



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

Title of evaluation report: **UNFPA Bhutan 7th Country Programme Evaluation 2019-23**

Overall quality of report: **Very good** Date of assessment: **13 March 2023**

Overall comments: This is a strong evaluation of the UNFPA Country Programme (CP) in Bhutan. The evaluators reconstructed the Theory of Change (ToC) through desk research and presentations by the Country Technical Team (CT), and then employed a mixed-methods approach that included desk review, semi-structured interviews, focus group discussions, and direct observations to collect both qualitative and quantitative data. The evaluators used purposive sampling based on clear indicators to select the main locations for data collection, although the sampling process for respondents was not provided. There was a careful description of the process used to analyze both qualitative and quantitative data, including content analysis for qualitative data and descriptive and contributory analyses for quantitative data. A wide range of stakeholders from the national level to the district level were consulted, including UNFPA, other UN agencies, national government, academia, national NGOs, and district officials, and triangulation is evident in the report. Focus group discussions were held to gather input from vulnerable groups including women, Youth Peer Education Networks (Y-PEER), Youth Centre participants, and the LGBTIQ community. The perspectives of different stakeholder groups are also captured in Annex Part 2-E, which includes notes from interviews, field visits, and focus group discussions. The findings are detailed and well evidenced and include a thorough analysis of gender and of cross-cutting themes (including disability inclusion). The recommendations are specific and appear very useful for the next CP. One limitation is that the report is text heavy and would benefit from the inclusion of more visual aids to succinctly communicate key issues and added design elements to better distinguish sections and headings. That said, a useful addition that goes beyond the annexes typically provided is Annex 2-B which includes graphic elements to depict the extent to which the CP is contributing to the relevant SDGs.

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

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

1. Structure and Clarity of Reporting Yes No Partial Assessment Level: **Very 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)? Partial The report follows the usual structure and the writing is easy to understand. However, the findings section in particular is quite text heavy. Although italicized key findings statements help to bring out the main points, further use of visual aids would increase the accessibility of this section, as would the use of colour for headings, shading of boxes, etc.

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) Yes The report, which is just over 57 pages long, is an appropriate length for a CPE.

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 All elements are included. A useful addition that goes beyond the annexes typically provided is Annex 2-B which clearly shows the extent to which the CP is contributing to the relevant SDGs.

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? Yes Executive summary includes all essential elements and is written as a stand-alone document that offers a thorough and informative overview.

5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)? Yes At 4 pages, the executive summary is well within the permitted length.

2. Design and Methodology Yes No Partial Assessment Level: **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? Yes The country context section clearly describes the development and institutional context of evaluation in Bhutan, including its population demographics, health sector, gender equality, and development challenges and opportunities. It also discusses Bhutan's national strategies for addressing these issues, as well as the impact of the COVID-19 pandemic and natural calamities. The information is supplemented by a detailed Key Facts table at the beginning of the report.

2. Does the evaluation report discuss and assess the intervention logic and/or theory of change? Partial Evaluators adopted a theory-based approach, which involved reviewing the ToC developed under the CP and reconstructing it based on desk review and presentations by CT. It is noted that the reconstructed ToC formed the basis for developing the evaluation matrix and was tested during the field phase. One issue not addressed by the evaluators was the minimal difference between the output and outcome statements for the Adolescent and Youth thematic area. A more precise outcome would be that this population uses the services.

