EQA for UNFPA Timor Leste Country Programme

Evaluation (2009-2013)

Title of Evaluation Report: Timor Leste, The UNFPA Second Country Programme (2009-2013): Evaluation Report of March 2013

Time-frame for the CPE stated in the ToR: 35 days
Cost of the CPE: $60,000[1]

Overall Assessment: The evaluation does not meet UNFPA quality standards. It does provide a useful look at the programme but there are notable divergences from the respective standards specified in the Quality Assessment criteria in each of the respective categories. The report **structure** includes the required chapters, but does not have Context as a separate chapter nor Transferable Lessons Learned. The **executive summary** provides an overview of the evaluation. The report explains **methodological** choices for data collection and selection of sites for visits. Nevertheless, methods for **data** collection and sources of data are not specified in detail. The **findings** stem from unclear data analysis and are supported by weak evidence. The links between findings and conclusions are not accurately explained; **conclusions** lack supportive arguments and do not make clear why the changes happened in the way they were observed. **Recommendations** are strategic and represent a general course of future actions.

Quality Assessment criteria	Assessment Levels			
	Very Good	**Good**	**Poor**	**Unsatisfactory**
1. Structure and Clarity of Reporting To ensure report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards. Checklist of minimum content and sequence required for structure: • i) Acronyms; ii) Exec Summary; iii) Introduction; iv) Methodology including Approach and Limitations; v) Context; vi) Findings/Analysis; vii) Conclusions; viii) Recommendations; ix) Transferable Lessons Learned	**Good** The evaluation report contains the required sections. Appendix 3 includes Evaluation Questions for Reproductive Health. The Evaluation Purpose clearly explains why the evaluation was undertaken. The "Context" is provided, although misplaced in the section labeled "Background" which is followed by Evaluation Purpose. However, methodological approach and instruments (protocols) are missing. Annex 6 List of Partners Met is not a list of interviews. Acronyms are sometimes used – e.g., "PD" on p 14 – without the full phrase in the text, making the reference obscure without			

	reference to List of Acronyms. The list contains two abbreviations in Portuguese without translation to English (language of the report). Some abbreviations in the main body of the report (e.g. STI, ARH) are not included in the List of Abbreviations and Acronyms.
(where applicable) • Minimum requirements for Annexes: ToRs; Bibliography List of interviewees; Methodological instruments used.	
2. Executive Summary To provide an overview of the evaluation, written as a stand-alone section and presenting main results of the evaluation. Structure (paragraph equates to half page max): • i) Purpose, including intended audience(s); ii) Objectives and Brief description of intervention (1 para); iii) Methodology (1 para); iv) Main Conclusions (1 para); v) Recommendations (1 para). Maximum length 3-4 page	**Poor** The Executive Summary provides an overview of the evaluation, is written as a stand-alone section and presents the main results of the evaluation. It contains all necessary parts. The Executive Summary is five pages long (which is slightly longer than necessary). There are issues with the length of some sections, for example the section on Evaluation Objectives and Interventions includes 6 paragraphs and is a page long. In contrast, the Evaluation Methodology is only one paragraph and includes general sentences that are relevant to all kinds of evaluations, for instance, "data to support the evaluation assessment were derived from primary and secondary sources representing both national and sub-national level stakeholders." Main Conclusions section is one and a half pages long and does not contain clear conclusions of the evaluation on the extent to which CP2 achieved its outputs and contributed to its intended outcomes, and the extent to which CP2 helped to enhance government commitment to the ICPD programme of action, Millennium Development Goals (MDGs) and other national priorities. The main conclusion in the Executive Summary about the design of the overall programme is not found in the body of the report. Recommendations section is one and a half pages long, and is written as a narrative text which is hard to read and does not reveal the key points.
3. Design and Methodology To provide a clear explanation of the following	**Poor** The design takes into account and builds on the programme's logic,

elements/tools Minimum content and sequence: - Explanation of methodological choice, including constraints and limitations; - Techniques and Tools for data collection provided in a detailed manner; - Triangulation systematically applied throughout the evaluation; - Details of participatory stakeholders' consultation process are provided. - Whenever relevant, specific attention to cross-cutting issues (vulnerable groups, youth, gender equality) in the design of the evaluation	explicitly noting specific outputs and outcomes. The report includes a general explanation of methodological choice, including constraints and limitations. The evaluation consultants claim that triangulation was systematically applied throughout the evaluation "through application of perceptions, validation, and documentation." However, the explanation of methodological choices is not detailed: - No explanation of chosen methods and approaches, especially those related to sampling, data collection, and data analysis were provided by the evaluation team (ToR requirement, p.87), and there is an absence of annexes with any kind of analyses. - There is no analysis of programme logic (for instance, no Theory of Change and the programme logframe analysis); no specific tables on cost efficiency analysis, and this data is presented in the text as a narrative description. - Methodological choice is not clear about (a) assessing programme relevance "at the outcome level" rather than the country level through the CPAP and (b) analysing programme effectiveness, efficiency and sustainability at the output level. This creates a disconnect between the specific purposes of the evaluation and the programme logic. Techniques and Tools for data collection are not provided in a detailed manner: "the type of data were based on a mix of quantitative and qualitative, derived from multiple sources." The main data collection method of the evaluation was interviews with the programme stakeholders and participants, but it is not clear from the report how many people were interviewed and how many meetings and focus groups were organized. The evaluation team did identify limitations, with time allowed for field work and time required to travel in country identified as major limitations which affected the extent to which they could deal with the various groups of stakeholders and how they could collect data from varying groups and validate it. The evaluation team attempted

