



Organizational unit:	Sao Tome and Principe	Year of report:	2021
Title of evaluation report:	EVALUATION DU 7 ^e PROGRAMME DE PAYS, SAO TOME ET PRINCIPE 2017-2021		
Overall quality of report:	Very good	Date of assessment:	21/02/2022

Overall comments: This is a very thorough evaluation of the 7th Country Programme of Sao Tome and Principe 2017-2021. Evaluators adopted mixed-methods to collect qualitative as well as quantitative data, including in-person data at the community level. Extensive stakeholder mapping was conducted to identify all relevant duty bearers and rightsholders and to inform the sampling procedures. The methodology was rigorous, particularly in the approach to sampling and analysis, and was adequate to assess cross-cutting issues although ethical considerations could have been more fully explained. The report includes a comprehensive evaluation matrix which also presents data obtained from documentary review and interviews for summarizing evaluation results, and shows the extent to which triangulation was done, demonstrating good practice. The findings are clearly organized by evaluation question and assumptions, and address both output and outcome level results. The conclusions are based on findings, well organized, and comprehensive although a more succinct presentation would make this section more accessible. Recommendations are also clearly formulated and appear to provide useful direction for the next CP. People with disabilities are covered as part of marginalized and vulnerable groups in two of the evaluation questions (under Relevance). As such, Findings reflect disability analysis, and one of the detailed recommendations in the annexes addresses PWD. The report makes good use of visual aids but would have benefited from a final review of their formatting.

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)*

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)?	Yes	The report is clearly structured with good definition between sections making it easy to follow. There is some, but limited, use of rights-based terminology. Evaluators included a good range of visual aids to helpfully convey key information, although there are some issues with formatting (i.e., some figures on p. 8 are not aligned with the text and are blurry, and figures 10 and 11 are compressed and difficult to read). It would be helpful if the final report was submitted as a pdf to ensure that formatting holds.
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	At 70 pages, the report is reasonable in length for a country programme 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	Annexes include all required elements.

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	The Executive Summary is well formulated. It provides a succinct presentation of the required elements and serves as a useful stand alone section for decision makers particularly due to the detail included in the recommendations.
5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	At 4 pages, the summary is reasonable in length.

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

To ensure that the evaluation is put within its context

1. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	A clear description of Sao Tome's country context is provided including a sociodemographic profile, country challenges, gender and equity development challenges, country progress towards SR / P&D / gender / HR development targets, and the role of external assistance.
2. Does the evaluation report discuss and assess the intervention logic and/or theory of change?	Yes	The evaluators adopted a solid theory-based approach to assess the extent to which results were achieved and to describe how and why an expected change took place. The evaluators assessed the ToC and reconstructed it.

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 matrix is discussed in the main report and provided in Annex 4. It includes assumptions to check, indicators, sources of information, and data collection methods and tools. The matrix is not only used to structure data collection but also for analysis purposes with the actual data from documentary review and interviews for summarizing evaluation results being inserted. Additionally, it shows the extent to which triangulation was done. This is a good practice since it helps to ensure a systematic approach is followed and it allows readers to see the evidence available through each data source.
4. Are the tools for data collection described and their choice justified?	Yes	Data collection methods were thorough and included desk review, direct observations, 13 individual interviews with duty bearers and rightsholders, 26 group interviews with duty bearers and rightsholders, and 5 FGDs with rightsholders including men, women, and teenagers. Although there is not explicit justification for the choice of each, the author emphasizes that the four techniques enabled the collection of qualitative and quantitative data, a participatory approach, and triangulation.
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	Evaluators conducted stakeholder mapping to identify all involved stakeholders. The map is provided in Annex 7 - it includes UNFPA, implementing agencies, other partners, rights holders, and marginalized groups including people with disabilities.

6. Are the methods for analysis clearly described for all types of data?	Yes	There is a careful explanation of the approach to data analysis which was based on the ToC and its hypotheses. Content analysis was employed with Figure 2 effectively illustrating the sequence description, inference, and interpretation processes that were part of that analysis. In addition, the annexes include a grid for document review and analysis.
7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?	Yes	Limitations and their mitigation strategies are clearly described on page 9. They are related to unavailability of data sources, access to sites, language challenges, and recording issues in the field.
8. Is the sampling strategy described?	Yes	Evaluators explain the use of 2-stage sampling (national and sub-national- district levels) involving stratification of 2 types (horizontal "thematic" or statutory "the human rights approach" and vertical "level of the pyramid of intervention").
9. Does the methodology enable the collection and analysis of disaggregated data?	Yes	The evaluation matrix and data collection tools are designed to collect data disaggregated by sex and age. The findings subsequently present gender disaggregated data.
10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)?	Yes	The methodology is appropriate - evaluation questions as well as data collection tools focus on cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights).
3. Reliability of Data		
	Yes No Partial	Assessment Level: Very good
<i>To ensure quality of data and robust data collection processes</i>		
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluators used different data sources including interviews with different stakeholders, desk review, and direct observations to allow triangulation. Additionally, the evaluation matrix shows the extent to which triangulation was possible through use of multiple data sources.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	Both qualitative and quantitative data was collected from sources appearing to be reliable. Evaluators highlight that the observation tool was a key tool for gathering quantitative data. Issues of data reliability (mainly in respect to availability and note-taking) are discussed under the section on evaluability, limitations and mitigation measures.
3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Partial	It is noted that evaluation adhered to guidelines and principles of the United Nations Evaluation Group and the UNFPA Evaluation Manual. There is no explanation of how these were applied although covid safety precautions are mentioned. All but one of the data collection protocols shows that informed consent and permission to take photos was to be obtained, and that confidentiality was to be addressed. It would be useful if there was a specific subsection within methodology or an annex that provided a fuller explanation of ethical considerations, including on how group discussions enabled free and full participation (or the limitations in this regard).
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	Findings are backed by adequate qualitative and quantitative evidence. Sources are noted in the report body as well as in footnotes.
2. Is the basis for interpretations carefully described?	Yes	The basis for the interpretation is consistently shown. For example, under Effectiveness it is noted that "At least 5 members per association (out-of-school and school) have been trained in CSE. The evolution of these indicators suggests an increase in knowledge aimed at young people in all their diversity".
3. Is the analysis presented against the evaluation questions?	Yes	The findings are organized by evaluation question and assumptions. Under each question, there is a brief description of the findings which adds to the clarity of this section.
4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes	Causal linkages between outputs and results are the focus of question 2 of Effectiveness with the evaluators carefully looking at where outcome level results can be seen and where indicators are not sufficient to capture this information. There is not explicit mention of unintended outcomes although these can be derived to some extent from the section on Adaptation to Crisis and in the discussion on SRH.
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	Findings clearly show outcomes for different groups and where more progress could be made. For example, the analysis takes up the specific SRH needs for LGBTI populations, for adolescents and isolated people, and for those in prisons.
6. Is the analysis presented against contextual factors?	Yes	Contextual factors are frequently identified. For example, in the section on sustainability of the SRH programme it is explained that, "The taboo of the distribution of condoms to teenagers by teachers led to their suspension at school with a decline in good RH practices".
7. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, disability inclusion, gender equality and human rights?	Yes	Cross-cutting issues including equity, vulnerability, disability inclusion, and HRGE were a major focus of the evaluation and are included in the analysis of Relevance, Effectiveness and Sustainability. The findings show how vulnerable groups were included in planning processes and where there were implementation gaps, for example that, "the inhabitants of the secondary island of Ilhéu das Rolas, LGBTI people, people with disabilities and indigents were excluded due to physical, psycho-social, physio-motor and economic barriers".
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?	Partial	The conclusions are based on findings and organized by strategic and programmatic conclusions. However their usefulness is somewhat limited by the length of this section which covers 5.5 pages. A more succinct presentation would make the main results of the evaluation more accessible, increasing the likelihood of this section being read in full.
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	Conclusions are adequately forward looking and thoroughly cover both strengths and weaknesses of the interventions. Cross-cutting issues of gender, equity, disability inclusion are addressed, and limitations in reaching all populations are brought out.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no indication of bias.

6. Recommendations		Yes No Partial	Assessment Level:	Very good
<i>To ensure the usefulness and clarity of recommendations</i>				
1. Do recommendations flow logically from conclusions?	Yes	Recommendations are clearly formulated and mention the conclusions on which they are based.		
2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes	Recommendations are directed towards intended users and provide information on financial, human, and technical implications.		
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	This section appears to be balanced and impartial. There are three recommendations that address cross-cutting issues (gender, human rights, equity, and communication). Detailed recommendations with costing of innovations are presented in Annex 16, and within these, people with disabilities are addressed as part of vulnerable and marginal groups.		
4. Are the recommendations prioritized?	Yes	Priorities are assigned on a scale of one to three. This section's formatting is very effective in terms of presenting and highlighting the key recommendations, and is an example of good practice.		
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?			<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 assessment of cross-cutting issues, including gender equality, is highlighted in the scope of the evaluation = 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 mainstreamed into the evaluation criteria as a cross-cutting issue = 3</p> <p>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) Yes, there is a specific question = 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) The background and findings section presents gender and age-disaggregated data. = 3</p>	
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?			<p>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex? (Score: 0-3) Data sampling processes and collection tools were designed to assess HRGE. However, there is not a specific discussion on gender responsive and the number of evaluation participants are not gender or age disaggregated= 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) An appropriate mixed-methods approach was used = 3</p> <p>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) A wide-range of data sources and stakeholders were covered to ensure triangulation, including both male and female rights-holders= 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) The stakeholder mapping included vulnerable groups (pregnant women, those with low literacy, GBV victims, unemployed people, teenage women and children GBV victims); and marginalized groups such as (youth and women with disabilities, PLHIV, street children, LGBTI). The difficulty in accessing all groups is noted as a limitation and the report could have been more clear on the extent to which representative participation was achieved = 2</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) Evaluators note that the evaluation was conducted in accordance with the procedures of the United Nations Evaluation Group and the UNFPA Evaluation Manual, and there is evidence of issues of informed consent and confidentiality being applied in the evaluation protocols. However, it would have been useful to more fully describe ethical considerations applied, particularly for the in-person group discussions. =1</p>	
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?			<p>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality? (Score: 0-3) Gender issues are covered well in the context section including gender inequality and equity in access to services. = 3</p> <p>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable? (Score: 0-3) There is strong gender analysis throughout the findings section. = 3</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) Unintended outcomes are shown. For example, it is noted that "The taboo of the distribution of condoms to teenagers by teachers led to their suspension at school with a decline in good RH practices". = 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) There are several recommendations directly addressing gender and human rights issues = 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	100			
Overall assessment level of evaluation report	Very good			