Organizational unit: **Evaluation Office - UNFPA**   **Year of report:** **2008-2015**

Title of evaluation report: **Evaluation of UNFPA Support to Adolescents and Youth**

Overall quality of report: | **Good** | **Date of assessment:** **12/19/2016**

Overall comments: This report is very clearly written, with a great use of graphics and tables to communicate information. The methodology is also well-conceptualized, particularly the selection of country case studies and the use of numerous typical and atypical sources of data including a country office survey and e-Roundtable with youth stakeholders. However, the report has two areas for improvement. First, data is often cited from internal documents that are not included as part of the report (i.e., country case studies prepared for the evaluation vs. the original data sources used in the case studies), or in generalities (i.e., "UNFPA interviews" vs. information about the interviewee's position), which prevents the reader from assessing its quality. Second, the evaluators include both a chapter on "main findings and analysis" that focuses primarily on outputs of UNFPA programming rather than cause-and-effect links, and a chapter on the reconstructed theory of change, which analyses cause-and-effect links from UNFPA programming to outputs but does not include any empirical evidence. These could have been better integrated in a way that used evidence to document the cause-and-effect links that are discussed in the reconstructed theory of change.

Assessment Levels

| **Very good:** | strong, above average, best practice | **Good:** | satisfactory, respectable | **Fair:** | with some weaknesses, still acceptable | **Unsatis-factory:** | 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	Assessment Level:	**Good**
To ensure the report is comprehensive and user-friendly • Is the report easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience)? • Is the report focused and to the point (e.g. not too lengthy)? • Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)? • Do the annexes contain – at a minimum – the ToRs; a bibliography, a list of interviewees, the evaluation matrix and methodological tools used (e.g. interview guides; focus group notes, outline of surveys)? _Executive summary_ • Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation? • 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)? • Is the executive summary reasonably concise (e.g. with a maximum length of 5-10 pages)?	Comment: The report is very well organized and is very easy to read. The use of graphics (e.g., in the executive summary p17) and tables (e.g., Table 1 on the integration of cross-cutting issues in various steps in the evaluation process) is generally very effective. In a few cases however figures could be better integrated with the text as they are not always referenced or given captions or legends (ex. the figure on the evaluation components on p30). The length of the report is consistent with other UNFPA evaluation reports (100-200 pages). The report's structure is logical, with fairly standard organization and clearly organized sub-sections. However, there is a section on the "ex-post theory of change" inserted between the findings and conclusions. This section would be more effective if it were better integrated with the empirical section, because as it is currently presented it the assertions about the causal relationships in the theory of change are not explicitly supported by any empirical findings. The annexes contain the required sections. The executive summary (pp15-25) is clear and includes subsections describing the purpose and audience of the evaluation, intervention, methodology, conclusions and recommendations. It is concise. The executive summary does not include the evaluation questions or criteria, however, which limits its ability to function as a stand-alone document.	

2. Design and Methodology	Assessment Level: **Good**
To ensure that the evaluation is put within its context • Does the evaluation describe whether the evaluation is for accountability and/or learning purposes? • Does the evaluation describe the target audience for the evaluation? • Is the development and institutional context of the evaluation clearly described? • Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change? • Does the evaluation explain any constraints and/or general limitations? *To ensure a rigorous design and methodology* • Is the evaluation approach and framework clearly described? Does it establish the evaluation questions, assumptions, indicators, data sources and methods for data collection? • Were the methods chosen appropriate for addressing the evaluation questions? Are the tools for data collection described and justified? • Is the methods for analysis clearly described? • Are methodological limitations acknowledged and their impact on the evaluation described? (Does it discuss how any bias has been overcome?) • Is the sampling strategy described? Does the design include validation techniques? • Is there evidence of involvement of stakeholders in the evaluation design? (Is there a comprehensive/credible stakeholder map?)	Comment: The evaluation describes clearly the objectives of the evaluation: to assess how frameworks have guided interventions, and to facilitate learning to inform further interventions (p16). The target audience for the evaluation is specified clearly as "UNFPA staff at all levels, UNFPA public and private sector implementing partners, civil society organisations, policy makers and donors as well as the end beneficiaries of UNFPA support" (p16) and discussed in detail throughout the report. A discussion of global context of adolescents and youth is included (p30), with effort made to place the evaluation in this development and institutional context providing relevant demographic data/trends (p42), definitions (p34), and reference to institutional/organizational responses (e.g. International Conference on Population and Development, p34). Limitations and associated mitigation strategies are discussed, with a focus on limitations of the data collection activities of the evaluation (Table 8, p29). The report references the reconstruction of the intervention logic/theory of change for UNFPA supports to adolescents and youth based on reviewed UNFPA documents together with the report's findings; however, a full discussion of the theory of change/intervention logic is included in Volume 2 (not Volume 1) of the report series (p18). For example, the report indicated that "[t]he quality and completeness of documentation provided at global and regional level varied"; in response, the evaluation conducted a "thorough review … to assess the documentation available…[and] Additional literature was sought external to UNFPA to provide

• Does the methodology enable the collection and analysis of disaggregated data? • Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	alternative sources of information where data gaps exist or external validation was required" (p29-30). The evaluation highlights several limitations and the steps that the evaluators took to minimize them (pp37-8). The evaluation framework is clearly described. The evaluation questions are outlined in the methodology section (p31). The methodology for data collection and analysis is clearly described and is rigorous. The evaluation uses multiple methods to collect data, ranging from standard methods like stakeholder mapping and document review, to a country office survey and e-roundtable on facebook with youth. Much of the collected data is disaggregated. The methods for analysing the data are also described in detail and appropriate. The countries selected for case studies were sampled to provide variation in the region and the UNFPA country quadrant classification system (pp33-4), as well as data availability and other criteria. The evaluators did a stakeholder mapping, and consulted internal and external stakeholders (pg37).

3. Reliability of Data	Assessment Level:	Good

To ensure quality of data and robust data collection processes

- Did the evaluation triangulate all data collected?
- Did the evaluation clearly identify and make use of qualitative and quantitative data sources?
- Did the evaluation make explicit any possible issues (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues? I.e. did the evaluation make explicit possible limitations of the data collected?
- Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?
- Is there adequate gender disaggregation of data? And if this has not been possible, is it explained?
- Does the evaluation make explicit the level of involvement of different stakeholders in the different phases of the evaluation process?

Comment:

The evaluation leverages both quantitative and qualitative data at appropriate points. There is evidence of triangulation in that the evaluators cite multiple different sources of information (often, reports and interviews with UNFPA staff) for the same piece of data. Qualitative (e.g. financial data) and qualitative (e.g. e-Roundtable responses) were used as appropriate and identified correctly throughout the report.

The evaluators note the limitations of their data sources, particularly gaps in coverage, and cross-check data to mitigate this bias (pg38).

Data is disaggregated by gender where possible.

The evaluation is explicit in when and how it involved stakeholders. However, more detail could be provided regarding the level of involvement for stakeholders across the phases of the evaluation process (p18)

4. Analysis and Findings	Assessment Level:	**Good**

To ensure sound analysis

- Is information analysed and interpreted systematically and logically?
- Are the interpretations based on carefully described assumptions?
- Is the analysis presented against the evaluation questions?
- Is the analysis transparent about the sources and quality of data?
- Are possible cause and effect links between an intervention and its end results explained?
- Where possible, is the analysis disaggregated to show different outcomes between different target groups?
- Are unintended results identified?
- Is the analysis presented against contextual factors?
- Does the analysis include reflection of the views of different stakeholders (reflecting diverse interests)? E.g. how were possible divergent opinions treated in the analysis?
- Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?

To ensure credible findings

- Can evidence be traced through the analysis into findings? E.g. are the findings substantiated by evidence?
- Do findings follow logically from the analysis?
- Is the analysis of cross-cutting issues integrated in the findings?

Comment:

The interpretation of data/information gathered is analysed in a systematic and logical manner, delineated by sub-sections and annotated with definitions and figures when necessary. Assumptions are clearly defined throughout, with analysis presented against the evaluation questions.

The report often remains at a level of generality that makes it difficult to assess the sources and quality of the data. For example, while the report includes copious citations, many of the citations are to internal documents that are not included as part of the evaluation report or annexes or merely cite the type of data rather than the actual source. For example, on pg54 the evaluators cite the following as sources for a shift in UNFPA policy: "Country case studies: Côte d'Ivoire, Egypt, Ethiopia, Kyrgyzstan, Lao PDR, Mozambique, Nepal, Nicaragua, Niger. Regional reviews: APRO, ESARO, LACRO. Interviews: UNFPA staff." These citations do not actually permit a reader to assess the quality of this evidence. The evaluators should consider including the country and regional case studies in the annexes if they want to cite them this way, and instead of citing the broad category of "UNFPA interviews," should cite specific types of interviewees by position and/or office. When examples are provided (ex. short discussions of the case studies of Kyrgyzstan and Nicaragua on pp59) it greatly improves the quality of the evidence.

The report identifies possible cause and effect links between the intervention and end results: e.g. the evaluation notes 'Good Practices' that influenced results, such as the case of Lao People's Democratic Republic in which a study on barriers to youth friendly

	health services informed evidence-based planning which improved service provision (p49, External alignment between A&Y programming, national priorities and A&Y needs). The evaluation is in some sections overly focused on outputs rather than outcomes. For example, evaluation question #3 asks to what extent UNFPA has contributed to an increase in the availability of sexual and reproductive health education and services (pg70). The second sub-section header claims that "strengthened national capacities contributed to increasing availability and use of integrated reproductive health services for adolescents and youth" (pg74). However, most of the information presented in this section is about UNFPA programming, rather than the effects of that programming with the exception of a quick mention of the Lao PDR program's effects on young people's access (pp77-8).
	When outcomes are discussed, the findings and analysis section generally does not include a rigorous discussion of cause-and-effect links. For example, on pp89-90, the effects of UNFPA's advocacy and technical assistance on national legislation are discussed, and the evaluators mention that in Senegal 5,470 communities have abandoned FGM, but the links from UNFPA programs to those changes are not described. Contextual factors are mentioned in a number of cases in the results section.
	In the section titled "ex-post theory of change," there is a thorough discussion of contextual factors and cause-and-effect links but it is not well-connected to the underlying data and analysis.
	There is little discussion of unintended results.

	The analysis reflects the views of a number of different stakeholders, including UNFPA staff, independent reports, youth, and partners in government.
	The findings do flow from the evidence, except in the case of discussions of effectiveness for the reasons outlined above.
	Cross-cutting issues were also presented clearly within the analysis of findings including Human rights (p52), gender responsiveness (p53), cultural sensitivity (p58), marginalized and vulnerable groups (Figure 16).Analysis of cross-cutting issues like equity, gender, and human rights is well-integrated into the findings.

5. Conclusions	Assessment Level:	**Good**
To assess the validity of conclusions • Are conclusions credible and clearly related to the findings? • Are the conclusions demonstrating an appropriate level of analytical abstraction? • Are conclusions conveying the evaluators' unbiased judgement of the intervention?	Comment: Conclusions are organized clearly by 'cluster': conclusions on the strategic positioning of UNFPA support to adolescent and youth; conclusions on program areas for UNFPA support to increase quality, effectiveness, and sustainability; conclusions on use of data for evidence-based programming and learning (p124). However, this could be made easier. The evaluators list the evaluation questions and criteria below each conclusion (i.e., EQ 1.2, EQ 7.1, and EQ 7.2 under conclusion 1) but these numbers seem to refer to assumptions in the evaluation matrix (i.e., A1.2, A7.1, and A7.2 on pp4-17 of the annexes). The conclusions demonstrate the appropriate level of abstraction. The conclusions highlight both strengths and weaknesses of UNFPA's programs and there is no evidence of bias.	

6. Recommendations	Assessment Level: **Very good**
To ensure the usefulness and clarity of recommendations • Do recommendations flow logically from conclusions? • Are the recommendations sufficiently clear, targeted at the intended users and operationally-feasible? • Do recommendations reflect stakeholders' consultations whilst remaining balanced and impartial? • Is the number of recommendations manageable? • Are the recommendations prioritised and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Comment: Recommendations flow logically from the conclusions and are linked to specific conclusions in the text. Recommendations flow logically from conclusions, with specific reference to which conclusion a recommendation relates to (e.g. Recommendation 1 connected to Conclusions 2, 3, 4, and 10) (p138). The rationale for each recommendation is presented, followed by 'Action Points' that are operationally feasible. The recommendations are targeted with intended users defined (and specified clearly in the 'Target Level' comments for each recommendation). The number of recommendations (eight) is reasonable, and they are broken down into 34 specific "action points". The report indicates the priority level of each recommendation clearly There is no evidence of bias in the recommendations, and they seem to reflect consultation with stakeholders.

7. Gender	Assessment Level:	**Fair**
To assess the integration of Gender Equality and Empowerment of Women (GEEW)[1] • Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected? • Do evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved? • Have gender-responsive evaluation methodology, methods and tools, and data analysis techniques been selected? • Do the evaluation findings, conclusions and recommendations reflect a gender analysis?	Comment: GEEW is included in the evaluation scope of analysis, with explicit mention of "initiatives to reach marginalized and disadvantaged adolescents and youth, especially girls" (p16) and specific attention given to the "integration of cross-cutting issues such as gender equity" (p17). Data collected is disaggregated by gender where appropriate, to allow GEEW to be integrated into analysis. The theory of change is not presented (rather, it is presented in Volume 2) – so it is not clear whether GEEW gender-sensitive indicators have been included. However, the report does discuss ways in which gender equality is integrated into the methodological approach and data collection (p20). In the e-Roundtable, a total of 190 youth leaders were contacted, however less than half participated. It was stated that attention was paid to the selection process to ensure gender balance, however the participants are only disaggregated by region, so it is unclear how many female youth leaders were invited to participate and how many actually participated. More efforts should be made throughout the report to disaggregate data by gender, especially when such data is available. For example, how many female youth leaders were available and what were the barriers, if any, to their participation? This type of disaggregated data would further clarify and substantiate their findings. The findings, conclusions, and recommendations take gender into account.	

---

[1] 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 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

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	11	82	7	
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 'finding and analysis' 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[2]:

- How it can be used?

- What aspects to be cautious about?

Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory[3]:

Consideration of significant constraints[4]

The quality of this evaluation report has been hampered by exceptionally difficult circumstances: X yes   no

If yes, please explain:

The scope of field studies was reduced as a result of the 2015 earthquake in Nepal. Nepal was subsequently included as a desk study.

---

[2] The purpose here is to clarify in what way the report can be used. This in order to assist the elaboration of a relevant Management Response and the wider use of the evaluation findings back into programming. When a report has been assessed as Fair, it is obligatory to fill this text box in.

[3] The purpose is, where relevant, to clarify for example severe unbalances in the report (for example, the report is good overall but recommendations very weak). Is optional to fill in.

[4] E.g. this should only be used in case of significant events that has severely hampering the evaluation process like natural disasters, evaluators falling sick, unexpected significant travel restrictions, etc. More 'normal' limitations should be mentioned under relevant section above.