

Organizational unit: UNFPA **Year of report:** 2019

Title of evaluation report: Government of the Kingdom of Eswatini/UNFPA 6th Country Programme Evaluation (2016 - 2020)

Overall quality of report: **Good** **Date of assessment:** 6 December 2019

Overall comments: This report provides a comprehensive evaluation of the UNFPA Eswatini country program. The overall evaluation report is written in a logical manner and is well-structured. While data reliability and analysis is strong, the methodology section could provide a more robust description of the sampling approach, more readily detail how data was collected and analyzed, and discuss the use of disaggregated data. Furthermore, systematically ensuring the meaningful participation of women and girls in the evaluation process could further strengthen the evaluation.

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	<i>Insert <u>assessment level</u> followed by main <u>comments</u>. (use 'shading' function to give cells corresponding colour)</i>		
I. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Very good
<i>To ensure the report is comprehensive and user-friendly</i>			
1. 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?	Yes	Yes, the report is easy to read and understand.	
2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPÉs; 80 for thematic evaluations)	Yes	The report meets the 70-page standard for country programme evaluations.	
3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?	Yes	The report follows the desired structure and is structured logically.	
4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process?	Yes	Yes, the annexes contain the listed material.	
<i>Executive summary</i>			
5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Yes	The report has a standalone executive summary, which includes the main results of the evaluation.	
6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?	Yes	It follows the prescribed structure.	
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Yes	The length of the summary is four pages (with one sentence carrying over to a fifth page), and is reasonably concise.	

2. Design and Methodology	Yes No Partial	Assessment Level:	Fair
<i>To ensure that the evaluation is put within its context</i>			
1. Does the evaluation describe the target audience for the evaluation?	Yes	The target audience was briefly described as follows: "The intended audience for the CPE report are UNFPA Eswatini CO, ESARO, and Headquarters; UN agencies; government; and non-government partners."	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes	At the onset, the evaluation report provides detailed contextual information and a key facts sheet of the Eswatini Kingdom. However, a comparison of the metrics in the fact sheet to either the world fact sheets or other regional standards could have given the reader a better sense of the significance of these metrics, including referencing it in the section on country and programmatic contexts.	
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Partial	The evaluation report provides a theory of change although figure 1.1 could but have been strengthened by a description of the results in the reconstructed theory of change, as well as its adequacy (i.e. the extent to which it logically depicts expected changes). As noted on p.7, the evaluators observed that "Establishing the causal linkages between interventions, outputs and indicators was not clear, with some subjectively described, limiting understanding	
<i>To ensure a rigorous design and methodology</i>			
4. 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 includes all necessary components.	
5. Are the tools for data collection described and their choice justified?	Yes	Each of the tools (documentary review, interviews, focus groups and observation) is described and why it has been chosen is explained.	
6. 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)?	Partial	Despite the requirement for a comprehensive stakeholder mapping, the evaluation report does not include a comprehensive map with the universe of stakeholders. While the evaluation report includes a list of stakeholders consulted, it does not detail the stakeholder consultation process nor a description of their consultation on draft recommendations specifically.	
7. Are the methods for analysis clearly described for all types of data?	Yes	Analysis methods were described by the type of data analyzed. For example, the report notes that content analysis was used to analyze documentation.	
8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	The methodological limitations were acknowledged in section 1.3.4, and mitigating measures proposed.	
9. Is the sampling strategy described?	Yes	The evaluation team reported using a purposive sampling approach to select a sample of sites and stakeholders for data collection based on four criteria: (a) the size of resource allocation, (b) type of implementing partners, (c) length of engagement with the country programme and (d) the interventions implemented during the period of review. However, it would have been important to include the criteria for selecting sites visited in the methodology section of the report.	
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Documentary data as well as statistical data could be disaggregated. For qualitative data some disaggregation was possible.	
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	Although the evaluation design and methodology did not explicitly state how cross-cutting issues will be assessed, the assumptions and indicators for the relevance criteria specifically referenced the needs of the of vulnerable and special groups, such as women and youth, and those from remote areas during planning and implementation. Three evaluation questions address cross-cutting issues of vulnerability, human rights and gender equality (EQ1, EQ2, and EQ4). The list of people consulted include representatives of members of youth groups in focus groups discussions. Furthermore, the evaluation team reported limitations in reaching indirect beneficiaries to assess outcome level results due to the nature of the design of the evaluation of the country program interventions whose focus is collecting data from governments and implementing partners (p. 6).	

3. Reliability of Data	Yes No Partial	Assessment Level:	Very good
<i>To ensure quality of data and robust data collection processes are used in the findings</i>			
1. Did the evaluation triangulate data collected as appropriate?	Yes	The evaluation team reported using triangulation of quantitative data to validate the findings from the qualitative data got from key informant interviews, focus groups, and site visits/observations. However, the evaluation team could have provided a description of how they used triangulation to validate qualitative findings.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluators were explicit about the source of qualitative data coming mainly from interviews, focus groups, and observations during site visits. However, the evaluators cautioned that the sources of the quantitative was limited to secondary data from the Country Office. They reported the "unavailability of up-to-date data and lack of a robust monitoring system for GBV and institutionalization of response system are still a gap" (p.36). The lack of national-level quantitative data might have limited their abilities to corroborate data sources to validate evaluation findings.	
3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues?	Yes	The evaluation team were explicit the possible bias introduced by the reliance on secondary data provided only by the Country Office and their struggle getting reliable and quality documents from other stakeholders. The authors of the evaluation report described key factors that constrained the evaluation, mainly contextual, operational, and methodical limitations. These factors included over reliance on qualitative data due to limited availability of secondary data, reports, and statistical data from government sources, and a weak monitoring and evaluation system for collecting evaluative evidence. The authors noted that over reliance on the secondary data from the country office was a possible source bias. The team used cross-validation from stakeholders and staff in addition to using expert	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	The evaluation team reported obtaining informed consent from interview and focus group participants in compliance with UNEG ethical requirements for confidentiality, avoidance of harm, and obligation to participants.	

4. Analysis and Findings	Yes No Partial	Assessment Level:	Good
<i>To ensure sound analysis and credible findings</i>			
1. Are the findings substantiated by evidence?	Yes	The evaluation findings are substantiated by evidence garnered from previous studies like recommendations from the mid-term review of UNDAF and other supportive documents and several footnotes on sources of data. For example, to answer the evaluation question on the relevance criteria, the team referenced the Country Program alignment to the UN Sustainable Development Goals, especially SDGs 3, 5, 10, 16, and 10 (p. 20), UNFPA Strategic Plan 2018 - 2021, and to government policies and programs such as the National Development Strategy (Vision 2022) linked to the Poverty Reduction Strategy and Action Plan (PRSAP) and sector programs (pp. 22-24). The evaluation team does a good job summarizing key findings per question in textboxes and tables (4.1, 4.2, 4.3, and 4.4), then provide detailed information. Nevertheless, the team could have substantiated evaluation findings further using direct quotes from the qualitative data - notwithstanding that the fact that the "CPE was based primarily on qualitative information collected from government counterparts and implementing partners ..." (p. 6). In fact, it appears that only four direct quotes were used in chapter four of the report.	
2. Is the basis for interpretations carefully described?	Yes	The evaluation team used outcomes and indicators reported in the evaluation matrix as a means for interpretation of the findings and carefully described how these were used.	
3. Is the analysis presented against the evaluation questions?	Yes	The evaluation team used program outcomes, outputs, indicators, and targets as a basis for presenting analyses against the evaluation questions.	
4. Is the analysis transparent about the sources and quality of data?	Yes	As noted earlier, the evaluation team referenced sources of the data, but could have made better use of direct quotes to further substantiate the findings. The team expressed concern about the lack of a robust monitoring and evaluation system to capture evaluative data both within the country office and among implementing partners.	
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Partial	As noted earlier, the evaluation team presented a reconstructed theory of change model the diagram on page 4. However, while the evaluations acknowledges in the narratives that UNFPA interventions contributed to key results and outcomes, the evaluation team could have further strengthened the analysis by offering a descriptive narrative of the cause and effect links. The evaluators raised a concern about not being able to collect data from the indirect beneficiaries to assess the achievement of program outcomes. This shortfall could have been addressed had the program maintained a robust monitoring and evolution system as noted earlier. Nevertheless, the evaluation report does not provide sufficient information on what did not work well and any negative effects or unintended consequences of the interventions. For example, the evaluators could have elaborated further their findings that the country did not adopt total market approach (TMA) to increase access to and uptake of FP services (Table 4.1, p. 25) despite stakeholder consensus on the importance of the TMA (p. 31).	
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings presented different outcomes for different target groups, especially in the four tables that summarize the key findings per program outputs for women, adolescents and youth, and sectors. In discussing the implementation of the Life Skills Education (LSE) program for out-of-school youth, the evaluators discussed program performance in specific geographical contexts (pp 34 - 35), for example.	
7. Is the analysis presented against contextual factors?	Yes	Contextual factors are reflected in the narrative sections of the report. For example, the evaluation team reported the alignment of the country program to the NDS/PRSAP, the 2005 Constitution, National Gender policy (2010), National Children's Policy 2009, National Plan of Action for children 2011-2015, Social Development Policy (2010); Education Sector Policy 2018, the National SRHR Strategy 2013, and the Kingdom of Eswatini Strategic Roadmap 2018 - 2022. The evaluators noted that "Government leadership and enabling policy environment facilitated the implementation of the CP strategic interventions" (p. 23). Furthermore, the evaluators reported that inadequate government commitment and accountability, and resource constraints limited the oversight role by government to attend stakeholder planning and review meetings and the full realization of population into planning processes.	
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	The evaluators outlined the different assumptions in the evaluation matrix and indicated the type of stakeholders to be consulted. For example, to assess the assumption that, "The evolving needs of the population, in particular those of vulnerable and special groups, such as women and youth, and those from remote areas, were well taken into account during the planning and implementation processes" (p. 102), the evaluators indicated in the evaluation matrix that they would collect data through document review, key informant interviews, focus groups, and site observations. Among the key indicators for this assumption was "Extent to which the interventions supported by UNFPA were targeted at most vulnerable, disadvantaged, marginalized and excluded population groups, and retargeted as needed." As noted earlier, the evaluators referenced previous studies, key national policies, strategies and action plans, and UN SDGs in their analyses of the relevance of the UNFPA interventions to evolving needs of the population. However, the evaluators reported methodological limitations including weak M & E system within implementation frameworks especially at the service delivery levels, missing national-level data such as recent demographic and health surveys to corroborate qualitative data, and inability to collect data from indirect beneficiaries due to the nature of CPEs.	

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 follow a clear structure and are divided into strategic and programmatic conclusions. They are linked to the evaluation questions/findings from which they flow, as well as the associated recommendations.	
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated?	Yes	The evaluation team provided six strategic conclusions and 12 programmatic conclusions organized by program outcomes. The categorization of the conclusions into strategic and programmatic levels underscores the evaluators' understanding of the country and program contexts and system being evaluated. This evaluation adopted a theory-based approach with the intention of assessing the extent to which the observed differences/results were, in part, a consequence of the 2016-2020 CP implementation (UNFPA contribution to the observed results). This analytic framework partially contributed to understanding the complex operational environment of program implementation.	
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	Biased judgement could be subjective and not obvious to detect. Nevertheless, the conclusions seems objective given the key findings of the evaluation.	
6. Recommendations	Yes No Partial	Assessment Level:	Fair
<i>To ensure the usefulness and clarity of recommendations</i>			
1. Do recommendations flow logically from conclusions?	Yes	The recommendations logically flow from the conclusions, organized by the evaluation questions, conclusion, and classified into high or medium priority for both the strategic and programmatic recommendations.	
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Partial	The recommendations are broadly written with explanatory notes targeting intended users. Nevertheless, some recommendations could be strengthened if targeted at a user. For example, of the 19 recommendations, these nine do not target a user to implement the recommendations - SR4, AYI, GEWE1, GEWE2, GEWE3, GEWE4, PD2, PD3, and SRHR1. There was also only partial reference to financial implications such at Recommendation SRHR 3 which states that "The RHCS funding needs to be revived in order to strengthen the consistent supply of FP commodities" but without further specification.	
3. Do recommendations appear balanced and impartial?	Yes	No partiality is detected.	
4. Is a timeframe for implementation proposed?	Yes	The high priority recommendations are those to be implemented within a 1-2-year period whilst medium priority refers to implementation within a 3-4-year period.	
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	The classification into high and medium seems appropriate to facilitate management response, considering that the evaluators got feedback from the CP stakeholders on the recommendations.	

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?		3	While the evaluation does not include an objective or evaluation criteria specific to the to assessment of human rights and gender equality considerations, gender equality was mainstreamed, as evidenced by the evaluation questions, with several evaluation questions dedicated to the assessment of gender equality and human rights. Furthermore, the evaluation assess the extent to which information/data is available/collected to assess progress against result indicators, including specifically vis a vis gender equality and human rights, noting where data was insufficient.
2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?		3	The evaluation employs a mixed-methods approach appropriate to assessing gender equality (and the extent to which the country programme has advanced it). The methods and sampling approach are designed in a way to ensure that a diversity of voices/data is included, and indeed diverse range of data sources are consulted. Data is disaggregated, where possible, and triangulation (as well as other validation methods) is used. There is no evidence to suggest that ethical standards were violated, nor confidentiality breached.
3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?		2	The background section of the evaluation report includes a description of the relevant normative instruments and policies related to human rights and gender equality. The evaluation findings include an analysis that triangulates the voices of different social groups, particularly through the use of focus group discussions. This, however, could have been further strengthened with additional disaggregation where possible. The evaluation did not draw out unintended or unanticipated effects of UNFPA support on gender equality specifically, should there have been any. Importantly, however, the report does provide specific recommendations addressing gender equality, both programmatically and strategically.
<p>(*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totalling the scores 8-9 = "very good"; 6-7 = "good"; 5-6 = "fair"; 1-0="unsatisfactory").</p> <p>(**) Scoring uses a four point scale (0-3).</p> <p>0 = Not at all integrated. Applies when none of the elements under a criterion are met.</p> <p>1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.</p> <p>2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.</p> <p>3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.</p>			

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	36	40	24	
Overall assessment level of evaluation report		Good		
	Very good very confident to use	Good confident to use	Fair use with caution	Unsatisfactory not confident to use

- (*) (a) Insert scoring points associated with criteria in corresponding column (e.g. - if 'Analysis and findings' has been assessed as 'Good', enter 40 into 'Good' column).
 (b) Assessment level with highest 'total scoring points' determines 'Overall assessment level of evaluation report'. Write corresponding assessment level in cell (e.g. 'Fair').
 (c) Use 'shading' function to give cells corresponding colour.

If the overall assessment is 'Fair', please explain

• How it can be used?

• What aspects to be cautious about?

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory

Consideration of significant constraints

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes

No

If yes, please explain: