

Organizational unit: _____ Year of report: **2021**

Title of evaluation report: **Evaluation of the 4th UNFPA Country Programme for Kyrgyzstan (2018-2022)**

Overall quality of report: **Good** Date of assessment: **9 November 2021**

Overall comments: This is a very detailed evaluation of the 4th UNFPA Country Programme for Kyrgyzstan (2018-2022). It covers the four focus areas: 1) Sexual and Reproductive Health (SRH) 2) Adolescent and Youth (A&Y), 3) Gender, and 4) Population and Development (P&D), each being individually assessed against four out of the five standard OECD- DAC criteria: relevance, effectiveness, efficiency and sustainability (due to nature of the intervention impact was not considered); and two criteria specific to UNFPA, (I) UNFPA coordination with UNCT; and (II) UNFPA Country programme added value. A wide range of stakeholders were engaged in data collection, however, rightsholders could not be interviewed due to COVID-19 pandemic and military incidents. The context section is comprehensive. The methodology section is generally strong, clearly describing ethical considerations and limitations as well as their mitigations strategies. The evaluation used a purposive and convenience sampling strategy; however it would have been useful to describe how the universe was determined. It would also have been useful to explain how the multiple types of data analysis processes were applied. The findings are supported by both qualitative and quantitative data, and are in-depth particularly for the question regarding achievement of outputs. Causal connections between outputs and end results can be observed, however, unintended outcomes are not captured. The evaluation also does well in respect to gender responsiveness. As noted in the methodology, a disability lens was purposely applied throughout the evaluation. Disability inclusion is helpfully defined, and is discussed under Country Context, Relevance, Effectiveness (for each programme area), however disability-related results are not carried through to Recommendations.

The evaluators were meticulous in collecting, documenting and incorporating into the report large amounts of data. However, the level of detail - particularly in the Executive Summary and much of the Findings - obscured some of the key results. Similarly, both the Conclusions and Recommendations sections include useful information for decision makers but could be more clearly presented to enable a management response.

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:	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 is structured in a logical way, although the formatting style of headers (i.e., larger of coloured font) could help to make clearer distinctions between each section. The report is generally well written but with some grammatical errors. More attention could have been paid to the explanation and presentation of several graphics (i.e., there is an error in the text referencing for figures 2 - 5, the chart legend for fig 2 uses acronyms that are not explained, the dates in fig 4 are cut off, the axis for fig 5 are not labeled, the SDG charts on p 25 and graphs on p 40 are too small and blurred to read). The report is very detailed, which to some extent obscures key results-related findings - for example, there is a set of lessons learned on p 45 that appear useful but could be more clearly highlighted.
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 68 pages, the report is reasonable in length.
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	There are total of 9 annexes and they include all required elements <input type="checkbox"/>

<i>Executive summary</i>		
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 complete as a stand-alone document, conclusions and recommendations are separated into strategic and programmatic categories. (findings are combined with conclusions)
5. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Partial	At 4.5 pages the summary is within the permitted length. However, the summary is text heavy and has a high-level of detail, particularly in the presentation of methodology and conclusions, which require a careful read to find the main points. More white space and formatting (such as bolded topics of each Conclusions and Recommendation) would increase readability.

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	There is a extensive description of the country background in terms of the development challenges, national strategies, and role of external assistance.
2. Does the evaluation report discuss and assess the intervention logic and/or theory of change?	Yes	The report discusses and assesses the intervention logic / theory of change. Annex 5 presents analysis of ToC for the CPD 2018-2022.

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 framework is described in the text with Annex 4 providing a very complete showing of the questions, assumptions, indicators, data sources and methods of data collection.
4. Are the tools for data collection described and their choice justified?	Yes	Evaluators have provided a list of data collection tools in Table 2, and the rationale for their selection is also described in the narrative.

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	A stakeholder map is provided in Annex 6. The Methodology section notes that a diverse range of stakeholders were identified through a stakeholder mapping exercise to ensure compilation of well-triangulated data.
<i>Remember: Please address all three aspects of this sub-criteria in the comment: 1) is a</i>		
6. Are the methods for analysis clearly described for all types of data?	Partial	The report lists various data analysis methods used by the evaluators; these included document analysis, ToC analysis, contribution, content and trend analysis, process mapping and the use of descriptive statistics. However, their application is not described.
7. Are methodological limitations acknowledged and their effect on the evaluation described? Does the report discuss what was done to minimize such issues?	Yes	Five key methodological limitations and their mitigation strategies are described.
8. Is the sampling strategy described?	Partial	The evaluators note that several types of samples were used depending on sub-questions, tools, and resources using purposive and convenience sampling strategy. However, the evaluators do not describe how the universe was determined.
9. Does the methodology enable the collection and analysis of disaggregated data?	Yes	Gender disaggregated data is presented in the report (including in the country context section), the total number of stakeholders consulted provided in Annex 6 also indicate the number of female and male stakeholders from participating groups.
10. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, disability inclusion, gender equality and human rights)?	Yes	Cross-cutting issues dealing with GEEW, leaving no one behind, human rights, and disability inclusion are integrated into the design.
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	Triangulation is evident from the methodology and findings.
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	The evaluation was careful to identify and describe the various sources and all were used, both qualitative and quantitative. The report also discusses the quality and discrepancies in the data available.
3. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes	The report mentions that the team closely adhered to the UN Evaluation Group Code of Conduct and Ethical Guidelines for Evaluations (2008) and adopted a person-centered approach by distinctly emphasizing respect, accountability, fairness and transparency. Cultural and religious sensitivities existing in Kyrgyzstan were also considered during data collection.
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 findings in the evaluation report are supported by a substantial amount of qualitative as well as quantitative data. There are 1051 endnotes that go as far as specifying the KII # from which the data was obtained.
2. Is the basis for interpretations carefully described?	Yes	The basis on which the interpretations are made are described. For example, one paragraphs note that "Document analysis, key informants' interviews and online survey shows the following achievements".
3. Is the analysis presented against the evaluation questions?	Yes	Findings are structured according to the criteria and start with the evaluation questions and then give the specific findings in summary form before providing the detailed evidence supporting those results. However, the findings in some cases are more than detailed than is normally found in an evaluation; those for EQ3 on extent of output achievement read more like a programme report. That also leads to overly detailed summary statements of the findings for most questions.
4. Are the cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Partial	Where causal connections can be observed, they are reported. It is noted that it was difficult to quantify contribution of various outputs to planned outcomes as indicators were hard to measure (p 42). However, although constraints and facilitating factors are discussed, there is no specific mention of unintended outcomes in the main report (the ToR specifies that data analysis should highlight any unintended outcomes).
5. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The evaluation findings show differences across relevant outcomes for adolescents, youth, and adult women, and for those in rural areas.
6. Is the analysis presented against contextual factors?	Yes	The evaluators have provided described contextual constraints and facilitating factors which affected the intervention.
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 such as GEEW, vulnerability, disability inclusion, and coordination are analysed in the findings section. The evaluation also addresses coordination. Additionally, the extent to which five cross-cutting issues are integrated into formulation of national policies are also presented by a 3-point scale.
5. Conclusions		
	Yes No Partial	Assessment Level: Good
<i>To assess the validity of conclusions</i>		
1. Do the conclusions flow clearly from the findings?	Yes	Conclusions are clearly derived from findings. The origin of each conclusion is provided by the number of the respective evaluation questions contained in the text of the finding section along with associated recommendations. They are organized as strategic conclusions and programmatic level conclusions.

<p>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?</p>	<p>Partial</p>	<p>A set of 20 Conclusions is presented in table format. They go beyond the Findings and reflect appropriate cross-cutting issues including vulnerability, disability inclusion, GEEW, and human rights. However, the large number of conclusions and the text-heavy presentation with frequent long sentences, affects readability. A more concise presentation with the subject of each in bolded text would make this section more impactful.</p>
<p>3. Do the conclusions appear to convey the evaluators' unbiased judgement?</p>	<p>Yes</p>	<p>There is no evidence of bias.</p>
<p>6. Recommendations</p>		
	<p>Yes No Partial</p>	<p>Assessment Level: Good</p>
<p><i>To ensure the usefulness and clarity of recommendations</i></p>		
<p>1. Do recommendations flow logically from conclusions?</p>	<p>Yes</p>	<p>Recommendations are derived from conclusions and are presented at both strategic and programmatic levels. The number of conclusion(s) from which each recommendation is drawn is also provided.</p>
<p>2. Are the recommendations targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?</p>	<p>Partial</p>	<p>There are 5 overall recommendations and the intended users at this level are shown. Each has a paragraph on operational implications which includes conclusions and sub-recommendations. This section would be more actionable if these paragraphs were more clearly formulated as sub-recommendations with each being numbered and directed to intended users. Budgetary implications are not described for each recommendation but Recommendation 2 does address the need for more predictable and flexible funding modalities.</p>
<p>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?</p>	<p>Partial</p>	<p>Recommendations appear balanced, impartial, and cover GEEW, however, it would have been useful to address the needs of other vulnerable groups, such as people with disabilities.</p>
<p>4. Are the recommendations prioritized?</p>	<p>Yes</p>	<p>The recommendations are prioritized into high and medium priority.</p>
<p>7. Gender</p>		
	<p>0 1 2 3 (**)</p>	<p>Assessment Level: Very good</p>
<p><i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i></p>		
<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) HRGE is not specifically mentioned in the objectives, however, it is integrated in the thematic scope of the evaluation process. = 2 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) HRGE is mainstreamed under various criteria including relevance and effectiveness. = 3 c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3) This is addressed in EQ3. = 3 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 evaluation matrix also provides several indicators and sources of data covering GEEW. = 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) It is noted that evaluation methodology adhered to guidelines outlined in UNFPA's 'How to Design and Conduct a Country Programme Evaluation at UNFPA' and 'UN Evaluation Group Code of Conduct and Ethical Guidelines for Evaluations'. Additionally, it is mentioned that collected data was analysed in a gender sensitive manner to reflect gender perspectives and decrease bias. = 3 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) A mixed methods and person centric approach is used. More than 80% of the evaluation participants were female. = 3 c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3) Sources are diverse as data was collected from an adequate range of stakeholder groups. Evaluators also note that in limitations section that due to COVID-19 pandemic and military incidents, it was not feasible to interview rightsholders. = 3 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) A range of stakeholders participated, although rights holders were not included in the sample due to COVID-19 pandemic and military incidents. = 3 e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3) The application of ethical considerations is clearly described in the methodology section. = 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 country context section focuses on HRGE and social issues including GBV, and includes statistics on issues including women's representation in parliament, child marriage, and abduction of girls. = 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 some presentation of results by stakeholder group; voices would be more apparent if there was further use of quotes that were attributed to stakeholder group. =2</p> <p>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3) ² Unanticipated effects are not discussed.= 0</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) Several recommendations address GEEW issues such as GBV, equitable norms, attitudes and behaviours. = 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	31	69		
Overall assessment level of evaluation report	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.

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: