and representative organizations of marginalized groups were consulted during data collection, and thus this can be inferred.</p>
<p>6. Are the methods for analysis clearly described for all types of data?</p>	<p>Yes</p>	<p>Methods and processes for analysis are clearly described in the report and supporting annexes. Contribution analysis was used, though with challenges due to the nature of a thematic evaluation with a broad and "amorphous" scope of activities and outcomes. Other methods include content analysis, comparative analysis, and quantitative analysis to develop descriptive statistics.</p>
<p>7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?</p>	<p>Yes</p>	<p>Limitations to the evaluation sampling approach, data collection tools, and analysis methods and analytical tools are clearly noted and appropriate mitigation strategies applied. It is clear that extensive efforts were undertaken to collect data from a broad range of stakeholders, adapting to the context presented by the pandemic, and to systematically analyse data and draw plausible conclusions.</p>
<p>8. Is the sampling strategy described?</p>	<p>Yes</p>	<p>There is a clear description of the sampling approach for selecting case studies provided in Annex 6. The sampling strategy is described for each of the methodological approaches. Case study countries and desk-based case studies were selected through a purposive sampling strategy; while the regional case study used a purposeful sampling approach based on five criteria. The selection criteria is also described in detail. The in-country case study sample was narrowed to 5 in consultation with the Evaluation Office, along with the 10 desk-based studies and the LACRO regional case study. The seven specific selection criteria are detailed in a text box, and limitations of this purposed approach described, notably COVID-19 and the resulting lack of representation from a country in Asia (at least in case studies, as stakeholders in Asia and the Pacific country offices were consulted through the survey).</p>
<p>9. Does the methodology enable the collection and analysis of disaggregated data?</p>	<p>Yes</p>	<p>The chosen methodology enables the collection and analysis of disaggregated data. Dedoose, the qualitative analysis software, was used to define and organize interviews for analysis of disaggregated data by "social identifiers" and sex; the survey tool also allowed for collection of disaggregated data.</p>
<p>10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)?</p>	<p>Yes</p>	<p>Table 6.1 in Annex 6 clearly define the ways in which the selected data collection and processing tools supported the assessment of human rights, gender equality and women's empowerment. Evaluation questions and indicators are explicit and adequate; they capture the extent to which the methodology was gender responsive. Nevertheless, the limitations of the COVID-19 pandemic limited the evaluation's ability to reach most vulnerable for input, as expressed in Table 4.1 in Annexes.</p>

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	Multiple sources (e.g. interviews, documents) and types (i.e. quantitative and qualitative) of data are drawn on in order to support the overarching findings.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluation clearly identifies sources and the limitations of some, such as the survey data, which had a low response rate, or the objectiveness and utility of the analysis drawn from the Gender Results Effectiveness Scale. The case studies are frequently cited, which themselves provide more supportive evidence for the main evaluation report findings. In particular, case study reports further draw on consultations with rightsholders.	
3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	There is substantial evidence in the report and supporting annexes (e.g. the detailed data collection approach and tools/protocols) which demonstrates attention was paid to maintaining sensitivity to discrimination. The ethical considerations prescribed by UNEG are referenced within section 3.4 and further described in Table 6.1 of Annex 6. The survey tool is reasonably short, and therefore demonstrates that the evaluation team considered the value placed on respondents' time, especially during a global pandemic when people were already stretched in terms of capacity. The script for obtaining consent is included within the interview protocol (included as Annex 13, 'Interview Logbook for Semi-structured Interview').	
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	All findings are clearly substantiated by evidence.	
2. Is the basis for interpretations carefully described?	Yes	The basis for interpretation of data is carefully described in the text and further justified within Annex 6.	
3. Is the analysis presented against the evaluation questions?	Yes	The analysis is presented against the evaluation questions and key assumptions tested.	

4. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	The findings very clearly unpack UNFPA's vision, mandate, policies and approaches, and their inherent complexities when seeking to have country-level impact in countries with diverse local contexts and norms.
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The analysis seeks to assess the ways in which UNFPA has considered and responded to the needs of different target groups in its policies and programming across regions, including the needs of indigenous populations, older women, adolescent girls, and persons with disabilities. Survey data is also disaggregated by gender and role, providing interesting insights; for example, fewer female staff than male staff agree that UNFPA's gender architecture is what is needed to meet GEWE needs of UNFPA staff.
6. Is the analysis presented against contextual factors?	Yes	The analysis is presented against contextual factors, most notably the affects of COVID-19 on the populations served by UNFPA programming and on the changing operational environment for UNFPA's work / strategy.
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Yes	The analysis is focused on cross-cutting issues such as equity and vulnerability, disability inclusion, gender quality and human rights.
5. Conclusions	Yes No Partial	<div style="display: flex; justify-content: space-between; align-items: center;"> Assessment Level: <div style="background-color: #00aaff; color: white; padding: 5px 20px; border-radius: 5px;">Very good</div> </div>
<i>To assess the validity of conclusions</i>		
1. Do the conclusions flow clearly from the findings?	Yes	The conclusions clearly flow from the findings. Each conclusion links explicitly the evaluation question(s) number(s) for its base.
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 8 conclusions bring together the 39 key findings to provide a clear summary response to the evaluation questions. Conclusions synthesize properly the underlying issues of the thematic assessment. Not only does it take into consideration gender equality and women's empowerment, but also vulnerabilities of girls and women with disabilities and in marginalized or vulnerable situations.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There appears to be no biased judgment within the conclusions, which are clearly derived from the extensive, and well-validated, findings.

6. Recommendations	Yes No Partial	Assessment Level:	Good
<i>To ensure the usefulness and clarity of recommendations</i>			
1. Do recommendations flow logically from conclusions?	Yes	The recommendations logically flow from the conclusions.	
2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Partial	The recommendations are clearly targeted towards the intended users, with a list of action items that would help to achieve the overarching recommendation. However, some action items are broad, or would be highly complex to implement as they sometimes involve external actors outside of UNFPA's direct sphere of influence. To increase usability, some operational suggestions could be written in more accessible language and may benefit from further explanation, i.e. "Seeking to clarify how UN Women (and likely also UNICEF) can strengthen the GEWE effects of the UNFPA mandate and, with input from this strategic analysis of contributions and value addition from each entity, presenting its priorities to UN Women as a basis for ongoing collaboration", and "Building a focus on deconstructing harmful masculinities in conflict prevention and analysis of the drivers of conflict".	
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	The recommendations appear balanced and impartial. The whole evaluation addresses GEWE and so this is integral to most of the recommendations. Equity dimensions are also embedded within them, and are explicit about the population groups at risk of being left behind, such as indigenous people, people of African descent, and persons with disabilities.	
4. Are the recommendations prioritized?	Yes	The recommendations are all prioritized by level of urgency (high or medium) and also by level of impact (high or medium).	

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?	3	<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) The subject of the evaluation and primary objective are aimed at assessing human rights and gender equality considerations. (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) Evaluation criteria were defined in terms of applicability to gender equality and women's empowerment, gender-based violence and harmful practices. (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) Questions and sub-questions (referred to as assumptions) make explicit reference to gender equality and empowerment. (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) There is less detail on this sub-criteria, though it is mentioned that a key limitation to the full integration of GEEW in the evaluation scope was the availability of sex disaggregated data. (2)</p>	

<p>2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?</p>	<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) Annex 4 specifically details the ways in which human rights and gender equality were integrated into the evaluation methodology, and notes any limitations to the full involvement of rights holders and the most vulnerable. (3)</p> <p>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analysing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3) A mixed-methods approach appropriate for evaluating GEEW considerations was applied. Annex 4 clearly defines limitations to the sample, and means for mitigating are sufficient (rights holders still consulted in case studies, for example). (3)</p> <p>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) Attention to triangulation and validation processes is clear. (3)</p> <p>d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3) Yes, and where there were limitations, they are clearly explained and were adequately mitigated. (3)</p> <p>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3) Yes, ethical standards are made explicit within the methodology chapter and Annex 4. (3)</p>
---	---

<p>3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?</p>	<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) The context section clearly sets out the relevant normative instruments and policies related to human rights and gender equality guiding UNFPA's work. There is also a clear analysis of effective strategies and progress towards the 'leaving no one behind' principle under 'Finding 2'. (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) The data analysis explicitly draws on diverse voices consulted primarily through the case studies. (3)</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) One of the evaluation's overarching objectives is to identify which factors have, positively or negatively, influenced UNFPA support to Gender Equality, which is then clearly responded to through the assessment of five assumptions. (3)</p> <p>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3) Evaluation recommendations all clearly address GEEW issues and priorities. (3)</p>
--	---

(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totaling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

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

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

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

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

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

Overall Evaluation Quality Assessment

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	82	18		
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.

OVERALL COMMENTS: Please explain the overall assessment.

Consideration of significant constraints (e.g. COVID-19 or civil unrest)

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

If yes, please explain: