

# Independent Quality Assessment of UNODC Evaluation Reports 2016

## **Synthesis Report**

Prepared for:

Independent Evaluation Unit, UNODC

by John Mathiason and Ann Sutherland

December 22, 2016

## CONTENTS

INTRODUCTION .....	3
METHODOLOGY.....	3
FINDINGS .....	4
A. Strengths of the Evaluation Reports .....	5
B. Weaknesses of the Evaluation Reports .....	6
C. Improvements from First Draft in Final .....	7
D. Best Practices Observed.....	8
E. Challenges for Evaluators .....	9
F. Mainstreaming Human Rights and Gender Considerations.....	9
RECOMMENDATIONS.....	11
Appendix 1. Comparison of Final Report with Draft Report	
Appendix 2. UNODC EQA Template as Used in the Review	
Appendix 3. Recommended Revisions to UNODC EQA Template	

## INTRODUCTION

The Independent Evaluation Unit (IEU) is leading and guiding evaluations in order to provide objective information on the performance of United Nations Office on Drugs and Crime (UNODC). As a member of the United Nations Evaluation Group (UNEG), IEU is following its Norms and Standards.

The number of independent project evaluations and in-depth evaluations has increased significantly over the past years at UNODC, generating important learning opportunities for all stakeholders, including the IEU. UNODC's efforts to create better mechanisms for discussing and propelling strategic issues forward are being informed by IEU, whereby important insights are generated in the planning, implementation and finalization of evaluations. As part of its efforts to ensure the agency's evaluations are providing credible information to inform planning processes, the IEU has commissioned independent evaluation quality assessments of evaluation reports produced since 2014.

This document synthesizes the Evaluation Quality Assessment (EQA) of all published evaluation reports in 2016 and makes comparisons with EQAs from 2014 and 2015. Moreover, the quality of a random sample of 30% of first draft evaluation reports has been reviewed and assessed to analyze the qualitative changes between the first draft report and the published version.

This assignment was carried out from mid-November to mid-December of 2016 by two independent consultants - Dr. John Mathiason (Team Leader) and Ann Sutherland (Team Member). Both have extensive experience in conducting evaluations and meta-evaluations for international organizations. They are the Managing Director and Principal Associate, respectively, for Associates for International Management Services (AIMS).

## METHODOLOGY

The first phase of this assignment involved making minor revisions to the EQA template. Check boxes for rating the extent to which each criterion was met (met, partially met, not met) were incorporated into the template. This was found to be advantageous in several respects: the comments are less redundant as it is no longer necessary to re-state the criteria; the reviewer is forced to consider all elements; and the comments section is used only to highlight issues that are particularly strong or weak, and provide any additional justification that may be needed to explain the overall score given for the respective section. The revised EQA template used for the assessment is shown in the Annex.

The reviewers then examined the quality of all of the evaluation reports published during 2016. The total number of evaluations for the year was 19. Three were in-depth evaluations and 16 were independent evaluations. The reviewers also compared the final drafts of a third of the evaluations, randomly selected, with the first drafts to observe any differences that could be attributed to the process of commenting on the first drafts by IEU. All were drafts of independent evaluations.

To ensure consistency of the review process, both reviewers independently assessed one In-Depth and three Independent Project Evaluation reports. These reports were also randomly selected and were rated according to the nine EQA criteria categories. The reviewers then compared their comments and scores for each criterion as well as the overall score. In all cases, the overall scores were the same for both reviewers. There were minor differences in criteria scores and there were non-material differences in comments. The differences were discussed and resolved. As it was evident that there was consistency between the team, the remaining reports were each rated by one person; assignments were based on each reviewer’s area of expertise and language skills. Overall, approximately one third of the evaluation reports were assessed by both reviewers.

The team used the “Unsatisfactory” rating only when the criteria elements were missing or very poorly addressed. Similarly, “Very Good” was only used when all criteria were fully met. As a result, “Fair” was used when the criteria were generally not met, and therefore that score can be understood to mean that the reports, or pieces of the reports, were not well done.

## FINDINGS

All of the 19 reports produced in 2016 were of adequate quality. As seen in Table 1, four reports received a Very Good rating and no reports were rated as Unsatisfactory.

**Table 1: Overall Rating of 2016 Reports compared with previous EQA cycles**

	Unsatisfactory	Fair	Good	Very Good	Total
# of Reports - 2016	0	8	7	4	19
% of Reports - 2016	0	42%	37%	21%	100%
# of Reports - 2015 <sup>1</sup>	0	12	9	1	22
% of Reports - 2015	0%	53%	41%	5%	100%
# of Reports - 2014/15 <sup>2</sup>	0	9	22	2	33
% of Reports - 2014/15	0%	27%	67%	6%	100%

Table 1 also shows there have been substantial improvements compared to 2015 with an increased number of Very Good reports and fewer rated as Fair. The drop in ratings for the latter half of 2015 compared to those of 2014-mid 2015 may be explained by changes made to the EQA template, in particular to the addition of a criteria of Reliability and an increased emphasis on gender analysis.

In 2016 there were, as last year, some differences by criterion. Table 2 shows that Executive Summary and Reliability sections were the most likely to have negative (Fair and Unsatisfactory) ratings. Recommendations, Lessons Learned and GEEW were also

<sup>1</sup> This included reports published from June through December 2015.

<sup>2</sup> This included the first batch of EQA assessments which considered reports published from January 2014 through May 2015.

less strong. The Conclusions and Presentation/Structure sections were the most likely to have positive ratings.

**Table 2: Report Rating by Criteria**

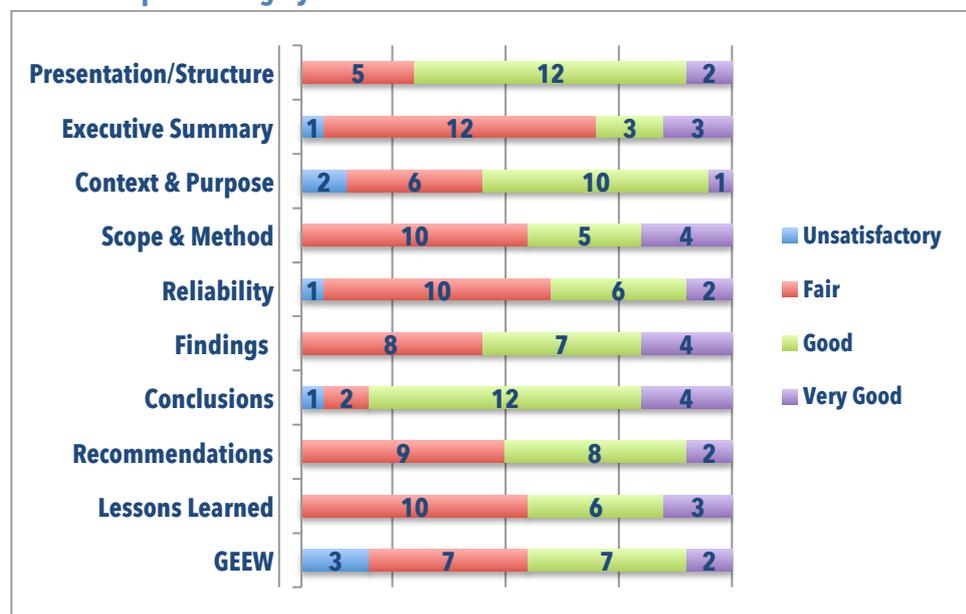
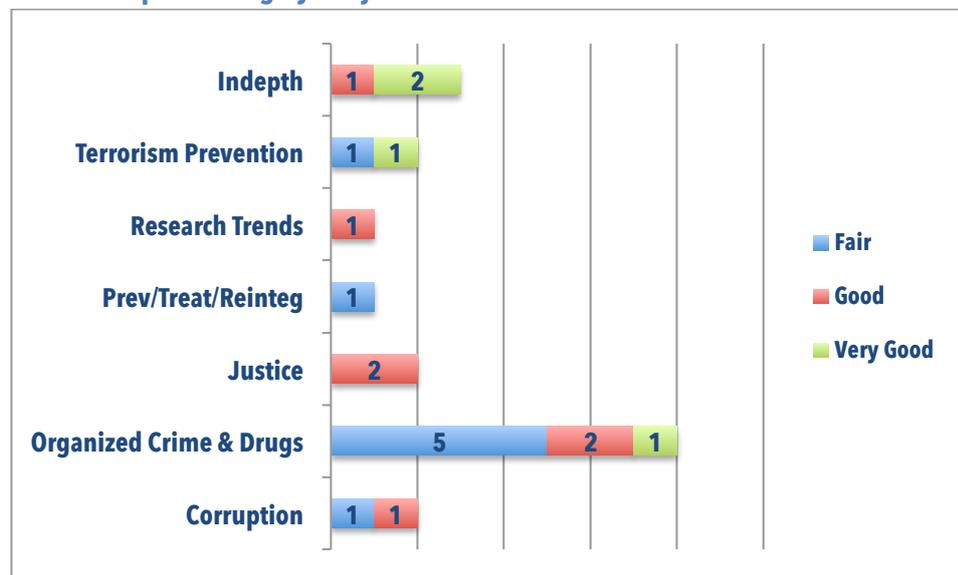


Table 3 shows the scores by subject area. Due to the low numbers of reports, there is not a specific sector that can be said to have stronger reports. However it is notable that two of the In-depth reports were rated as Very Good and one as Good. This is an improvement from 2015 where, of the eight In-depth reports, 25% were rated as Fair and 75% were rated as Good.

**Table 3: Report Rating by Subject**



### A. Strengths of the Evaluation Reports

As mentioned above, the evaluation reports for 2016 showed improvement with 58% receiving a rating of Good or Very Good. There were four primary strengths observed across those reviewed:

**Consistency with DAC and UNEG Norms:** As was the case for 2015, the reports generally conformed to the accepted norms and guidelines.

**Conclusions:** The conclusions section was the best in meeting criteria, showing that the evaluators were able to synthesize the results effectively in 16 of 19 cases (84% were rated as Good or Very Good).

**Presentation and Structure:** This section of the reports received the second-highest score with 14 of 17 reports receiving a rating of Very Good or Good.

**Inclusion of Gender Analysis:** There were notable improvements in the 2016 reports and these are discussed in the “Mainstreaming of Human Rights and Gender” section below.

## B. Weaknesses of the Evaluation Reports

There continue to be areas in which UNODC evaluations could improve upon overall. These include:

**Executive Summaries:** This was the weakest section of the 2016 evaluation reports. The main issues were that the section often (a) exceeds the maximum page length, (b) does not include reference to the evaluation methodology used, (c) is overly focused on findings rather than conclusions, and (d) the Summary Matrixes include too many and/or unclear recommendations.

**Reliability of Data:** There are a number of good practices in the field of development evaluation that are not sufficiently used in the reports reviewed. Common are:

- The absence of sampling frames to address reliability of data. Without this, it is not possible to know if the data are representative.
- The sources of data from stakeholders consulted during the evaluation process not being evident in the Findings. In approximately one third of reports, the findings are presented without reference to where the perspectives came from or with only very general reference to source (i.e. “interviewees said . . .”). This can leave the impression that the Findings are solely or mostly from the perspective of the evaluator.
- Responses not being broken down by stakeholder group or by gender.
- Data that was collected but not reported on. In a number of cases, the methodology included data collection processes such as surveys or Most Significant Change (MSC) questions but there was no specific reference to these in the Findings. BOL/Y15 is a case in point whereby the methodology included interviews with 467 students, however none of the data from this exercise was shown. In such instances, it is unknown to the reader whether the data was not analyzed or whether it was not useful.
- Lack of quantitative data collected during the evaluation (primary source).
- The overuse of data from document review (secondary source).
- No, or overly vague, explanations for how data was analyzed which suggests that data is not being reviewed in a systematic manner.

**Absence of Logframes:** As was the case for 2015, the most significant concern is the absence of results frameworks and clear indicators to measure the progress or achievement of the intervention against its planned results. Logframes were rarely included in the reports but in four cases the evaluators prepared an improved framework (and these were evaluation reports rated Very Good). Logframes should be mandatory components of evaluations since they are intended to express what has been promised in terms of outcomes and objectives to be achieved. If evaluators find existing logframes inadequate, then there should be some attempt to set one up in order to be clear about what the evaluation process is measuring. At a minimum, the evaluators should articulate the proposed chain of results or program theory. This practice should be suggested in the inception reports.

**Inadequate attention to Effectiveness:** Effectiveness should be the main summative focus of the Findings. Often, more attention is given to at least one or more of Relevance, Design, Efficiency, and Impact (when incorrectly used). While all are important to comment on, these should be less of a focus than progress towards achieving results of the intervention.

**Varied understandings of Effectiveness and Impact:** Frequently, the Effectiveness section is used to present the activities that have been carried out without mentioning or linking them to the outcomes they are expected to contribute towards. This happens even though, in most cases, the main evaluation questions in the ToRs are explicit about the need for the Effectiveness section to focus on outcomes. An additional concern is that the Impact section is frequently pitched at the level of outcome results, and sometimes at the output level (i.e. increased knowledge gained from training workshops). If the project has not yet had the opportunity to show substantial progress towards results, in most cases it would be sufficient for the evaluator to just state that in the Impact section.

**Other common shortcomings:** These include lack of consultation with, or involvement of, CLPs or other stakeholders beyond their being interviewed; not providing information about evaluators and their suitability for the assignment (in one case the name of the evaluator was not provided); not providing the dates/timeframe of the evaluation process; minimal use of visual aids; and, using interview protocols that are not adapted for specific stakeholder groups.

### C. Improvements from First Draft in Final

As requested, the reviewers compared the final reports of just over one third (seven) of the evaluations with their first drafts on which IEU will have commented. These were randomly selected from the first batch of 17 evaluations received. The results of the comparison are shown in Table 4 and the specific analysis is shown in Annex 1.

**Table 4. Comparison of First and Final Drafts of a Sample of Evaluations**

Evaluation	Date of first draft and final draft	Substantive Changes	Rating
MEXZ44 Preventing Corruption Mexico	October 2015 October 2015	Few	Fair
XAWU72 Airport Communications Project Final	July 2016	Many	Good

	September 2016		
LAOX26 Criminal Justice Responses to Human Trafficking in Lao	Aug/Sept 2016 September 2016	Many	Very Good
PERG34 Proyecto Sistema de Monitoreo de Cultivos Ilícitos SIMCI	July 2015 February 2016	Some	Fair
BRAX16 Expressive Youth - Citizenship, Access to Justice, Peace Brazil	November 2014 June 2015 (But issued in 2016)	Some	Good
PSEX02 Forensic Human Resource & Governance Palestine Midterm	February 2016 June 2016	Some	Good
XAPX37 Counter-Terrorism East and Southeast Asia Partnership Criminal Justice Responses	April 2016 June 2016	Few	Very Good

The final documents showed general improvement from the draft versions. Although it did not appear that the changes would have affected the overall EQA ratings of any of the reports, there were several cases where the category scores would have improved, particularly GEEW scores. There was considerable variation in the extent of changes between the draft and final versions. Of the seven reports, two had minimal changes, two had substantial changes and three were in between. In the two reports that were in Spanish, the main difference was that the Executive Summary (and summary matrix) was translated into English. Overall, the revisions were mostly helpful in one or more of the following ways: increasing readability through editing and formatting changes, providing more evidence to back up findings, disaggregating the list of stakeholders consulted by gender, including more gender analysis, and reformulating the Conclusions. In some instances the revisions were not beneficial, most notably a report where the Executive Summary was increased from four to seven pages.

There were substantial differences in time between the first draft and the publication of the final version of the report. The range was from the same month to over a year.

#### D. Best Practices Observed

**Revisions to logical frameworks:** Where the evaluators reviewed and then, when necessary, revised the logical framework of the evaluation, they were able to obtain data that show outcomes and give an initial sense of impact, or indicate where data could not be obtained. The good logical frameworks showed clearly the expected causal connection of output with outcomes. The two In-depth evaluations rated as Very Good provide good examples of this practice.

**Evaluation Matrix:** The matrix that is included in the report for *LAO X26 Strengthening Criminal justice Responses to Human Trafficking* clearly shows the indicators/benchmarks and findings in relation to each of the evaluation questions.

**Robust methodology:** *RASH13 Prevention of transmission of HIV Among Drug Users* is exemplary for its strong methodology which is able to show effectiveness, relevance and sustainability.

**Treatment of findings:** *Cluster Evaluation of SMART (GLOJ88) & Forensic (GLOU54)* has strong presentation of Effectiveness, Efficiency, Impact, and Partnerships.

**Presentation of Evaluation Tools:** *TKMX57 Strengthening Customs Service in Turkmenistan* is an exemplary report for the inclusion of evaluation tools tailored to different stakeholder groups.

### E. Challenges for Evaluators

Although some progress has been made in this regard, the ToRs mostly continue to have too many questions to be answered by the evaluations. The stronger evaluations are generally those where the evaluators have narrowed down the number of questions.

### F. Mainstreaming Human Rights and Gender Considerations

As in 2015, the evaluations this year consistently included dedicated sub-sections within the Findings for Human Rights and Gender.

Table 5 shows that the average scores for considering GEEW have improved according to the UN-SWAP criteria. The SWAP tool assesses the extent to which gender equality and the empowerment of women (GEEW) is integrated into evaluation processes. There are four criteria and each is rated on a scale of 0-3 with 0 being awarded when there is no integration, 1 when there is partial integration, 2 when there is satisfactory integration, and 3 when the evaluation process exceeds requirements. The overall scores for 2016 would have been even higher if it was not for one report that received an Unsatisfactory rating.

**Table 5: Average scores for the integration of GEEW (UN-SWAP)**

Quality Assessment Criteria	Average Score (0-3)	
	2015	2016
a. GEEW is integrated in the evaluation scope of analysis and indicators are designed in a way that ensures GEEW-related data will be collected.	1.125	1.63
b. Evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved.	1.625	1.63
c. Gender-responsive evaluation methodology, methods and tools, and data analysis techniques are selected.	.875	1.37
d. Evaluation findings, conclusions and recommendations reflect a gender analysis.	1.875	1.84
<b>Overall Score</b>	<b>5.5</b>	<b>6.53</b>
<b>Overall Rating</b>	<b>Fair</b>	<b>Fair</b>

Reports that were particularly strong in analysing gender issues included:

**PSEX02 Forensic Human Resource & Governance Palestine Midterm** – this is a good example of an analysis that goes beyond looking at representation of women in events and career positions to look at broader issues of GEEW.

**XAPX37 Counter-Terrorism East and Southeast Asia Partnership Criminal Justice Responses** and **LAOX26 Criminal Justice Responses to Human Trafficking in Lao** were both exemplary for the explanation in the methodology section of how gender was considered. The latter report referenced the UN Women’s handbook, “How to Manage A Gender-Responsive Evaluation” and used this guidance to structure the evaluation approach (see Figure 1).

**Figure 1: Excerpt of Methodology Section from LAOX26 Evaluation Report**

The UN Women’s handbook ‘How To Manage A Gender-Responsive Evaluation provided a useful context for the approach the evaluation team took to the evaluation. The following excerpt summarises this approach:

Gender-responsive evaluation has two essential elements: what the evaluation examines and how it is undertaken. It assesses the degree to which gender and power relationships—including structural and other causes that give rise to inequities, discrimination and unfair power relations, change as a result of an intervention using a process that is inclusive, participatory and respectful of all stakeholders (rights holders and duty bearers).<sup>4</sup>

What did the evaluation examine? While assessing the project in the context of its design, it was noted that the project has a specific focus on addressing, with *authorities* and in terms of *victims and victim support*, gender-based violence in the form of human trafficking. The focus here was on the gender-focus and gender-balance in the activities and outputs/ outcomes of the project, with an emphasis on the structural aspects. Enquiry also focused on societal inequities and what, if anything, is shifting in perceptions about and provision of services/ support to victims of trafficking.

How was the evaluation undertaken? The evaluation was gender sensitive in its approach. A focus of the inception and field phases was paying particular attention to ensuring a balance of perspectives (victim, victim’s family, support organisation/ activist organisation and authority) when planning how to enquire into the effectiveness of project activities and the effectiveness of results. The evaluation team emphasised the importance, to the evaluation, of the responses of victims and victim groups, working to ensure that these target groups were included in field processes. It is noted however that no discussion/ enquiry was actually undertaken with victims, due to actual project implementation status. Irrespective of project design, there has been no emphasis to date on specific work with victims. The evaluation team sought to understand the issues, the project approaches and the outcomes from a perspective beyond that of authorities (municipal authorities, prosecutors, police). As victims were not part of enquiry, the evaluation also sought the perspective of support services, NGOs and community representatives in order to understand more than structural, legislative and policy questions.

## RECOMMENDATIONS

The main recommendations for the IEU to consider are:

**Updated guidance for managers of evaluation processes and for evaluators:** It is suggested that the review and revision of UNODC's guidance for evaluation processes take into account the recommendations made in the EQA Synthesis Reports for both 2014 and 2015 as the issues raised continue to be relevant. The guidance should include the expectations for reviewing draft reports and should be available in Spanish and French.

Additional recommendations beyond those in past reports are:

**More emphasis on robust methodology:** Evaluators should be including qualitative data collection and analysis processes. If surveys are not feasible, interview protocols and other tools should be designed to elicit at least some responses that can be more easily quantified through content analysis processes. Evaluators also need to be transparent about how they have analyzed the data collected through both quantitative and qualitative processes. Additionally, evaluators should be disaggregating responses by stakeholder groups and by gender. It is unlikely that all stakeholders hold the same views and the varying perspectives should be illuminated in the Findings.

**Further refinement of the EQA template:** Further refinement of the template is suggested. The new design with a box to rate each criteria element has shown to be helpful but its use has revealed some redundancies to be addressed. The recommended changes are shown in Annex 3.

## Appendix 1. Comparison of Final Report with Draft Report

### MEXZ44

Added English-language Executive Summary and Summary Matrix. Very minor editorial changes.

### XAWU72

There were extensive changes between the versions – most but not all were improvements. The Executive Summary was increased from 4 to 7 pages, in part due to an overly detailed description of the project and changes, and the excessive number of footnotes was increased. The ES improvements included criteria subheadings for the findings/conclusions, and a note about the role of the CLP. The text in the Summary Matrix is somewhat more clear and concise. In the Background section, a description of the project was removed which negatively affected the EQA rating for that section. In the Methodology, Conclusions and Recommendations sections references to HR and Gender were added, and in the annexed list of stakeholders, the numbers interviewed were gender disaggregated. Text was adjusted throughout the Findings – mostly for clarification or additional evidence. There was a substantial amount of editing done (mostly minor issues), long paragraphs were broken up and some were rearranged. The Conclusions were completely changed to include all criteria. One lessons was moved to become a conclusion. There were a number of formatting changes through out the document – most helpful but in a couple of places maps/figures ended up covering text which wasn't the case in the earlier draft.

### LAOX26

There were a number of changes that improved the final version, and the GEEW score. The final version noted that the consultants were independent and one was a substantive expert. Text of Executive Summary not included in draft. Summary Matrix adjusted in final version to include to whom recommendations directed to and more in evidence column, and one recommendation (and corresponding lesson) about handing back victims of trafficking was dropped. In several cases direct excerpts from documents (some quite lengthy) are included in the draft but were removed or shortened and the main point from those were integrated into the main text. This was an improvement. More specificity was added about institutions that were interviewed and part of the Project Steering Committee. The full set of evaluation questions (41) was removed from the methodology section and reference was made to where they could be found in the annex. The Efficiency section was increased to include more financial information (budget/outcome and spending rate), more analysis of the project logframe, and more evidence to support the shortcomings identified. The Effectiveness section included additional information about the outcomes of the training. The Conclusions section, which was organized by criteria, originally did not address Design or Human Rights & Gender. There was minor re-wording of the Recommendations. The names of people consulted during the evaluations were removed from the listing in the annex and added was the gender disaggregation of this group.

## PERG34

Added English-language Executive Summary and Summary Matrix. Moved a table on Evaluación de Hallazgos from the Findings section to an annex. Added a recommendation for UNODC Vienna: “ 4. Se recomienda a la sede en Viena establecer un programa de entrenamiento de nuevas tecnologías, dirigido al personal técnico del proyecto, entre otros, en el uso y manejo de productos satelitales de última generación, en los nuevos sistemas de medición de la extensión, producción y definir su aplicabilidad en el país.” Added five recommendations for the project team and edited the results shown in the Summary Matrix. Added a table on budgetary contributions by source. Added a paragraph in the methodology section that described limitations in the interviews that were made. Added material in the Findings dealing with design, relevance, efficiency, partnerships and effectiveness were all increased in the final. There were no changes in the conclusions or recommendations.

## BRAX16

The first draft did not have an executive summary. The final draft translated the front page from Portuguese to English. Annexes were added to the final draft.

## PSEX02

There were minor changes throughout most of the document. The font used in the draft was according to the Guidelines but this was changed in the final. Minor revisions included moving one recommendation from ‘important’ to ‘key’ in the Summary Matrix and noting at the beginning of the Recommendations section who the majority of recommendations were directed towards. More substantial editing changes included moving portions of the text to the footnotes (footnotes were already used rather frequently), gender disaggregation of the list of stakeholders, adding more explanatory details in some cases, and making some paragraphs/sections more succinct. The latter was particularly evident in the Efficiency section where the financial discussion was shortened and the lengthy subsection on “Reasons explicating Efficiency” was reduced from 3.5 to 2 pages. Although most of the editing changes appeared helpful, in a few cases reference to the perspectives of stakeholders were removed and it was noted in the EQA that this was a shortcoming of the report. The draft just included the ToR of the lead evaluator but the final version included the full ToR (with ToRs for both consultants), a practice that is noted in the EQA Summary Reports as being redundant.

## XAPX37

There were some editorial changes (including making some statements less definitive). The section on human rights and gender equality was expanded. Text on partnerships was moved up ahead of effectiveness in the findings.

## Appendix 2. UNODC EQA Template as Used in the Review

General Project Information	
Project/Programme Number and Name	
Thematic Area	
Geographic Area (Region, Country)	
Approved budget of the time of the evaluation (USD)	
Type of Evaluation (In-Depth/Independent Project; final/ midterm; other)	
Evaluator(s)	
Date of Evaluation (from MM/YYYY to MM/YYYY)	
Date of Evaluation Report (MM/YYYY)	
Quality Assessment conducted on/by	

### OVERALL QUALITY RATING:

### SUMMARY:

Quality Assessment Criteria	Assessment Levels: Very Good - Good - Fair - Unsatisfactory	
	Meets Criteria: Y = Yes N = No P = Partially	
<b>1. Structure, Completeness And Clarity Of Report</b>	<b>RATING:</b>	
a. Format (headings, font) accords to IEU Guidelines and Templates for Evaluation		

Reports.	
b. Structure accords to IEU Guidelines for Evaluation Reports with the following sequence: Executive Summary; Summary Matrix of Findings, Evidence and Recommendations; Introduction (Background and Context, Evaluation Scope and Methodology, Limitations to the Evaluation); Findings (Design, Relevance, Efficiency, Partnership and Cooperation, Effectiveness, Impact, Sustainability, Human Rights and Gender Equality/mainstreaming, Innovation); Conclusions; Recommendations; Lessons Learned.	
c. Language is empowering and inclusive avoiding gender, heterosexual, age, cultural and religious bias, among others.	
d. Report is easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience). Visual aids, such as maps and graphs, are used to convey key information. List of acronyms is included.	
e. Report is free from any grammar, spelling, or punctuation errors.	
f. Objectives stated in the terms of reference are adequately addressed.	
g. Issues of human rights and gender equality/mainstreaming are adequately addressed	
h. Report contains a logical sequence: evidence-assessment-findings-conclusions-recommendations.	
i. Composition of Evaluation Team is included and has gender and geographic expertise. Preferably it is gender balanced and includes professionals from countries or regions concerned.	
j. Annexes include at a minimum: evaluation terms of reference; list of persons interviewed and sites visited; list of documents consulted; evaluation tools used.	
<b>2. Executive Summary</b>	<b>RATING:</b>
a. Written as a stand-alone section that provides an overview of the evaluation and presents its main results.	
b. Generally follows the structure of: i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology); iv) Main Conclusions; v) Recommendations.	
c. Summary Matrix presents only the key and most important recommendations from evaluation report.	
d. Findings, sources and recommendations in the Summary Matrix are clear and cohesive, and specify the stakeholder to whom they are addressed.	

e. Maximum length 4 pages, excluding the Summary Matrix.		
<b>3. Evaluation Context And Purpose</b>		<b>RATING:</b>
a. Clear description of the project evaluated is presented.		
b. Logic model and/or the expected results chain, and /or program theory (that at a minimum identifies and links objectives, outcomes and indicators of the project) is clearly described		
c. Context of key cultural, gender related, social, political, economic, demographic, and institutional factors are described, and the key stakeholders involved in the project implementation and their roles are identified.		
d. Project's status is described including its phase of implementation and any significant changes (e.g. strategies, logical frameworks) that have occurred.		
e. Purpose of the evaluation is clearly defined, including why it was needed at that point in time, who needed the information, what information is needed, how the information will be used, and the target audience.		
<b>4. Scope And Methodology</b>		<b>RATING:</b>
a. Evaluation scope is clearly explained including the main evaluation criteria, questions and justification of what the evaluation did and did not cover.		
b. Transparent description presented of methodology applied; how it was designed to address the evaluation purpose, objectives, questions and criteria is explained.		
c. Methodology allows for drawing causal connections between output and expected outcomes		
d. Gender sensitive methodology aware of power relations during an evaluation process, inclusive and participatory.		
e. Data collection methods and analysis, and data sources are clearly described; as are the rationale for selecting them, and their limitations are clearly described. Reference indicators and benchmarks are included where relevant.		
f. Sampling frame clearly described and includes area and population to be represented, rationale for selection, mechanics of selection including whether random, numbers selected out of potential subjects, and limitations of sample.		
g. Methods are appropriate for analysing gender equality/mainstreaming and human rights issues identified in evaluation scope		
h. High degree of participation of internal and external stakeholders, including the Core		

<p>Learning Partners, throughout the evaluation process is planned for and made explicit. When there are thematic or approach gaps (i.e. gender equality/mainstreaming) among stakeholders, other key informants not directly involved in the project were invited for consultation.</p>		
<p><b>5. Reliability of Data</b> <i>To ensure quality of data and robust data collection processes</i></p>		<p><b>RATING:</b></p>
<p>a. Triangulation principles (using multiple sources of data and methods) were applied to validate findings.</p>		
<p>b. Qualitative and quantitative data sources were used, and included the range of stakeholder groups and additional key informants (when necessary) defined in evaluation scope.</p>		
<p>c. Limitations that emerged in primary and secondary data sources and collection processes (bias, data gaps, etc.) are identified and, if relevant, actions taken to minimize such issues are explained.</p>		
<p>d. Evidence provided of how data was collected with a sensitivity to issues of discrimination and other ethical considerations.</p>		
<p>e. Adequate disaggregation of data by relevant stakeholder undertaken (gender, ethnicity, age, under-represented groups, etc.). If this has not been possible, it is explained.</p>		
<p><b>6. FINDINGS AND ANALYSIS</b> <i>To ensure sound analysis and credible findings</i></p>		<p><b>RATING:</b></p>
<p><i>Findings</i></p>	<p>-</p>	
<p>a. Have been formulated clearly, take into account any identified benchmarks, and are based on rigorous analysis of the data collected.</p>		
<p>b. Address all evaluation criteria and questions raised in the ToR including relevance, efficiency, effectiveness, impact and sustainability, as well as UNODC's additional criteria of design, partnership and cooperation, innovation, and the cross-cutting themes of human rights and gender.</p>		
<p>c. Address any limitations or gaps in the evidence and discuss any impacts on responding to evaluation questions raised in ToR.</p>		
<p>d. Discuss any variances between planned and actual results of the project (in terms of objectives, outcomes, outputs).</p>		
<p>e. Are presented in a clear manner.</p>		

<i>Analysis</i>	-	
a. Interpretations are based on carefully described assumptions.		
b. Contextual factors are identified (including reasons for accomplishments and failures, and continuing constraints).		
c. Cause and effect links between an intervention and its end results (including unintended results) are explained.		
d. Includes substantive analysis of gender equality/mainstreaming issues		
e. Includes substantive analysis of human rights issues.		
<b>7. CONCLUSIONS</b>		<b>RATING:</b>
a. Take into consideration all evaluation criteria and questions, including human rights and gender equality/mainstreaming criteria.		
b. Have been formulated clearly and are based on findings and substantiated by evidence collected.		
c. Convey the evaluators' unbiased judgement of the intervention		
d. Developed with the involvement of relevant stakeholders.		
e. Present a comprehensive picture of both the strengths and weaknesses of the project.		
f. Go beyond the findings and provide a thorough understanding of the underlying issues of the project and add value to the findings.		
<b>8. RECOMMENDATIONS</b>		<b>RATING:</b>
a. Are clearly formulated, based on the conclusions, and substantiated by evidence collected.		
b. Address flaws, if any, in project's data acquisition processes.		
c. Are specific, realistic, time-bound and actionable, and of a manageable number.		
d. Are clustered and prioritized.		
e. Reflect stakeholders' consultations whilst remaining balanced and impartial		
f. Clearly identify a target group for action.		
<b>9. LESSONS LEARNED</b>		<b>RATING:</b>
a. Are correctly identified, innovative and add value to common knowledge.		
b. Are based on specific evidence and analysis drawn from the evaluation.		
c. Have wider applicability and relevance to the specific subject and context.		

## SCORING

Element Of The Evaluation	Points Per Category	Points Awarded			
		Very Good	Good	Fair	Unsatisfactory
Presentation And Completeness	10				
Executive Summary	5				
Evaluation Context And Purpose	5				
Evaluation Scope And Methodology	10				
Reliability Of Data	5				
Findings And Analysis	35				
Conclusions	10				
Recommendations	15				
Lessons Learned	5				
<b>Total Maximum Score</b>	<b>100</b>				
		<b>Very Good</b> -> very confident to use	<b>Good</b> -> confident to use	<b>Fair</b> -> use with caution	<b>Unsatisfactory</b> -> not confident to use

## ASSESSMENT OF THE INTEGRATION OF GENDER EQUALITY AND EMPOWERMENT OF WOMEN (GEEW) for UN-SWAP

Quality Assessment Criteria	Comments	Score (0-3)
a. GEEW is integrated in the evaluation scope of analysis and indicators are designed in a way that ensures GEEW-related data will be collected.		
b. Evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved.		
c. Gender-responsive evaluation methodology, methods and tools, and data analysis techniques are selected.		
d. Evaluation findings, conclusions and recommendations reflect a gender analysis.		

<b>Overall Score</b>		
<b>Overall Rating</b>		

**UN-SWAP Scoring System**

<b>Exceeding Requirements</b>	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
<b>Meeting Requirements</b>	2 - Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done
<b>Approaching Requirements</b>	1 - Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.
<b>Missing</b>	0 - Not at all integrated. Applies when none of the elements under a criterion are met.
<b>Overall Calculation</b>	11-12 = very good      8-10 = good      4-7 = Fair      0-3=unsatisfactory

### Appendix 3. Recommended Revisions to UNODC EQA Template

General Project Information	
Project/Programme Number and Name	
Thematic Area	
Geographic Area (Region, Country)	
Approved budget of the time of the evaluation (USD)	
Type of Evaluation (In-Depth/Independent Project; final/ midterm; other)	
Evaluator(s)	
Date of Evaluation (from MM/YYYY to MM/YYYY)	
Date of Evaluation Report (MM/YYYY)	
Quality Assessment conducted on/by	

**OVERALL QUALITY RATING:**

**SUMMARY:**

Quality Assessment Criteria	Assessment Levels: Very Good - Good - Fair - Unsatisfactory Meets Criteria: Y = Yes N = No P = Partially <u>N/A- Not Applicable</u>
<b>1. Structure, Completeness And Clarity Of Report</b>	<b>RATING:</b>

a. Format (headings, font) accords to IEU Guidelines and Templates for Evaluation Reports.	
b. Structure accords to IEU Guidelines for Evaluation Reports with the following <u>logical</u> sequence: <u>List of acronyms</u> ; Executive Summary; Summary Matrix of Findings, Evidence and Recommendations; Introduction (Background and Context, Evaluation Scope and Methodology, Limitations to the Evaluation); Findings ( <u>Design</u> , Relevance, Efficiency, Partnership and Cooperation, Effectiveness, Impact, Sustainability, Human Rights and Gender Equality/ <u>Mainstreaming, Innovation</u> ); Conclusions; Recommendations; Lessons Learned.	
c. <u>Objectives stated in the terms of reference are adequately addressed.</u>	
d. <u>Issues of human rights and gender equality/mainstreaming are adequately addressed</u>	
e. Report is easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience)..	
f. <u>Language is empowering and inclusive avoiding gender, heterosexual, age, cultural and religious bias, among others.</u>	
g. Report is free from grammar, spelling, or punctuation errors.	
h. <u>Visual aids, such as maps and graphs, are used to convey key information</u>	
i. <del>Report contains a logical sequence: evidence-assessment-findings-conclusions-recommendations.</del>	
j. Composition of Evaluation Team is included and has gender and geographic expertise. Preferably it is gender balanced and includes professionals from countries or regions concerned.	
k. Annexes include at a minimum: evaluation terms of reference; <u>logic model and/or evaluation matrix</u> ; list of persons interviewed and sites visited; list of documents consulted; evaluation tools used.	
<b>2. Executive Summary</b>	<b>RATING:</b>
a. Written as a stand-alone section that provides an overview of the evaluation and presents its main results.	
b. Generally follows the structure of: i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology); iv) Main Conclusions; v) Recommendations.	

c. Summary Matrix presents only the key and most important recommendations from evaluation report.		
d. Findings, sources and recommendations in the Summary Matrix are clear and cohesive, and specify the stakeholder to whom they are addressed.		
e. Maximum length 4 pages, excluding the Summary Matrix.		
<b>3. Evaluation Context And Purpose</b>		<b>RATING:</b>
a. Clear description of the project evaluated is presented.		
b. Logic model and/or the expected results chain, and /or program theory (that at a minimum identifies and links objectives, outcomes and indicators of the project) is clearly described.		
c. Context of key cultural, gender related, social, political, economic, demographic, and institutional factors are described, and the key stakeholders involved in the project implementation and their roles are identified.		
d. Project status is described including its phase of implementation and any significant changes (e.g. <u>to</u> strategies, logical frameworks) that have occurred.		
e. Purpose of the evaluation is clearly defined, including why it was needed at that point in time, <del>who needed the information</del> , what information is needed, <del>how the information will be used</del> , and the target audience.		
<b>4. Scope And Methodology</b>		<b>RATING:</b>
a. Evaluation scope is clearly explained including the main evaluation criteria, questions and justification of what the evaluation did and did not cover.		
b. Transparent description presented of methodology applied, <u>including</u> how it was designed to address the evaluation purpose, objectives, questions and criteria <del>is explained</del> .		
c. Methodology allows for drawing causal connections between output and expected outcomes		
d. <u>Methods are appropriate for analysing gender equality/mainstreaming and human rights issues identified in evaluation scope;</u> <del>Gender sensitive</del> methodology <u>takes into account</u> power relations during an evaluation process; <u>is</u> inclusive and participatory.		
e. Data collection methods and analysis, and data sources are clearly described; as are the rationale for selecting them, and their limitations are clearly described. Reference indicators and benchmarks are included where relevant.		

f. Sampling frame clearly described and includes area and population to be represented, rationale for selection, mechanics of selection including whether random, numbers selected out of potential subjects, and limitations of sample.		
g.		
h. High degree of participation of internal and external stakeholders, including the Core Learning Partners, throughout the evaluation process is planned for and made explicit. When there are thematic or approach gaps (i.e. gender equality/mainstreaming) among stakeholders, <u>external</u> key informants <u>not</u> directly involved in the project were invited for consultation.		
<b>5. Reliability of Data</b> <i>To ensure quality of data and robust data collection processes</i>		<b>RATING:</b>
a. Triangulation principles (using multiple sources of data and methods) were applied to validate findings.		
b. Qualitative and quantitative data sources were used, and included the range of stakeholder groups and additional key informants (when necessary) defined in evaluation scope.		
c. Limitations that emerged in primary and secondary data sources and collection processes (bias, data gaps, etc.) are identified and, if relevant, actions taken to minimize such issues are explained.		
d. Evidence provided of how data was collected with a sensitivity to issues of discrimination and other ethical considerations.		
e. Adequate disaggregation of data by relevant stakeholder undertaken (gender, ethnicity, age, under-represented groups, etc.). If this has not been possible, it is explained.		
<b>6. FINDINGS AND ANALYSIS</b> <i>To ensure sound analysis and credible findings</i>		<b>RATING:</b>
<i>Findings</i>	-	
a. <u>Are clearly formulated and presented</u>		
b. <u>Are based on rigorous analysis of the data collected; take into account any identified benchmarks.</u>		
c. Address all evaluation criteria and questions raised in the ToR including relevance, efficiency, effectiveness, impact and sustainability, as well as UNODC's additional criteria of design, partnership and cooperation, innovation,		

and the cross-cutting themes of human rights and gender.	
d. Address any limitations or gaps in the evidence and discuss any impacts on responding to evaluation questions raised in ToR.	
e. Discuss any variances between planned and actual results of the project (in terms of objectives, outcomes, outputs).	
<del>f. Are presented in a clear manner.</del>	
<i>Analysis</i>	-
a. Interpretations are based on carefully described assumptions.	
b. Contextual factors are identified (including reasons for accomplishments and failures, and continuing constraints).	
c. Cause and effect links between an intervention and its end results (including unintended results) are explained.	
d. Includes substantive analysis of gender equality/mainstreaming issues	
e. Includes substantive analysis of human rights issues.	
<b>7. CONCLUSIONS</b>	<b>RATING:</b>
a. Take into consideration all evaluation criteria and questions, including human rights and gender equality/mainstreaming criteria.	
b. Have been formulated clearly and are based on findings and substantiated by evidence collected.	
c. Convey the evaluators' unbiased judgement of the intervention	
d. Developed with the involvement of relevant stakeholders.	
e. Present a comprehensive picture of both the strengths and weaknesses of the project.	
f. Go beyond the findings and provide a thorough understanding of the underlying issues of the project and add value to the findings.	
<b>8. RECOMMENDATIONS</b>	<b>RATING:</b>
a. Are clearly formulated, based on the conclusions, and substantiated by evidence collected.	
b. Address flaws, if any, in project's data acquisition processes.	
c. Are specific, realistic, time-bound and actionable, and of a manageable number.	
d. Are clustered and prioritized.	
e. Reflect stakeholders' consultations whilst remaining balanced and impartial	

f. Clearly identify a target group for action.		
<b>9. LESSONS LEARNED</b>	<b>RATING:</b>	
a. Are correctly identified, innovative and add value to common knowledge.		
b. Are based on specific evidence and analysis drawn from the evaluation.		
c. Have wider applicability and relevance to the specific subject and context.		

## SCORING

Element Of The Evaluation	Points Per Category	Points Awarded			
		Very Good	Good	Fair	Unsatisfactory
Presentation And Completeness	10				
Executive Summary	5				
Evaluation Context And Purpose	5				
Evaluation Scope And Methodology	10				
Reliability Of Data	5				
Findings And Analysis	35				
Conclusions	10				
Recommendations	15				
Lessons Learned	5				
<b>Total Maximum Score</b>	<b>100</b>				
		<b>Very Good</b> -> very confident to use	<b>Good</b> -> confident to use	<b>Fair</b> -> use with caution	<b>Unsatisfactory</b> -> not confident to use

## ASSESSMENT OF THE INTEGRATION OF GENDER EQUALITY AND EMPOWERMENT OF WOMEN (GEEW) for UN-SWAP

Quality Assessment Criteria	Comments	Score (0-3)
a. GEEW is integrated in the evaluation scope of analysis and indicators are designed in a way that ensures GEEW-related data will be collected.		

b. Evaluation criteria and evaluation questions specifically address how GEEW has been integrated into design, planning, implementation of the intervention and the results achieved.		
c. Gender-responsive evaluation methodology, methods and tools, and data analysis techniques are selected.		
d. Evaluation findings, conclusions and recommendations reflect a gender analysis.		
<b>Overall Score</b>		
<b>Overall Rating</b>		

#### UN-SWAP Scoring System

<b>Exceeding Requirements</b>	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
<b>Meeting Requirements</b>	2 - Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done
<b>Approaching Requirements</b>	1 - Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.
<b>Missing</b>	0 - Not at all integrated. Applies when none of the elements under a criterion are met.
<b>Overall Calculation</b>	11-12 = very good    8-10 = good    4-7 = Fair    0-3=unsatisfactory