



Organizational unit:  Year of report: **2022**

Title of evaluation report: **Formative evaluation of the UNFPA engagement in the reform of the United Nations development system**

Overall quality of report: Very good Date of assessment: **19 December 2022**

**Overall comments:** This is an institutional evaluation of UNFPA's contribution to the UNDS reform and of the effects of UNDS reform on UNFPA's strategic positioning. Overall, this is an excellent evaluation report that can be used by decision-makers with confidence. The evaluation draws on a mixed-methods approach combining both qualitative and quantitative data. The methodology used a range of methods to gather the perspectives of more than 250 UNFPA staff and external stakeholders at global, regional, and country levels. The evaluation used a strong participatory approach, with findings validated not only through consultations with the ERG but also through the use of short surveys administered to UNFPA country staff. Even so, the methodology presents a few gaps, namely because it does not clearly explain ethical safeguards and the sampling strategy. This said, the evaluation presents robust findings that are substantiated by a strong evidence base and answer all of the evaluation questions. The findings systematically cite data sources and the data triangulation process is clear. Likewise, the evaluation presents a strong set of conclusions and recommendations, each of which are strategic and propose concrete actions to improve UNFPA involvement in the UNDS reform going forward. Finally, the evaluation satisfactorily integrates cross-cutting issues of gender equality and equity. However, the cross-cutting issue of disability is not adequately addressed. Indeed, the methodology explains that the evaluation is aligned with the UNFPA EQAA system and the UN Disability Inclusion Strategy, and that the evaluation examines disability inclusion across all questions. However, aside from mentioning briefly that three UNFPA country offices engaged in an inter-agency working group on disability, the findings do not discuss disability inclusion. In addition, disability inclusion is not addressed in the conclusions or recommendations.

**Assessment Levels** Very Good strong, above average, best practice Good satisfactory, respectable Fair with some weaknesses, still acceptable Unsatisfactory weak, does not meet minimal quality standards

**Quality Assessment Criteria** Insert 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*

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	Yes	The report is logically structured, starting with the introduction, context and methodology, which are then followed by the findings, conclusions, and recommendations. Overall, the report is well written, easy to understand, and is free from grammatical, spelling and punctuation errors.
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

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	At 76 pages, the report surpasses the length requirements for institutional evaluations (established at 60 pages).
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----	--------------------------------------------------------------------------------------------------------------------

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 complete. These include, among others: the evaluation ToR, a list of persons met, a bibliography, an evaluation matrix, data collection tools, an evidence matrix, and additional details on the methodology.
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

*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	Overall, the executive summary summarizes fairly well the evaluation report and includes most required elements. These include the purpose, objectives, scope and intended users as well as the methodology. Findings, conclusions and recommendations are also well summarized. However, the introductory paragraphs of the executive summary could have provided more information on what the UNDS reform is.
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	At 4 1/2 pages, the executive summary is concise.
-----------------------------------------------------------------------------------------	-----	---------------------------------------------------

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

*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?	Yes	The context provides an excellent overview of reforms undertaken in the UN since 1997, with a particular focus on the latest UNDS reform. It provides a good overview of the new Resident Coordinator (RC) system and its management and accountability framework. The context also explains the structural changes in the UN system at regional and country level as well as processes and mechanisms to enhance efficiency. Finally, this section describes the legal framework (GA resolutions) underpinning the UNDS reform.
----------------------------------------------------------------------------------------------------------------	-----	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2. Does the evaluation report discuss and assess the intervention logic and/or theory of change?	Yes	The report clearly describes the logic underpinning the UNDS reform. A reconstructed theory of change (Figure 4) depicts the contribution of UNFPA to the UNDS reform on the one hand and, on the other hand, the effects of the UNDS reform on UNFPA. It should also be noted that the theory of change informed the development of the evaluation questions, which are themselves embedded in the ToC.
--------------------------------------------------------------------------------------------------	-----	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

*To ensure a rigorous design and methodology*

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 evaluation framework is clearly described. An evaluation matrix is presented in Annex V. This evaluation matrix outlines the evaluation questions and key assumptions to be assessed for each question, as well as a series of indicators, sources of information and data collection methods. In addition, Annex IV - which provides additional detail on the methodology - gives a thorough explanation of key issues to be explored under each question.
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

4. Are the tools for data collection described and their choice justified?  <i>Remember: Please address both aspects of this sub-criteria in the comment: 1) are data collection tools described (i.e. documentary review, interviews, focus group discussions etc.) and 2) is the rationale for their selection detailed</i>	Yes	Data collection tools are clearly outlined. These include: document review, a review of monitoring and survey data, KIIs at HQ and RO, and country studies. Overall, the evaluation team reviewed more than 200 documents and consulted with 266 stakeholders at global, regional, and country level (covering a total of nine countries). Data collection tools are well justified.
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)?  <i>Remember: Please address all three aspects of this sub-criteria in the comment: 1) is a comprehensive stakeholder map included (in either the report itself or the annexes) 2) Is the overall stakeholder consultation process described and 3) within the consultation process were key</i>	Yes	A comprehensive stakeholder map is presented in Annex VIII. The stakeholder map identifies key stakeholders at global, regional and country level, as well as their responsibilities within UNFPA and their roles in the UNDS reform more broadly. The report also specifies that evaluation findings, conclusions and recommendations were validated with the ERG. In addition, short online surveys were sent to select stakeholders to validate findings from the country studies and fill gaps. This can be considered good practice.
6. Are the methods for analysis clearly described for all types of data?  <i>Remember: This sub-criteria is asking about the data analysis methods used and whether they are clearly described - was contribution analysis used, or qualitative comparative analysis, for example, or descriptive statistics? Triangulation is not a method of analysis; it is a validation technique.</i>	Yes	The report clearly describes the methods of analysis (which are further detailed in Annex IV, Section F). The evaluation matrix and assumptions were used as a framework to analyze data (Annex VII includes the evidence matrix showing the triangulation process for each evaluation question). Analysis by differentiated contexts based on business model tiers, income-levels, vulnerability and humanitarian response was also conducted. In addition, Discussion Paper 2 (found in Volume 3) includes an in-depth quantitative analysis of the positioning of UNFPA's transformative results at country level. The methods of analysis for this analysis are explained.
7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?	Partial	The methodology clearly describes the evaluation's limitation and their effects on the evaluation. However, mitigation strategies are not discussed for all limitations.
8. Is the sampling strategy described?  <i>This criteria is asking whether the methodological approach to determining the sample of stakeholders consulted and the sample of site visits is described.</i>	No	The sampling strategy is not described. The methodology (see Annex IV) identifies the case study countries but in a footnote makes reference to the Inception Report for information on the selection criteria. Good practice is for the sampling strategy to be explained in the evaluation report itself.
9. Does the methodology enable the collection and analysis of disaggregated data?  <i>Remember: The default should be to disaggregate by sex. Whenever possible, this sub-criterion is also asking systematic disaggregation of data related to population groups (e.g. persons with</i>	Yes	This criterion does not fully apply for institutional evaluations. However, the evaluation presents disaggregated data wherever possible (for example, data from the online survey with UNFPA staff [Annex IX] has several levels of disaggregation - e.g. by region, country income level, humanitarian vs. development status). Findings also disaggregated between country and regional offices.
10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)?  <i>Remember: This sub-criteria is asking about the evaluation methodology itself – specifically does the evaluation's design/methodology ensure that the evaluation is able to assess the extent to</i>	Yes	The report confirms that the evaluation adheres to the UNEG gender-related norms and standards, meets UN-SWAP requirements, and is aligned with the UN Disability Inclusion Strategy. The report specifies that "human rights, gender equality, disability and other elements of leaving no one behind (LNOB)" were examined across all seven evaluation questions. Cross-cutting issues are well integrated across the evaluation matrix.
<b>3. Reliability of Data</b>	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	Findings are effective at identifying data sources from documents, KIIs and the online survey, making the data triangulation process very clear. In addition, Annex VII includes an evidence matrix that demonstrates, for each evaluation question, how evidence from different methods was triangulated to come up with the findings.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?  <i>Remember: Please address both parts of this sub-criteria, namely do the evaluators identify the sources of the qualitative and quantitative data they used and do they discuss the reliability (or lack thereof) of both?</i>	Yes	The evaluation used mostly qualitative data which were quantified as relevant, and also presents some quantitative data from a staff survey. In addition, the findings used quantitative data from UN-Info on cost avoidance accrued from the implementation of the Business Operations Strategy. Finally, a discussion paper (see Vol. 3, Annex 2) presents an in-depth quantitative analysis of the positioning of UNFPA's transformative results at country level. Data presented is reliable.
3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?  <i>Remember: Ensure that reviewers are assessing the sub-criteria is concerned with whether there is</i>	Partial	The methodology states that the evaluation "abides by the UNEG Ethical Guidelines and Code of Conduct. However, the report does not describe the ethical safeguards put in place for this evaluation nor does it address confidentiality issues.
<b>4. Analysis and Findings</b>	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	Key findings are clearly presented using bold finding statements, making it easy for the reader to understand the evaluation's main messages. These finding statements are supported by a robust evidence based elaborated through a coherent narrative. The narrative presents the views of multiple stakeholder groups at global, regional and country level, making the evidence particularly strong.
2. Is the basis for interpretations carefully described?	Yes	The findings carefully present the evidence used for interpretation and data gaps potentially making the evidence inconclusive are outlined.
3. Is the analysis presented against the evaluation questions?	Yes	Findings are presented according to and systematically address all the evaluation questions.
4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?  <i>Remember: Please address both parts of the sub-criteria in the comment, namely: are the</i>	Yes	This criterion does not fully apply as it is a process-oriented evaluation. However, the use of the reconstructed theory of change to understand causal linkages regarding the effects of the UNDS reform on UNFPA's strategic positioning is clear. And, although unintended results are not explicitly assessed, the evaluators do examine risks factors in the achievement of results.
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	This is an evaluation of UNFPA's contribution to the UNDS reform and of the effects of UNDS reform on UNFPA's strategic positioning. Therefore, this criterion is not fully applicable. However, the findings clearly examine the effects of the UNDS reform on the functioning of different target offices, namely regional and country offices.

6. Is the analysis presented against contextual factors?	Yes	The findings thoroughly discuss enabling and hindering factors affecting the effectiveness of UNFPA's contribution to the UNDS reform as well as factors influencing the effects of the UNDS reform on UNFPA's strategic positioning. For example, the findings examine how the Covid-19 pandemic has adversely affected and provided opportunities for inter-agency collaboration and the implementation of the UNDS reform.
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Partial	The analysis adequately integrates most cross-cutting issues (see Finding 19, among others). For example, the findings note that the UNDS reform has strengthened the strategic positioning of UNFPA as a leader on gender equality and youth in the UNCT. Similarly, it found that UNFPA has been taking a leadership role at country level on issues of Protection from Sexual Exploitation and Abuse (PSEA). However, even though the methodology says that the evaluation addresses disability inclusion, the findings do not meaningfully consider this issue. Indeed, Table 9 indicates that three UNFPA country offices engaged in an inter-agency working group on disability, but the findings do not provide additional detail on this limited engagement.

<b>5. Conclusions</b>	Yes No Partial	Assessment Level:	<b>Good</b>
<i>To assess the validity of conclusions</i>			
1. Do the conclusions flow clearly from the findings?	Yes	The evaluation presents six conclusions, each of which are explicitly linked to their corresponding findings.	
<b>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?</b>	Partial	Structured around six issues, the conclusions are pitched at a high level and go beyond a simple summary of findings. The conclusions also discuss the implications of findings for the future of UNFPA's involvement in the UNDS reform. However, the conclusions do not adequately reflect on cross-cutting issues; while gender and youth are very briefly mentioned, disability inclusion is not addressed.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions make reference to the evidence presented in the findings and convey the evaluators' unbiased judgement.	
<b>6. Recommendations</b>			
Assessment Level: <b>Very good</b>			
<i>To ensure the usefulness and clarity of recommendations</i>			
1. Do recommendations flow logically from conclusions?	Yes	Each recommendation is explicitly linked to a conclusion.	
<b>2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?</b>	Yes	Each recommendation is targeted at specific users (i.e. divisions or offices within UNFPA). The implications for implementing the recommendations are discussed and specific actions are proposed to guide their operationalization.	
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?	Partial	Recommendations are balanced and impartial. Recommendations also address gender equality by recommending the development of corporate guidance to ensure that indicators on female genital mutilation and child marriage are incorporated in UNSDCF's. However, disability inclusion is not addressed.	
4. Are the recommendations prioritized?	Yes	Each recommendation is prioritized (i.e. very high/high/medium).	
<b>7. Gender</b>			
Assessment Level: <b>Very good</b>			
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) <sup>(#)</sup></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><b>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)</b> The evaluation does not include a specific objective assessing GEEW. However, GEEW is integrated in the scope in that the evaluation examined "the engagement of UNFPA in the UNDS reform to ensure that sexual and reproductive health rights (SRHR) and responses to gender-based violence (GBV) are integrated into emergency preparedness and responses on the ground." (Score: 2)</p> <p><b>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)</b> The evaluation does not include a separate criterion on GEWE, but the methodology confirms that "human rights, gender equality, disability and other elements of leaving no one behind (LNOB)" were examined across all seven evaluation questions. (Score: 3)</p> <p><b>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3)</b> The evaluation does not have a specific question or sub-question on GEEW. (Score: 0)</p> <p><b>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)</b> This criterion does not apply given the process-oriented nature of this evaluation (Score: n/a)</p>		
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?	<p><b>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)</b> The report confirms that the evaluation adheres to the UNEG gender-related norms and standards, meets UN-SWAP requirements, and is aligned with the UN Disability Inclusion Strategy. The report specifies that "human rights, gender equality, disability and other elements of leaving no one behind (LNOB)" were examined across all seven evaluation questions. Cross-cutting issues are well integrated across the evaluation matrix. (Score: 3)</p> <p><b>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)</b> The evaluation used mixed-methods with both qualitative and quantitative data (Score: 3)</p> <p><b>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3)</b> The evaluation used multiple data sources and methods to ensure data triangulation (Score: 3)</p> <p><b>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)</b> The evaluation consulted with a wide variety of stakeholders at global, regional, and country level. Considering that this is a process-oriented evaluation, vulnerable groups were not consulted (Score: 3)</p> <p><b>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3)</b> The report does not describe the ethical safeguards put in place for this evaluation nor does it address confidentiality issues.</p>		

<p>3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?</p>	<p><b>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)</b> The context does not address gender-related issues (Score: 0)</p> <p><b>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)</b> This criterion does not apply considering the process-oriented nature of this evaluation. (Score: n/a)</p> <p><b>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3)</b> This criterion does not apply considering the process-oriented nature of this evaluation. (Score: n/a)</p> <p><b>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)</b> Gender is integrated in both the findings and recommendations (Score: 3)</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			
<b>Total scoring points</b>	<b>82</b>	<b>18</b>		

**Overall assessment level of evaluation report** **Very good**

The evaluation integrates adequately cross cutting issues of gender equality, human rights and disability inclusion, even though is not included as part of the evaluation objective.

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

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

If yes, please explain:  
 The evaluation was undertaken under covid travel restrictions which limited broad stakeholder engagement, especially with governments.