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 evaluators are clear about using a non-experimental design. The evaluation matrix is discussed in the report body and included in Annex 2. The matrix includes the assumptions to be assessed, indicators, sources of information, and methods and tools for data collection. Although main findings collected during data collection are not included within the evaluation matrix, they are provided in a separate annex, Annex 2-E.
4. Are the tools for data collection described and their choice justified?	Yes	The evaluation report describes the data collection tools used, including desk review, semi-structured interviews, focus group discussions, and direct observations. It is noted that the selection of these tools was based on the evaluation questions and ToC.
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)?	Yes	Annex 4 contains a list of stakeholders categorized by organization, gender, and title. Similarly, table 3 in the report's body provides a breakdown of the people interviewed by institutional affiliations, disaggregated by gender, with the total number of males and females and their corresponding percentages. This is good practice. Additionally, it is noted that draft findings, conclusions, and recommendations were shared with the CO and key stakeholders, including the External Reference Group (ERG), for validation.
6. Are the methods for analysis clearly described for all types of data?	Yes	Evaluators have clearly described the methods used to analyze both qualitative and quantitative data under section 1.3.2. Qualitative data was analyzed using content analysis to identify emerging trends, supplemented with analysis of quantitative data. Descriptive analysis was used to identify the context and types of interventions, while it is noted that contributory analysis was conducted to assess the programme's contribution to expected results.
7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?	Partial	Limitations related to purposive sampling and availability of quantitative data and data disaggregated by age, sex, and geography is noted under section 1.4. However, it would be useful to also discuss what was done to mitigate the effect of these issues.
8. Is the sampling strategy described?	Partial	Evaluators used purposive sampling to determine locations for data collection based on number of interventions being implemented, types of beneficiaries, socio-economic characteristics, and ease of travel. The criteria for selection is provided. However, it would be useful to provide information on how the sample was determined for the KIs and FGDs.
9. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The evaluation matrix includes indicators on disaggregated data. Evaluators note unavailability of recent disaggregated data limiting analysis.
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 evaluation report includes EQ4 which assesses the integration of human rights, gender perspectives, and disability inclusion in the country program. The evaluation matrix includes indicators that cover gender and vulnerable groups including persons with disabilities (PWDs). Data collection tools were designed to collect data on disability and human rights. However, it would have been useful if the report provided more details on how the data methodology was gender-responsive and whether any accommodations were made to include vulnerable groups, such as persons with disabilities (for example, whether and how PWDs or their representatives included as evaluation participants).
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	Evaluators used quantitative data sources, such as annual reports, studies, reports, and financial data, to cross-check and validate the qualitative data collected during the evaluation. Triangulation processes are explained.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	A wide range of stakeholders from national level to district level were consulted, including UNFPA, other UN agencies, national government, academia, national NGOs, and district officials. It is also noted that focus group discussions were held with members of organizations of persons with disabilities and LGBTIQ.
3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	It is noted that the evaluation followed UNFPA evaluation policy and UNEG guidance. Evaluators obtained informed consent before all interviews and ensured confidentiality of information during interviews. Consent was also obtained from teachers for students below 18 years in school settings. The FGD protocol also has instructions about avoiding 'hierarchical' levels amongst participants.
4. Analysis and Findings		
	Yes No Partial	Assessment Level: Very good
<i>To ensure sound analysis and credible findings</i>		
1. Are the findings substantiated by evidence?	Yes	The findings are supported by both qualitative and quantitative evidence. In many cases, the sources are referenced very generally with the statement, "The sources of information for the findings in this section are mainly from the document review, stakeholder interviews and focus group discussions (Annex Part 2 E) and other sources as specified in the footnote ", but in most sections more specific documents and stakeholder groups are also cited. The inclusion of direct quotes from participants is also helpful in this regard.

2. Is the basis for interpretations carefully described?	Yes	Evaluators are careful in interpreting data. For example, under effectiveness criteria it is noted "A critical gap that was observed during stakeholder interviews is the lack of support for strengthening the RMNCAH indicator data collection through the DHIS 2, which could have provided more recent data (though the data has its limitations). The reason could be the lack of coordination by WHO (lead agency for DHIS 2) and MOH in the initial stages of development and implementation of DHIS 2."
3. Is the analysis presented against the evaluation questions?	Yes	This section provides analysis against the evaluation questions, and a summary at the start of the each question gives an overview of its main findings. In addition, key findings statements are highlighted throughout the narrative.
4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	Evaluators provide causal linkage between outputs and outcomes. This is particularly evident in the effectiveness section. Additionally, Annex 6 shows the progress of outcome and output indicators. Four positive unintended effects are listed in section 4.3.b.
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	This is done. For example, under effectiveness criteria it is noted that "UNFPA promoted the inclusion of the LGBTQ+ and PWDs in the service delivery, including advocating for their access to the services. There was however a confirmation from interviews with stakeholders that there was inadequate integration of PWDs in programming which may need to be enhanced."
6. Is the analysis presented against contextual factors?	Yes	The analysis in the findings section takes into account contextual factors. For example, it is noted that due to COVID-19 pandemic, activities and progress were affected due to focus on pandemic mitigation. In addition, section 4.3a has a discussion on facilitating and hindering factors.
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Yes	The analysis considered several groups, including adolescents, youth, persons with disabilities, LGBTQ individuals, women and girls, and low-income groups.
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 flow clearly from the findings, and indicate the evaluation questions they are based upon.
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	Yes	The conclusions are well presented. They are pitched at a higher level than findings and provide a good overview of the underlying issues including cross-cutting themes.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	The conclusions appear to be unbiased as they clearly respond to the evidence and analysis.
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	The recommendations are based on conclusions and indicate the conclusions they are linked to.
2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	The recommendations are targeted at the intended users and come with detailed action plans. Some of the recommendation indicate the requirement of additional funding and technical assistance.
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. Several recommendations address cross-cutting issues such as equity and vulnerability, disability-inclusion, gender equality, and human rights.
4. Are the recommendations prioritized?	Yes	Recommendations are prioritized as medium or high. They are categorized as strategic or programmatic, with one overall recommendation related to preparations for the design stage of the next CP!
7. Gender		
	0 1 2 3 (**)	Assessment Level: Very good
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (**)</i>		

<p>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>	<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 evaluation scope includes cross-cutting issues such as human rights, gender equality, and disability. =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 is not covered as a separate criterion but is mainstreamed into other criteria. The evaluation matrix includes assumptions and indicators covering HRGE. =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) Gender is covered under two sub-questions EQ1 (Relevance) and EQ4 (Effectiveness). =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) Evaluators note "availability of disaggregated data on SRHR needs of PWD and other vulnerable populations is a serious gap that also makes it difficult to advocate for the needs of the vulnerable groups." and provide a recommendation to build capacity to generate credible disaggregated data (including for vulnerable populations). =3</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) The evaluators note partnerships with religious institutions, organizations representing PWD, volunteer/youth networks, and LGBTIQ communities. However, the report lacks a clear description of how gender was factored into the data collection and analysis methods, or whether any accommodations were made on the basis of gender, vulnerability, or disability. =1</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) The evaluation employs a mixed-methods approach that is suitable for assessing gender equality and women's empowerment (GEEW) considerations. The sample size was appropriate. =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) A wide range of respondents were consulted including organizations representing persons with disabilities (PWD), volunteer/youth networks and Lesbian, Gay, Bisexual, Transgender, Intersex, Queer (LGBTIQ) communities. Triangulation is evident. =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) Annex 4 presents a comprehensive list of stakeholders who were interviewed, highlighting their designation, sex, organization, and thematic area. Furthermore, the total number of respondents is provided by category and sex at the end of the list. In addition to this, focus group discussions (FGDs) were conducted with various vulnerable groups, such as women (7), Youth Peer Education Network (Y-PEER) (2 females and 6 males), Youth Centre participants (6 females and 1 male), and those identifying as LGBTIQ (6). =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) It is noted that the evaluation was carried according to the UNFPA evaluation policy and UNEG guidance. Evaluators describe that prior to conducting any interviews, informed consent was taken, and in the case of school settings where the students were below 18 years of age, consent was sought from their teachers. The evaluators ensured adherence to other ethical considerations, and during each interview, they guaranteed the confidentiality and purpose of the information. Furthermore, confidentiality and consent are also highlighted during data collection. =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 background section notes that women in rural communities are more vulnerable and have limited access to resources, and that more than two in five Bhutanese women have experienced abuse by a spouse at some point in their lives. It also highlights the challenges faced by adolescents and youth, particularly in terms of access to employment, education, and comprehensive sexuality education, as well as high rates of early marriage, teenage pregnancy, and HIV/AIDS prevalence. The background section also spells out relevant normative instruments and policies related to human rights and gender equality, including the Bhutan Gender Policy Note 2013. =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 analysis covers various stakeholder groups, including vulnerable groups such as persons with disabilities, victims of gender-based violence, young people, and the LGBTIQ community. The perspectives of these groups are further captured in Annex Part 2-E, which includes notes from interviews, field visits, and focus group discussions. =3</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) The report describes unanticipated effects of the intervention on human rights and gender equality. For example, the focus on preventing teenage pregnancy led to consultations on MTP and advocacy for broadening the scope of the guidance. Similarly, consultations on the prevention of GBV revealed issues such as sexual exploitation, abuse, and harassment, resulted in the development of PSEAH policies in academic and monastic institutions, as well as at the national level. = 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) Recommendations cover gender, disability, and human rights. For example, Recommendation 1 focuses conducting a stakeholder analysis and developing SMART indicators with a focus on gender sensitivity and human rights-based approaches. Whereas Recommendation 6 proposes that CP 8 should focus on GBV prevention and management, with a multi-sectoral and multi-partnership approach, and support research on sociocultural reasons underlying GBV. =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			
Total scoring points	87	13		
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.

The report includes a solid analysis of cross-cutting themes, including disability.

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:

