

EQA for UNFPA Bosnia Herzegovina Country Programme Evaluation (2010-2013)



Title of Evaluation Report: UNFPA Country Programme Evaluation: Bosnia Herzegovina

Time-frame for the CPE stated in the ToR: 36 days

Cost of the CPE: \$47,122.82¹

Overall Assessment: The report is logically **structured**. However, while it includes the required sections and annexes and additional sections which contribute to the report’s quality, the clarity of reporting is slightly undermined by inconsistent writing quality, use of acronyms, and somewhat poorly-structured tables. The evaluation **methodology** is explained in detail, including methodological choice, but lacks simple stakeholder classification. **Findings** stem from rigorous data analysis. Some **conclusions** are non-specific and this section also includes recommendations. The **recommendations** flow logically from conclusions and are strategic and targeted. The report meets most of the requirements outlined in the **ToR**.

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
<p>I. Structure and Clarity of Reporting <i>To ensure report is user-friendly, comprehensive, logically structured and drafted in accordance with international standards.</i> Checklist of minimum content and sequence required for structure:</p> <ul style="list-style-type: none"> • i) Acronyms; ii) Exec Summary; iii) Introduction; iv) Methodology including Approach and Limitations; v) Context; vi) Findings/Analysis; vii) Conclusions; viii) Recommendations; ix) Transferable Lessons Learned (where applicable) • Minimum requirements for Annexes: ToRs; 	<p>Good The report has the minimum content and sequence required. However, some sections are missing or combined. It is the case that the conclusions and recommendations were combined under one chapter; however, the two sections within that chapter were clear and separate, so this did not affect readability. Also, regarding Transferable Lessons Learned (where applicable), there is no such a chapter in the evaluation report (or mentioned in the ToR). The table of contents is complete and is supported by a list of tables, diagrams and graphs. Annexes are numbered, making it easy for the reader to refer to them when prompted by the report. Whilst the annexes did not include a bibliography, sources and annotations were listed as footnotes in their associated sections. Overall, the report is user-friendly and clear.</p>			

¹ Source: Evaluation Office ‘Country Programme Evaluation Survey’ 2012-13

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
Bibliography; List of interviewees; Methodological instruments used.	The excessive use of acronyms did detract from the quality of the report by undermining clarity. Also, at times, the report did not include the full version of the term before employing an acronym in place of the full term.			
<p>2. Executive Summary <i>To provide an overview of the evaluation, written as a stand-alone section and presenting main results of the evaluation.</i> Structure (paragraph equates to half page max):</p> <ul style="list-style-type: none"> i) Purpose, including intended audience(s); ii) Objectives and Brief description of intervention (1 para); iii) Methodology (1 para); iv) Main Conclusions (1 para); v) Recommendations (1 para). Maximum length 3-4 page. 	<p>Poor The executive summary provides an overview of the evaluation and presents the main results. However, the executive summary was incomplete and at times unclear: it did not include a description of the actual intervention or a full description of the purpose and intended audience. Also, an executive summary should be able to serve as a ‘standalone’ and autonomous document, which this does not. The text is full of general phrases and many abbreviations. The executive summary refers to acronyms that have not yet been defined, which means that readers are dependent upon the main report itself for comprehension. These negative aspects detract from the quality of the report. The length of executive summary is between 3-4 pages, which is appropriate. Specific issues on sections:</p> <ul style="list-style-type: none"> i) Purpose did not include a description of the actual intervention or a full description of the purpose and intended audience; ii) The paragraph “Objectives and scope” clearly describes the subject of the evaluation, “this evaluation reviews the UNFPA BiH Country Programme Action Plan (CPAP) using two separate components”, but does not state clear objectives; iii) The Methodology section describes sampling methods, the collection of evaluation data, logic model, general approach for the analysis, and limitations. iv) Main Conclusions section includes overview of achieved results, strategic level overview, and programmatic-level achievements. The text is full of abbreviations, and references to the outcomes, outputs, and activities are not understandable, for instance, “Outcome I, Increased access to quality FP, and its associated Output I, have been partially achieved.” Output I 			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
	<p>does not have any text. The claim “partially achieved” is not criteria-based argument: the level of achievement is not specific (numbers or percentage).</p> <p>v) Recommendations section provides an overview of recommendations at strategic and programmatic level, including cross-cutting aspects. Recommendations include clear objectives and activities such as “UNFPA BiH should increase staffing to permit greater engagement at local level”. At the same time, the text is full of general phrases: greater engagement (number or percentage would be better), local level (name or other characteristics of agencies or stakeholders which report refers to would add clarity).</p>			
<p>3. Design and Methodology <i>To provide a clear explanation of the following elements/tools</i> Minimum content and sequence:</p> <ul style="list-style-type: none"> • Explanation of methodological choice, including constraints and limitations; • Techniques and Tools for data collection provided in a detailed manner; • Triangulation systematically applied throughout the evaluation; • Details of participatory stakeholders’ consultation process are provided; • Details on how cross-cutting issues (vulnerable groups, youth, gender, equality) were addressed in the design and the conduct of the evaluation. 	<p>Good</p> <p>The evaluation team provided a clear definition of the evaluation’s purpose, objectives and scope, and then provided a detailed explanation of the methodology. The evaluation report explains the methodological choice for data collection strategies. Details and examples of participatory stakeholder consultation are also provided.</p> <p>The authors acknowledge the main constraints and limitations of the study, namely that the small and non-random sampling leads to a non-representative and qualitative dataset. The authors explain that the site selection, though non-random, was purposive, resulting from a stakeholder consultation. It is not clear if the people interviewed are fully representative of the stakeholders who should be interviewed. The evaluation report acknowledges the poor response rate of the study, which amounted to only nine out of thirty-six beneficiary interviews being conducted. The authors also explained how the adverse timing of the study (in June school holidays) likely worsened the response rate for certain interview categories.</p> <p>Explanation of methodological choice is given in the text as “justification for the evaluation methods used.” The evaluation consultants used three criteria to choose “interviews” as the main method for data collection: “appropriate,</p>			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
	<p>economical and efficient given the time and resource constraints”. Nevertheless, there are no criteria in the text for the appropriateness.</p> <p>Techniques and Tools for data collection are provided in a detailed manner: desk review; site visits; interviews with stakeholders, UNFPA BiH programme clients, and former trainees. Quantifiable data were provided with regards to the number of persons interviewed and number of interview sessions (tables 3-5).</p> <p>Triangulation was systematically applied throughout the evaluation “desk review, the interviews (stakeholder, training follow-up and client/beneficiary), financial data, and other documentation” and is evident from Annex 5.</p> <p>Details of participatory stakeholders’ consultation process are provided in the section Stakeholder Involvement.</p> <p>Details on how cross-cutting issues (vulnerable groups, youth, gender, equality) were addressed in the design and the conduct of the evaluation is provided: “Meetings were also arranged with...women’s rights advocates, and youth representatives. These meetings ensured an opportunity for stakeholders to participate in the design of the evaluation.” However, the evaluation team provides a general number of people interviewed per programmatic component without details of how cross-cutting issues were addressed during the conduct of the evaluation, except for the following: “Because of the especially sensitive nature of their vulnerability, the evaluation team concluded that it was not feasible to obtain valid informed consent for victims of CRSV. For this reason, this group of client/beneficiaries was not interviewed”.</p>			
<p>4. Reliability of Data</p> <p><i>To clarify data collection processes and data quality</i></p> <ul style="list-style-type: none"> • Sources of qualitative and quantitative data have been identified; • Credibility of primary (e.g. interviews and focus groups) and secondary (e.g. reports) data established and limitations made explicit; 	<p>Poor</p> <p>Sources of qualitative and quantitative data have been identified and clearly explained in the Annex 2: Evaluation Site Visit Schedule, Annex 3: List of Persons Contacted, Annex 4: Data Collection Instruments, and Annex 5: Evaluation Matrix. Nevertheless, the annexes do not include any contact information (e.g. name of the department, website) which would be useful to ensure reliability of data.</p>			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
<ul style="list-style-type: none"> Disaggregated data by gender has been utilized where necessary. 	<p>In terms of credibility, the evaluation team also screened respondents for having a minimum knowledge of the UNFPA BiH CPAP, but also included a separate group of respondents reporting insufficient knowledge of the CPAP as they were deemed to provide valuable information regarding the context surrounding the UNFPA BiH CPAP. This method improves the reliability of data by screening out irrelevant responders, but also including those that yield insight into context. However, whilst the authors acknowledge that, although informed consent procedures are applied consistently and uniformly for all interviews, respondents participating in group interviews may not have been honest in their responses given that they were in the presence of their peers. The authors also acknowledged that respondents likely exhibited a bias towards providing positive responses so as to give a favorable impression of UNFPA activities in order to increase their chances of receiving continued funding. The authors' do not present how the evaluation team took account of this challenge to the validity of the interviews, other than to say that repeated attempts were made by the interviewer to probe responses during the data collection phase.</p> <p>Disaggregated data by gender has been utilized. For instance, Figure 3 describes the ratio of female to male labour force participation (in %) at the national level. However, whilst the authors point to the diverse profile of respondents, they do not disaggregate their own data by gender and it should have been made clear whether data is geographically disaggregated.</p>			
<p>5. Findings and Analysis <i>To ensure sound analysis and credible findings</i> <u>Findings</u></p> <ul style="list-style-type: none"> Findings stem from rigorous data analysis; Findings are substantiated by evidence; Findings are presented in a clear manner <p><u>Analysis</u></p>	<p>Good</p> <p>The report includes sound analysis (developed evaluation matrix, identified contextual factors, complex desk review, valid sampling, triangulation of information) which supports the credibility of findings.</p> <p>Findings stem from rigorous data analysis. Each focus area was reviewed for progress within their respective component activities; each activity was assessed based on the desk review criteria matrix, interview data and analysis of related financial information. Findings were validated across all data sources, with an</p>			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
<ul style="list-style-type: none"> • Interpretations are based on carefully described assumptions; • Contextual factors are identified. • Cause and effect links between an intervention and its end results (including unintended results) are explained. 	<p>attention to all of the evaluation criteria specified in the TOR (Annex 5: Evaluation Matrix). The analysis is based on triangulating information obtained from the desk review, the interviews (stakeholder, training follow-up and client/beneficiary), financial data, and other documentation.</p> <p>Findings are substantiated by evidence. For example, in the Section 3.3: The UNFPA programmatic response, the consultants provide evidence of UNFPA influence on their target audience. For instance, the establishment of the Parliamentary Group for Population in Development, working with ministries on a draft of a Programme for victims of wartime rape, sexual abuse and torture, and their families in Bosnia and Herzegovina 2013-2016.</p> <p>Interpretations are based on carefully described assumptions. Contextual factors are identified. Political, social, and economic contexts are considered, as well as the responses of the people that were interviewed. For instance, the report says “Based on the current policy context of the State and Entity level governments, as well as the current trends in sexual and reproductive health in BiH, including the adverse profile of methods of contraception use, especially a high reliance on withdrawal as a method of contraception, and an associated high prevalence of abortion, all of the proposed activities related to Output 1 and Outcome 2 are highly relevant,” “...on the basis of stakeholder interviews familiar with UNFPA SRH activities, as well as an assessment of the current context...”</p> <p>In order to demonstrate intervention logic, every section has a logic model for a focus area describing one outcome, one output and one activity (Figures 13, 15, 17). There are relevant figures to explain Budget and Expenditure distribution. In spite of this, it is not easy to identify and analyze cause and effect links between an intervention and its end results (including unintended results). The evaluation consultants often refer to outcomes, outputs, and activities without providing a quote and number of the activity according to the logic model they refer to. For instance, it is not clear from the text which activities they refer to “the proposed activity for cervical cancer screening is highly relevant... Among the activities for Outcome I, the most impressive by far has been related to Cervical Cancer (CC)</p>			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
	<p>screening.”</p> <p>The cause and effect links between activities and outputs, and outputs and outcomes, are discussed. A clear Logic Model is presented that explains the theoretical link between activities, outputs and outcomes. This programme theory is broken down by outcome, which aids clarity. Examples are given of the 'effectiveness' of given activities in terms of their contribution to outcome measures. Examples and explanations are also given of outcomes that were not achieved. External factors/contexts and internal limitations are discussed with respect to whether outcomes were achieved or not. However, unintended consequences are not discussed or explained by the report.</p>			
<p>6. Conclusions</p> <p><i>To assess the validity of conclusions</i></p> <ul style="list-style-type: none"> • Conclusions are based on credible findings; • Conclusions are organized in priority order; • Conclusions must convey evaluators' unbiased judgment of the intervention. 	<p>Poor</p> <p>Conclusions are based on credible findings, referencing the analysis of findings from previous chapters, and flow logically from the general results. Moreover, the conclusions do not convey a particular bias or tendency of the evaluators towards the intervention. The conclusions present both the negatives and positives associated with the intervention.</p> <p>However, sub-section “Overview of achieved results” presents the general opinion of the consultants expressed in non-specific terms e.g. “there is still a feeling that UNFPA office missed opportunities...” (opportunities are not clarified), “Many of the activities appear to have been “one-off.” This sub-section also includes recommendations: “UNFPA needs to give more focus on reaching for and establishing key partnerships at all levels.”</p> <p>Strategic level conclusions present an overview of the context that adds clarity, but the section is written as recommendations: “the UNFPA BiH needs to be funded and staffed... The UNFPA BiH needs to work more on strengthening capacity.” Key conclusions are presented after some recommendations: “The UNFPA BiH CPAP has fallen into a pattern of implementing short-term activities, such as roundtables and regional conferences, without a planned strategic sequence and follow up...”. Numbering conclusions would add clarity and order as</p>			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
	<p>conclusions are presented as a discussion. It is not clear if conclusions are organized in priority order.</p> <p>Programmatic level conclusions highlight activities and key factors which led to success and failure, including social, political and economical factors e.g. “Lost momentum in SRH and Youth SRH Counselling and Services”.</p>			
<p>7. Recommendations</p> <p><i>To assess the usefulness and clarity of recommendations</i></p> <ul style="list-style-type: none"> • Recommendations flow logically from conclusions; • Recommendations must be strategic, targeted and operationally-feasible; • Recommendations must take into account stakeholders’ consultations whilst remaining impartial; • Recommendations should be presented in priority order 	<p>Good</p> <p>Recommendations flow logically from conclusions. Recommendations are also strategic, targeted, sensible, operationally-feasible and realistic in nature. Stakeholder consultation is suggested within recommendations at various levels. Recommendation are numbered, presented in priority order and include activities and objectives which may be accomplished by these activities, for instance, “Increase staffing to permit engagement at local level.”</p> <p>Most recommendations are strategic, targeted and operationally-feasible. For instance, the first recommendation says that “UNFPA BiH needs to include implementing partners early on in the design of any proposed CPAP 2 project activities. This includes the establishment of a regular UNFPA presence in both Entities at minimum on a quarterly basis.” Another example of operationally-feasible recommendation is “UNFPA should not continue to invite elderly and youth to roundtables and other events unless UNFPA is a) serious about providing a genuine role for them at the event and b) has concrete plans for useful and compelling follow-up activities.”</p> <p>Some recommendations need to be more specific. The report does not clarify stakeholders they refer to at State, Entity and local levels. Therefore, the phrase “UNFPA BiH needs to move beyond State and Entity level to local-level engagement” is unclear, as is the phrase “Future work on RMs should have stronger focus on ensuring creation of innovative solutions within the framework of minimal standards for referral mechanisms.”</p> <p>Recommendations are not presented in priority order but in the same order as conclusions: 'Overview of Achieved Results', 'Strategic Level', Programmatic Level',</p>			