	to mitigate the limitations, including in terms of the "lack of counterfactuals…" which was appropriate. However, no specific mention is made of who they interviewed at the NGO implementing partners or target/beneficiary groups. Another limitation noted was the inability "to cover the long-term impacts of the interventions" but this was not included in the ToR.
	Details of the participatory stakeholders' consultation process are not provided. Beneficiaries, as well as programme implementing partners, have not been included in the interview schedule although brief reference is made to "a convenient sample of beneficiaries was also used for focus group discussions" and Annex 6. No reason was given as to why this was less important than government officials. This gave the government officials a disproportionate weight and would make efforts to triangulate difficult.
	Details on how cross-cutting issues (vulnerable groups, youth, gender, equality) were addressed in the design and during the evaluation are not described in the report. The design deals primary with the three programme components, which includes gender, but has no methodological considerations for dealing with the other cross-cutting issues, which is not included in the design and methodology.
4. Reliability of Data To clarify data collection processes and data quality • Sources of qualitative and quantitative data have been identified; • Credibility of primary (e.g. interviews and focus groups) and secondary (e.g. reports) data established and limitations made explicit;	**Poor** Sources of qualitative and quantitative data are provided in general in the report, for instance, "key stakeholders including officials from SEPI, MSS, and implementing partners", etc. On the credibility of primary and secondary data, the evaluation team states that "the limited timeframe of this evaluation exercise did not allow the team to collect primary quantitative data of related areas" (p. 21). Collected evaluation data are of relatively low value for evaluation of UNFPA programme achievements, using outcomes' and outputs' indicators, because it is not clear how many interviewees were questioned, how these interviewees were selected, and who the interviewees were (there are no names and

	contact details in the appendix 6). Documents reviewed during the evaluation (Appendix 7) do not contain full citation and often have only brief documents titles or even abbreviations. To find out "how UNFPA activities have influenced the planned objectives and vice versa" community group leaders, key respondents "were identified at the national and district levels" (no lists with names and location are attached). Documents for the interviews, "consent form" and "a questionnaire protocol", are not attached. The evaluation team declared that disaggregated data by gender was utilized during the survey, but this statement is not supported by the data provided with the report.
5. Findings and Analysis To ensure sound analysis and credible findings <u>Findings</u> • Findings stem from rigorous data analysis; • Findings are substantiated by evidence; • Findings are presented in a clear manner <u>Analysis</u> • Interpretations are based on carefully described assumptions; • Contextual factors are identified. • Cause and effect links between an intervention and its end results (including unintended results) are explained.	**Poor** Evaluation findings include basic data on outputs achieved during the project but little data on outcomes. The findings begin with a description and assessment of the country programme documentation and each programme section provides a programme context and information on programme management and partnerships. Both are helpful to the reader in understanding the evaluation findings that ensue. However, findings stem from weak data analysis. Collected qualitative data are presented generally as a description of different activities without evaluation of their contribution to achieving planned outputs and outcomes. The evaluation findings contain mainly qualitative data. Findings are not substantiated by evidence and are more descriptive and less analytical, limited to reporting information/data collected. This affected the ability to base the findings on substantial evidence gathered from multiple sources and rigorous data analysis. The evaluation team members differ in the way they treat information which impairs the analysis of the findings. Interpretations are not based on carefully described assumptions.

	Some cause and effect links are well explained in the report:
	- Reproductive Health subsection provides a brief, helpful assessment of the output, outcomes and indicator as elements in the programme design. - Hindering factors used in the respective programme components is useful for future consideration. - In the Reproductive Health subsection, the issue of relevance is appropriately dealt with at the country level. However, other cause and effect links between an intervention and its end results (including unintended results) are not clearly explained. - Achievements in terms of outcomes and outputs are phrased generally, e.g. "significant progress" or "considerable achievement', without explanation of the real meaning according chosen indicators of progress. For example, it is not clear from the report how the country programme affected the MMR. - In the Reproductive Health subsection, focus is placed on assessing "outputs" at great length, one at a time, focusing on implementation of intervention outputs, with little attention given to outcomes. This results in a lack of links between "an intervention and its end results", which presumably means between outputs and outcomes and hence programme objectives. - The Population & Development subsection gives attention to the outcomes primarily in terms of their relevance, but little in terms of whether they were achieved, and focuses mostly on output in terms of the effectiveness, efficiency and sustainability. The findings section also includes some conclusions, for instance "because of the above mentioned interventions, we may come to a conclusion that CP has contributed considerably in improving comprehensive capacities especially in EmOC." (p 30).
6. Conclusions To assess the validity of conclusions	**Poor** Conclusions are organized and address all programme components.

• Conclusions are based on credible findings; • Conclusions are organized in priority order; • Conclusions must convey evaluators' unbiased judgment of the intervention.	In general, the links between findings and conclusions are not clear. It is not clear which evidence about programme outcomes and progress toward of attainment of objectives the conclusions are based on. The conclusion section includes some findings. For instance, "There was an inadequate number of qualified local staff to provide SRH care. Technical supervision was not functioning regularly." Conclusions lack supportive arguments e.g. • "The RH component made great efforts in policy advocacy for changes" (it is not clarified why the results are considered by the evaluators as "great efforts"). • "The RH component …. has made an effective contribution to the improvement of service provision" (it is not clear from the statement how much the RH components contributed to achievement of planned outputs and outcomes and how the effectiveness of the contribution was assessed (p. 77). In addition to conclusions for each of the programme components, conclusions about the overall programme success would add value for the programme evaluation. How well the programme components fit and works together is a critical part of the relevance issue which left unaddressed in this section. Most conclusions of the evaluation are described in general terms, for instance: • "The country has achieved a significant measure of success in addressing GBV and DV made possible because of financial and technical support from UNFPA under CP2" (it is not clear how to measure the "significant measure of success") (p.78). • "The Family Planning interventions have made remarkable contribution in strengthening capacity of the current system and staff members in provision of services and information" ("remarkable contribution" is not measurable).
7. Recommendations To assess the usefulness and clarity of recommendations	**Poor** Recommendations are strategic and targeted, as they represent

- Recommendations flow logically from conclusions; - Recommendations must be strategic, targeted and operationally-feasible; - Recommendations must take into account stakeholders' consultations whilst remaining impartial; - Recommendations should be presented in priority order	general course of future actions for UNFPA programme in the target region. However, it is not evident from the recommendations chapter how the recommendations flow from conclusions. Recommendations are presented as a narrative text with a mixture of objectives and activities. Even though the recommendations are strategic (they suggest continuing efforts of the UNFPA programme in the country within chosen course of action), they are not specific e.g. (pp.80-82): "continue to provide refresher training to midwives and nurses who have received training on BemOC" (it is not clear how many more midwives and nurses should be trained; and the evaluation consultants do not explain the reasons for such a recommendation); "continue to work on establishment of the National Population Commission" (recommendations do not clarify what other work is necessary for establishment of National Population Commission). It is not clear from the text whether the recommendations take into account stakeholders' consultations or not. The logic used in setting the priority order of the recommendations is hard to identify.
8. Meeting Needs To ensure that Evaluation Report responds to requirements (scope & evaluation questions/issues/DAC criteria) stated in the ToR (ToR must be annexed to the report). In the event that the ToR do not conform with commonly agreed quality standards, assess if evaluators have highlighted the deficiencies with the ToR.	**Poor** The evaluation report presents assessment of effectiveness, efficiency and sustainability with regards to outcomes and outputs and includes examples and data but, in terms of addressing effectiveness, the extent (quantitative data, costs analysis) that the expected outputs and outcomes were achieved or are likely to be achieved is difficult to understand from the evaluation report. There are no specific tables showing logical connections between expected outputs and outcomes, performance indicators and their values and data obtained with regards to achieved outputs and

	outcomes. According to the ToR for each outcome and output, issues of programme design should be assessed by using the Results and Resources Framework of the CP Action Plan (p. 85) but there is no explicit analysis of the programme design and logic (SMART formulation of outcomes and outputs) in the report and annexes.

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Unsatisfactory	Poor	Good	Very good
1. Structure and clarity of reporting (2)			2	
2. Executive summary (2)		2		
3. Design and methodology (5)		5		
4. Reliability of data (5)		5		
5. Findings and analysis (50)		50		
6. Conclusions (12)		12		
7. Recommendations (12)		12		
8. Meeting needs (12)		12		
TOTAL		98	2	

(*) Insert the multiplying factor associated with the criteria in the corresponding column e.g. - if "Finding and Analysis" has been assessed as "good", please enter the number 50 into the "Good" column. The Assessment level scoring the higher number of points will determine the overall quality of the Report

OVERALL QUALITY OF REPORT: Poor