Title of evaluation report: Botswana 5ᵗʰ Country Programme End of Programme Evaluation

OVERALL QUALITY RATING: Good

Summary: The evaluation covers a period in which UNFPA transitioned its approach because Botswana is a middle-income country and the role of development assistance has changed. It is a thorough review, largely based on documentary analysis and interviews with implementers and stakeholders. It points out problems in monitoring and evaluation in the context of delivering as one. Its analysis of the context and efficiency of the program is very solid, but as it points out, evaluation of effectiveness is more complex. It finds that programs of population and development are particularly effective, and makes a large number of practical recommendations for the next programme period.

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 (where applicable) • Minimum requirements for Annexes: ToRs; Bibliography; List of interviewees; Methodological instruments used.		**Good** The evaluation was structured according to UNFPA standards, was clearly drafted and contained all of the required annexes. There is no separate chapter "Transferable Lessons Learned." The evaluators incorporated lessons learned where appropriate in the programme findings (good practices and success stories) and in the recommendations (lessons learned).		
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.		**Good** The executive summary was thorough and self-standing. It was appropriately structured. Its length was three pages, conforming to the norm. The paragraph which is supposed to describe "Objectives and Brief description of intervention" does not have brief description of intervention.		

3. Design and Methodology *To provide a clear explanation of the following 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; • Details on how cross-cutting issues (vulnerable groups, youth, gender, equality) were addressed in the design and the conduct of the evaluation.	**Good** The evaluation was done at a time when there was limited availability of informants and this was noted. However, how those to be interviewed (other than the UNFPA Office and direct counterparts) were selected or which sites were selected for visits (other than proximity to Gambarone) was not clear in the report (although it was said to have been included in the inception report). The report explains the methodological choices, including constraints and limitations. For instance, the evaluators said that "…actual site visits were mainly limited to head offices of IPs. This was because of the distance involved to most project sites… ." Triangulation was systematically applied throughout the evaluation "In particular, quantitative and some qualitative data came from the many documents reviewed, both qualitative and quantitative data from KI interviews, and the two focus group discussions and site visit particularly provided qualitative information." The consultation process was well described. An effort to include cross-cutting issues was described. The sources of data and the collection protocols are well-described and included in the annexes.
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; • Disaggregated data by gender has been utilized where necessary.	**Very Good** The sources of data were identified, usually through informants and document analysis, and when used were reliable. The data were credible since limitations were indicated. Some gender disaggregated data were used. Credibility of primary and secondary data is established and limitations are made explicit. For instance, the reports says that "The CPE included quantitative and qualitative data from both primary and secondary sources. Findings were

	weighted and systematically triangulated from the various sources to ensure robustness." Limitations are discussed in the sub-chapter 1.3.6 "Limitations." Disaggregated data by gender are utilized in the description of county context "percent of women reported" (p. 10) and in presenting the findings "The ICPD focus also includes interrelationships between… and the empowerment of women" (p. 20), "Production of reports related to ICPD ensures that there are disaggregated data on vulnerable groups such as women and girls" (p. 27). The list of Persons/Institutions Met do not include gender component.
5. Findings and Analysis *To ensure sound analysis and credible findings* Findings • Findings stem from rigorous data analysis; • Findings are substantiated by evidence; • Findings are presented in a clear manner Analysis • 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.	Good The findings were carefully drawn, based on the data sources. The analysis noted where the system of indicators at the level of Delivering as One were inadequate and that in a number of cases causal connections with UNFPA output could not be drawn easily. The findings are organized by evaluation question and are clear. The findings include qualitative and quantitative data which are clearly presented in the Tables (4.2.1. – 4.2.3) and have references to the sources of information in the text. A minor issue is that some interpretations lack of SMART-ness. For instance, the report says "UNFPA is reported as contributing valuable insights across the areas of its mandate." It is not said which kind of insights were noted. Other example is that "UNFPA contributed to resource mobilisation together with other agencies and provided TA for the drafting of the proposal." There is no reference on discussion of particularly UNFPA input/contribution.
6. Conclusions *To assess the validity of conclusions* • Conclusions are based on credible findings; • Conclusions are organized in priority order; • Conclusions must convey evaluators' unbiased judgment of the intervention.	Poor Conclusions are divided into two parts "5.1 Strategic Level" and "5.2 Programme Level: Additional Conclusions by Programme Component." Most of conclusions summarize the findings, but conclusions

are organized with regards to the evaluation questions and are not in priority order. Moreover, there are relatively few conclusions about effectiveness compared with, say, efficiency, which understates the findings. There are references to the related recommendations. However, conclusions are written as 4-11 lines paragraphs without titles or the key idea of the conclusion.

Conclusions are based on interviews and data analysis, but there are no links to the relevant parts of the findings in the text (cross-references are absent). Some conclusions lack specific logical connections with the findings on which the conclusion was based. For instance, the authors say that "the evaluation found that the CO is insufficiently tracking the effectiveness of its programmes, and partner coordination is insufficiently effective, despite the existence of DaO mechanisms for coordination." Such words are too general: "insufficiently tracking," "insufficiently effective."

Another example of a too general conclusion is that "Opportunities are being missed for documenting good practice and sharing lessons learned among IPs. No IP site visits for quality assurance are reported to take place, for example89, since the cutting of funds." The evaluators do not specify kinds of "opportunities" and the context and do not provide the reference on appropriate finding. Also, it is not clear how opportunities relates to the phrase in the next sentence "No IP site visits."

Conclusions do convey evaluators' unbiased judgment

	of the interventions as the consultants always refer to the evaluation findings, but often without a reference on the related part in the findings. For instance, the authors say that "The evaluation found that…".
7. Recommendations *To assess the usefulness and clarity of recommendations* • 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	**Poor** Recommendations are divided into two parts "Strategic Level" and "Programme Level." Recommendations flow logically from conclusions, and there are references to the conclusions from which it they are derived. Recommendations are presented with priorities but are structured by evaluation question. Each recommendation has the assessment of priority, usually, "High priority." However, there are too many recommendations to allow for a proper follow up by the country office (35 recommendations). Recommendations are generally strategic and targeted but who is supposed to implement them is not precise. For instance, the evaluators say "Overarching Recommendation 1 to UNFPA/UN Corporate Level." The issue with the recommendation is the lack of clarity in some cases. Some words are general as in the conclusions, for instance, "The priority in the 6th CP should be, across all programme areas, high level advocacy and providing high level technical assistance." The consultants do not provide guidance to what assistance is supposed to be considered as "high level." The evaluation methodology explains that recommendations took into account stakeholders' consultations.

8. Meeting Needs	Good
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.	The evaluation meets all of the needs. It notes some issues with the way results-based management takes place in the UN Country Office. There is no separate chapter "Transferable Lessons Learned." Instead, lessons learned are incorporated into findings (for instance, as success stories) and the chapter "recommendations."

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Very good	**Good**	**Poor**	**Unsatisfactory**
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	5	71	24	

(*) 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