Quality Assessment criteria	Assessment Levels			
	Very good	Good	Poor	Unsatisfactory
	and 'Cross Cutting Aspects' and broken down further by focus area.			
<p>8. Meeting Needs To ensure that Evaluation Report responds to requirements (scope & evaluation questions/issues/DAC criteria) stated in the ToR (ToR must be annexed to the report). In the event that the ToR do not conform with commonly agreed quality standards, assess if evaluators have highlighted the deficiencies with the ToR.</p>	<p>Good The report responds to the requirements in the ToR. Moreover, the report responds to weaknesses of the ToR and made efforts to correct these deficiencies. For example, the evaluation team notes that redundant or duplicate evaluation questions have been omitted from consideration. Annex 5: Evaluation Matrix is a useful way to address all the questions in the ToR.</p>			

Quality assessment criteria (and Multiplying factor *)	Assessment Levels (*)			
	Very good	Good	Poor	Unsatisfactory
1. Structure and clarity of reporting (2)		2		
2. Executive summary (2)			2	
3. Design and methodology (5)		5		
4. Reliability of data (5)			5	
5. Findings and analysis (50)		50		
6. Conclusions (12)			12	
7. Recommendations (12)		12		
8. Meeting needs (12)		12		
TOTAL		81	7	

(*) Insert the multiplying factor associated with the criteria in the corresponding column e.g. - if “Finding and Analysis” has been assessed as “good”, please enter the number 50 into the “Good” column. The Assessment level scoring the higher number of points will determine the overall quality of the Report

OVERALL QUALITY OF REPORT: Good