

Organizational unit:		Year of report:	2017
Title of evaluation	Formative Evaluation of the UNFPA Innovation Initiative		
Overall quality of report:	Good	Date of assessment:	31 July 2017

Overall comments: The report is called a "formative evaluation" in which the focus would be on why things happened rather than what, but in fact it also includes, as it must, summative data. It is written to UNFPA standards and uses graphics well. While it is not structured according to the evaluation questions the logic for not doing so is given by the evaluators. The data acquisition approach is at standard, the findings are very numerous and are not easy to read. The conclusions, however, partly solve this problem by consolidating the findings in such a way that recommendations can be made that are realistic, even if not always clearly time-specific. Gender is dealt with relatively well given the complexities of 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	Insert <u>assessment level</u> followed by main <u>comments</u> . (use 'shading' function to give cells corresponding colour)		
I. Structure and Clarity of Reporting	Yes No Partial	Assessment Level:	Fair
<i>To ensure the report is comprehensive and user-friendly</i>		Comment:	
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	1. The report is easy to read. There are a few small grammatical errors (e.g. page xv reads "It has also helped staff reconnected and realign with the organization's mandate.") The evaluation has clear structure, presented in an easy to read and understandable, non-technical manner. It is well organized, thorough, focused and to the point.	
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	2. At about 74 pages (excluding annexes), the report is reasonably long.	
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	3. The report is structured according to UNFPA guidelines, but while mentioning lessons learned in several places does not include a section on them.	
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	4. All required annexes are available, including a long separate annex on the reconstruction of the theory of change. Stakeholders and the stakeholder consultations are described in several annexes (in particular annexes 2, 5 and 7). However, information on the stakeholder consultation process could have benefited from a specific annex.	
<i>Executive summary</i>		5. The Executive Summary is written as a stand-alone document and follows the guidance from UNFPA.	
5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?	Yes	6. The Executive Summary follows the required structure.	
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	7. At 7 pages, the executive summary is a bit too long.	
7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?	Partial		

2. Design and Methodology	Yes No Partial	Assessment Level:	Fair
<i>To ensure that the evaluation is put within its context</i>		Comment:	
1. Does the evaluation describe the target audience for the evaluation?	Yes	<p>The evaluation is entitled a "formative evaluation" and an effort is made to show what this is and why it is better in this case than a "summative evaluation". However, all evaluations are both summative (showing what happened) and formative (why or why not things happened) and this is no exception. Most of the findings and conclusions are based on data about what happened and from this conclusions about why they happened or not are drawn.</p> <p>1. While the target audience is not formally stated, it is clear that it is the management of the Innovation Fund, UNFPA human resources and higher management at headquarters.</p> <p>2. Both the context and background of the evaluation are thoroughly described in the chapter 2 (pp. 15-19). Constraints are explained in the section 1.4.10 (pp. 12 - 14).</p> <p>3. The evaluation includes a separate volume describing the methods for reconstructing the theory of change. What is particularly notable in this evaluation is its explanation of the reconstruction of the implicit theory of change: detailed description of the theory of change (p. 3) and comprehensive analysis of the theory of change (pp. 6-10). The analysis presents critical overview of the assumptions that have not held true either partially or entirely, as well as Outcome pathways that have not unfolded as expected, including the revised version of the theory of change. In the meantime, the evaluators refer to the section 2.4., section 3.4., and the figure 3.1. which cannot be found in the report (p. 6)</p> <p>4. The evaluation framework meets UNFPA standards. The evaluation questions are clear and are derived from UNEG standards. The evaluation matrix is found in the annexes, but uses a slightly different structure than would be suggested by the UNFPA Handbook by using categories like dimensions and pointers and fairly large numbers of indicators not well-connected with the expected results. How it translated into the data collection process is not well explained in the main text.</p> <p>5. Tools for data collection are clearly described in the section 1.4.6. "Tools for data collection" (pp. 9-10). The evaluators briefly justify their choice on the pp. 7-8. They clearly explain that "Cyber search data collection ... was applied systematically for the comparative analysis with other UN agencies as well as when assessing visibility and branding aspects included in evaluation questions" (p. 9). In other cases, the evaluators describe the data collection tools without justification (pp. 9-11).</p> <p>6. There is a stakeholder map in Figure 5 and the consultation process, through a committee and through organized interaction with stakeholders, is clear.</p> <p>7. The methods for analysis are clearly described for all types of data in the section 1.4.7.</p>	
2. Is the development and institutional context of the evaluation clearly described and constraints explained?	Yes		
3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?	Yes		
<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?	Partial		
5. Are the tools for data collection described and their choice justified?	Partial		
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)?	Yes		
7. Are the methods for analysis clearly described for all types of data?	Yes		

8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)	Yes	<p>8. The methodological limitations are well-described in section 1.4.10 Risks, limitations and quality assurance and in each case how the limitation was mitigated is also described. The evaluators state that "Limitations were mostly related to data collection" (p. 12). Effects of evaluation limitations are described and mitigating measures are implemented, for instance, the evaluators clarify that "...online survey's low response rate for non-applicants ... limited the generalizations of findings..." (p. 13). The report discusses how any bias has been overcome, for instance: "...this finding should be treated with caution...Self-selection bias could conceal that survey respondents implement more innovation projects than the rest of non-applicants" (p. 64).</p> <p>9. The sampling strategy is well-described and justified and, in one case, where a random stratified sample was replaced by a census, the reason for so doing is explained. Of interest is how a low response in one group (field offices submitting proposals that had not been funded) was addressed by moving from a sample to a census.</p> <p>10. The evaluation shows clearly how disaggregated data were collected and analyzed. 11. Two cross-cutting issues were noted: gender and youth. Gender was consistent with normal standards. The issue of youth, however, was slightly problematic in that the method was "to gather data on variables related to youth (e.g. the distribution of UNFPA staff by age range and region and average age of the workforce), staff attrition rates by age range and region, and number of interns and the proportion of them who stayed in UNFPA after the internship." Since youth is normally considered to be persons between 15 and 24 years of age, very few UNFPA staff are likely to be youth. And, as the evaluators noted, they did not have good data on interns (many of whom would be "youth").</p>
9. Is the sampling strategy described?	Yes	
10. Does the methodology enable the collection and analysis of disaggregated data?	Yes	
11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?	Yes	

3. Reliability of Data	Yes No Partial	Assessment Level:	Very good
<i>To ensure quality of data and robust data collection processes</i>		Comment:	
1. Did the evaluation triangulate data collected as appropriate?	Yes	1. The evaluation consistently used a triangulation process in making its findings on specific question.	
2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?	Yes	2. The data sources, mostly a combination of interviews, case studies and use of quantitative data, were used effectively. 3. The data limitations were clearly indicated (p. 13) and the mitigating measures for each described.	
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	4. There is evidence that UNFPA and UNEG ethical standards were used such as, for example, interviewees were kept confidential.	
4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?	Yes		

4. Analysis and Findings	Yes No Partial	Assessment Level:	Good
<i>To ensure sound analysis and credible findings</i>		Comment:	
1. Are the findings substantiated by evidence?	Yes	<p>1. The findings are always shown with the evidentiary basis clear.</p> <p>2. The basis for interpretations are carefully described throughout the analysis which is particularly notable in this evaluation. For instance, the evaluators discuss that “whereas interviews pointed at the widespread perception that the number of staff in country offices mattered when deciding whether to apply to the Innovation Fund, quantitative data reveals no clear evidence of such an effect. It cannot be concluded that office size (in terms of staff) explains Innovation Fund participation” (p. 23). Another example: “To test the hypothesis, the evaluation team ran a logistic regression model, calculating the statistical significance of the odds of being awarded if one is a member of the IDWG against the odds of not being a member” (p. 31).</p> <p>3. The evaluators made a decision not to present the analysis in terms of the evaluation questions, although they claim that all questions were addressed. The report noted (p. 20) that “Within each level, the aspects addressed correspond to the topics that were considered most relevant and useful to evaluation users, given the decision-making processes going on at the time of report writing (UNFPA Innovation Business Case, input to the Strategic Plan 2018-2021). This analysis chapter is thus not structured around answers to the evaluation questions but around topics of interest for evaluation users. Some of the questions in the original evaluation matrix do not appear in this chapter because they were not considered to be a priority at the time of report writing.” 4. 14. The analysis is transparent about the sources and quality of data, for instance, the consultants argue that “All relevant documentation on the Innovation Fund is available on the Innovation Hub. However, the site is not hosted as a regular MyUNFPA site, which makes it difficult to find, as shown by the scant number of informants reporting visits to the site” (p. 30).</p> <p>5. The evaluation makes a clear effort to show causal links. In a number of cases, where a causal link could be based on a number of different factors, the evaluators applied a multivariate analysis to see whether a causal link was supported by the data.</p> <p>6. The evaluators were careful to identify different target groups, by type of country, by region and by headquarters office. Unintended outcomes are discussed in the report, for instance: “Beyond the initial effects it had in building momentum for the buy-in of innovation in UNFPA, the project had additional positive unintended effects” (p. 46), “Yet the current activity- and expected output-based systems makes it challenging to capture unintended and unexpected effects” (p. 53).</p> <p>7. The contextual factors in findings were clearly shown. 1. The analysis is disaggregated to show different outcomes for different countries. Figure 21 presents comparison of funded projects by Thematic areas.</p> <p>8. While gender was not well covered in the evaluation, an effort was made to include the youth factor, particularly when an innovation was related to information technology. One of the cases examined in detail was Hack for Youth.</p>	
2. Is the basis for interpretations carefully described?	Yes		
3. Is the analysis presented against the evaluation questions?	Partial		
4. Is the analysis transparent about the sources and quality of data?	Yes		
5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?	Yes		
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes		
7. Is the analysis presented against contextual factors?	Yes		
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes		

5. Conclusions	Yes No Partial	Assessment Level:	Very good
<i>To assess the validity of conclusions</i>		Comment:	
1. Do the conclusions flow clearly from the findings?	Yes	<p>1. While the findings are organized around a different structure than the evaluation questions, the conclusions succeed in arriving at general findings that indicate how well the innovation programme is working. There is a rough match between the eight conclusions and the ten evaluation questions.</p> <p>2. The conclusions organize the findings in a more coherent package, since there were many more findings than the evaluation questions.</p> <p>3. The conclusions appear to convey the evaluators' unbiased judgement. Even though the conclusions lack of references to the analysis section, they look specific and objective.</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?	Yes		
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes		

6. Recommendations	Yes No Partial	Assessment Level:	Good
<i>To ensure the usefulness and clarity of recommendations</i>		Comment:	
1. Do recommendations flow logically from conclusions?	Yes	<p>1. The nine recommendations flow from the conclusions, although there are no references to specific conclusions, and are mostly related to planning, monitoring and administration.</p> <p>2. The recommendations specify the intended user in terms of whether headquarters or regional and, within headquarters, by type of unit.</p> <p>3. The recommendations are balanced and impartial.</p> <p>4. While some recommendations have a time horizon (e.g. incorporation into the 2018-2021 Strategic Plan), others are general (e.g. 9. "Develop a learning-for-impact framework").</p> <p>5. The recommendations include operational suggestions that can be used to facilitate the management response. The recommendations are clearly presented to facilitate appropriate management response and follow up on each specific recommendation. The evaluation stated that "In some cases, ...the actions called for in the recommendations have already been acted upon or were being addressed at the time of writing this report" (p. 75).</p>	
2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?	Yes		
3. Do recommendations appear balanced and impartial?	Yes		
4. Is a timeframe for implementation proposed?	Partial		
5. Are the recommendations prioritised and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes		

7. Gender	0 1 2 3 (***)	Assessment Level:	Good
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) (*)</i>		Comment:	
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	1. GEEW was included into the evaluation scope as one of the UNFPA program/project areas (pp. 44-46). The evaluation matrix does not have actual indicators, but it has GEEW into the Dimensions category D1.3 "Incorporation of key crosscutting themes: gender-sensitivity; human rights..." (p. 22). Section P.3.2.7. of the column "Pointers" states that "Evidence of the participants in activities having experienced discrimination or lost opportunities due to age and/or gender" (p. 25). During the design stage of the evaluation they developed an "Innovation project tracking sheet", ... with information on all innovation project analysed. Empowerment of young people and gender equality for example, where categories used" (p. 33). One major gender element is the analysis of a project entitled "Gender Transformative Media Programming" as well as references to the experience of UN Women,	
2. Do evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved?	1	2. Evaluation criteria and evaluation questions The evaluation criteria and evaluation questions do not address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved.	
3. Have gender-responsive evaluation methodology, methods and tools, and data analysis techniques been selected?	2	3. The evaluation report explains how ethical standards were applied throughout the evaluation and that all stakeholder groups were treated with integrity and guaranteed protection and respect for confidentiality (p. 12). Nevertheless, the Annex 7 "Interview logbook" does not show evidence of application of the ethical standards. The Annex 9 "Online survey..." contains information with regards to confidentiality. 4. Annex 6 "The Evidence Table" shows that the evaluators assessed progress on GEEW wherever possible. They present GEEW-related data in the column "Evidence collected to respond to the evaluation question/dimension" (pp. 46 - 48). The evaluation conclusions reflect a gender analysis, for instance: "... the Innovation Fund has helped staff to reconnect and realign with the mandate of the organization in a tangible way by offering opportunities to bring about change for and with adolescents and women" (p. 44). Evaluation recommendations reflect a gender analysis, for instance: "...innovative solutions with a direct impact on the lives of women and young people"	
4. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?	2		

(*) 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).

(**) Scoring uses a four point scale (0-3).

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

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	22	58	20	
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: