



UNODC

United Nations Office on Drugs and Crime

Independent Evaluation Section



Independent Quality Assessment of UNODC Evaluation Reports 2019

Synthesis Report



CONTENTS

1. INTRODUCTION	3
2. METHODOLOGY	3
3. FINDINGS	4
3.1 EQA Ratings.....	4
3.2 General Strengths and Improvements of the Evaluation Reports.....	6
3.3 General Weaknesses of the Evaluation Reports	10
3.4 Mainstreaming Human Rights and Gender Considerations	11
3.5 Assessing Contributions to SDGs.....	14
3.6 Evaluation Follow Up	17
3.7 Evaluation Complexity	18
4. RECOMMENDATIONS	20
Annex 1. List of evaluations reviewed	23
Annex 2. Proposed Complexity Assessment Framework for Evaluation Planning	24
Annex 3. UNODC EQA Template as Used in the Review	25

1. INTRODUCTION

The Independent Evaluation Section (IES) is leading and guiding evaluations in order to provide objective information on the performance of United Nations Office on Drugs and Crime (UNODC). Evaluation is one of the key factors in the Secretary-General's UN reforms that were considered by the General Assembly in 2019.¹ While most evaluations in the Secretariat are done by the Office of Internal Oversight Services (OIOS), only a few entities considered to be part of the Secretariat have independent evaluation offices. UNODC is one these and its progress in undertaking quality evaluations is an important contribution to the larger effort.

As a member of the United Nations Evaluation Group (UNEG), the IES is following its Norms and Standards. IES's work is based on three pillars: 1) National Evaluation Capacity Building and SDGs; 2) Evaluation results; 3) Evaluation knowledge products, communication and innovation.

As part of its efforts to ensure the office's independent evaluations (pillar 2) are providing credible information to inform planning processes, the IES has commissioned independent evaluation quality assessments of evaluation reports produced since 2014. The IES also seeks to improve evaluation practice and to create better mechanisms for tracking and using the knowledge gained from evaluations as part of ongoing organizational learning.

Building on the previous Evaluation Quality Assessment (EQA) report, this document synthesizes the EQA results of all published UNODC evaluation reports in 2019 (the list of reports is provided in Annex 1) and makes comparisons with EQAs since 2014/15.

This assignment was carried out from mid-January 2020 to end-March of 2020 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).

2. METHODOLOGY

EQA Template: The first deliverable on the annual evaluation quality assessment review was to assess the EQA template used for the 2018 reports and to propose any needed adjustments. Changes were made to nine questions. In six cases, this was to further clarify the intent of the question. Three questions were deleted in order to minimize duplication of question areas and as part of efforts to shorten the EQA tool. In addition, a field was added to the profile section of the form to capture evaluation complexity. The final version of the template used for the 2019 review is attached as Annex 2.

2019 EQA Process: The reviewers then examined the quality of all of the evaluation reports published during 2019. The total number of evaluations for the year was 11. Two were in-depth evaluations and nine were independent project evaluations.

To ensure consistency of the review process, two of the reports were selected for assessment by both reviewers. 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

¹ Eighth progress report on the accountability system in the United Nations Secretariat: strengthening accountability under the new management paradigm (A/73/688)

differences were discussed and resolved. The remaining reports were each rated by one person with assignments being based on each reviewer's area of expertise and language skills.

Consideration of SDGs and HRGE: The reviewers were requested to pay specific attention to the extent to which the Sustainable Development Goals and issues of Human Rights and Gender Equity were considered as part of evaluation processes. This report includes a separate chapter on each of these areas.

Consideration of Evaluation Follow up and Complexity: Additional components of the assignment included an assessment of the extent to which recommendations from 2019 evaluation reports were addressed. This entailed a review of available Management Responses (provided in table form by IES); data was available for eight of the reports. The team also looked at how issues of complexity of UNODC-commissioned evaluations could be captured and addressed in evaluation planning processes. The method used for this part of the assignment included consultation with IES staff and the team leader of the IISG evaluation, which was identified as a good case example of a complex evaluation process. An approach paper to complexity was prepared separately and a summary is included in Section 3.7 of this report.

3. FINDINGS

3.1 EQA Ratings

The 11 published evaluation reports in 2019 were rated highly and reveal the continuous improvement in the quality of UNODC evaluations since 2015². Both Figure 1 & 2 and Table 1 show this trend. Figure 2 also shows the significant increase in Very Good reports this year. All of the reports were rated as Very Good or Good, with 73% being rated as Very Good.

Figure 1: Percent of combined Very Good and Good reports by year

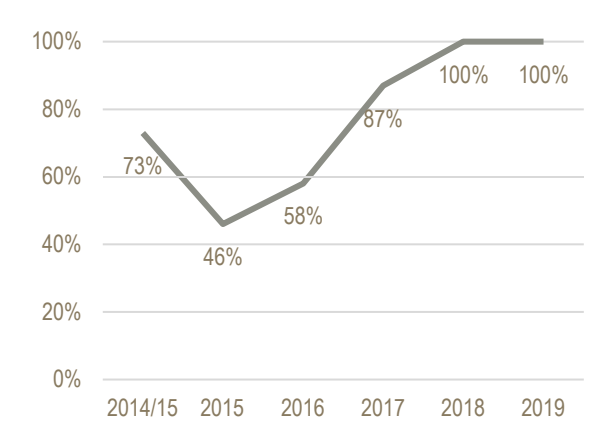
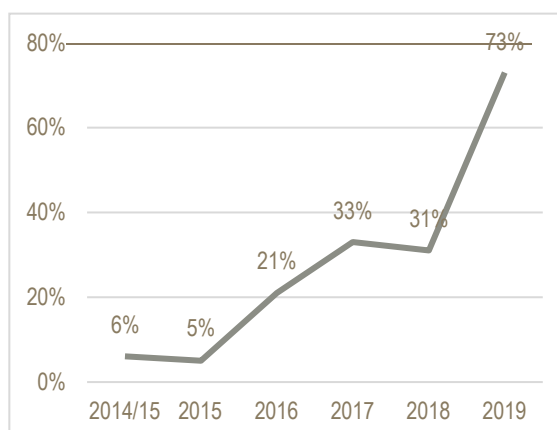


Figure 2: Percent of Very Good reports by year



The numerical scoring that was introduced to the EQA template for review of the 2017 and 2018 reports, and again used in 2019, brings an additional level of preciseness to the ratings. The ratings are based on the overall scores of each report: Very Good (90%+), Good (70-89%), Fair (50-69%), and Unsatisfactory (<50%).

² It should be noted that since 2016 the IES includes the EQA template as an attachment to introductory emails to all evaluators. The EQA criteria and template are also found in the [Evaluation Handbook: Guidance for designing, conducting and using independent evaluation at UNODC](#) (October 2017).

In 2019, eight reports (73%) received a Very Good rating and no reports were rated as Fair or Unsatisfactory. The 2019 scores ranged from a low of 77% to a high of 96%. Particularly notable is that the average score was 88, which is excellent and 2 points higher than in 2018.

Table 1 also shows that the number of commissioned evaluations has decreased substantially with 11 evaluations undertaken in 2019 compared with period between 2014 – 2016 when there was an average of over 20 evaluations each year. During this time, the quality has improved suggesting a probable link between IES staff time and evaluation quality. The IES reports that fewer evaluations were done in 2019 due to the increased complexity of these assignments.

Table 1: Overall rating of 2019 reports compared with previous EQA cycles

	Unsatisfactory	Fair	Good	Very Good	Total
# of Reports - 2019	0	0	3	8	11
% of Reports - 2019	0%	0%	27%	73%	100%
# of Reports - 2018	0	0	11	5	16
% of Reports - 2018	0%	0%	69%	31%	100%
# of Reports - 2017	0	2	8	5	15
% of Reports - 2017	0%	13%	54%	33%	100%
# of Reports - 2016	0	8	7	4	19
% of Reports - 2016	0%	42%	37%	21%	100%
# of Reports - 2015 ³	0	12	9	1	22
% of Reports – 2015	0%	53%	41%	5%	100%
# of Reports – 2014/15 ⁴	0	9	22	2	33
% of Reports – 2014/15	0%	27%	67%	6%	100%

As with previous years, there were some notable differences by criterion. Table 2 shows that as was the case in 2018, Recommendations tended to have lower ratings. In 2016, the lower ratings were found for the Executive Summary and Reliability sections. In 2017, Presentation & Structure rated the most highly. In 2018 Lessons Learned ranked highest. In 2019 while the largest number of Very Good was in Lessons Learned, it had two evaluations with problems in this section.

Table 2: Report rating by criteria

	Unsatisfactory	Fair	Good	Very Good
Presentation/Structure	0	0	4	7
Executive Summary	0	0	7	4
Context & Purpose	0	1	3	7
Scope & Method	0	2	4	5
Reliability	0	0	4	7
Findings	0	2	2	7
Conclusions	0	0	3	8
Recommendations	0	3	6	2
Lessons Learned	1	1	0	9
Consideration of GEEW	0	2	4	5

Table 3 shows the scores by type of report: Independent Project Evaluations – midterm and final – and programmatic In-depth Evaluations. Both of the In-depth evaluations were Very Good.

³ This included reports published from June through December 2015.

⁴ This included the first batch of EQA assessments, which considered reports published from January 2014 through May 2015.

Table 3: Rating by type of evaluation

Type of evaluation	Good	Very Good	Grand Total
Independent Project Final	2	5	7
Independent Project Midterm	1	1	2
Indepth Formative	0	1	1
Indepth Final	0	1	1
Grand Total	3	8	11

Table 4 shows the scores by subject area. There were too few reports in any of the areas to draw conclusive findings about which had the strongest reports.

Table 4: Rating by subject of project

Subject	Good	Very Good	Grand Total
Corruption	1	1	2
Drug Trafficking	0	1	1
Drug Prevention, Treatment & Care	0	2	2
Crime Prevention & Criminal Justice	1	1	2
Human Trafficking & Migrant Smuggling	1	2	3
Organized Crime	0	1	1
Grand Total	3	8	11

3.2 General Strengths and Improvements of the Evaluation Reports

This section looks first at areas where UNODC reports have commonly excelled over the past three years. It then considers the extent to which the recommendations from the 2018 Synthesis Report appear to have been taken up and highlights examples of good practice found in the 2019 reports.

Consistency with DAC and UNEG Norms: The reports continue to generally conform to the accepted norms and guidelines. The DAC undertook a review and updating of the criteria for assessing evaluation quality in 2019, but many of the main changes being considered – such as more focus on gender equity and human rights, contribution to SDGs, partnerships, design rigour, and learning – were already incorporated into UNODC evaluation guidance, the EQA process, and most agency evaluations. The DAC’s newly introduced criterion, Coherence, will need to be incorporated into the 2020 EQA template.

Inclusion of Gender Analysis: This section has consistently shown improvement and in 2019 nine of the 11 reports were rated as Very Good or Good. Further discussion is provided in the “Mainstreaming of Human Rights and Gender” section below.

Executive Summaries: This section was highlighted as being the weakest section of the 2016 reports but has since shown marked improvement. For 2019, all of the 11 reports received a section score of Very Good or Good. The most significant improvement has been that these sections are more concise, with most reports adhering to the maximum length of four pages. A previously observed issue - of the emphasis, and most of the space, being given to summarizing the findings of each criteria rather than to the overall conclusions of the evaluation process – was less evident in 2019. As a result, the reader is more easily able to get a sense of the overall accomplishments, strengths and weaknesses of the evaluation object.

Lessons Learned: As in 2018, this section of the reports received the highest number of Very Good ratings (9). But it also was the only section to receive an Unsatisfactory (1). This was the XASV23 evaluation which did not have a special section on lessons learned. As the EQA said “*The lessons are woven into the final recommendation which is titled Lessons Learnt and Best Practices. The*

description lists key topic areas from which lessons could be drawn but there are no details about what the lessons actually are.”

Good practice example of presenting Lessons Learned: The XSPZ91 report succinctly states eight lessons and best practices that highlight factors that have helped drive project success. These included keeping the focus on outputs and outcomes (the bigger picture), ensuring the flexibility to respond to beneficiary needs, and making partnerships inherent to the methodology and not a by-product of the project. These lessons have clear applicability to other interventions that work at the regional level with a range of government and non-government actors.

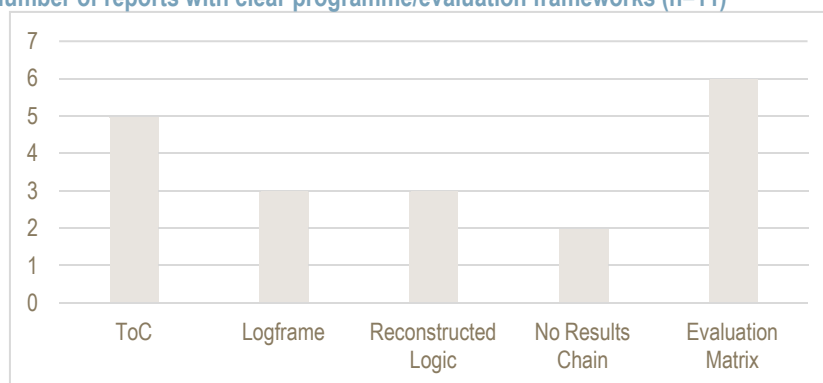
Uptake of Recommendations from 2018 Synthesis Report

1. **Cause-effect links to be more clearly reflected in the evaluations:** Improvements were seen in the ways evaluators showed how the intervention’s activities and outputs contributed to its intended results. These connections were evident in eight of the 11 reports undertaken in 2019; they were very clearly shown in three and one report did not address the linkages. A particularly good example was GLO.ACT where the EQA noted, *“This was well done, mainly due to careful reconstruction of the ToC with stakeholder input”*.

The most effective way to show causal relationships is through use of the programme theory as the basis for the analysis. As shown in figure 3, five reports included a theory of change, three included a logical framework, and two did not provide any information to show the intended chain of results. Good practice is for evaluators to assess the intervention’s programme theory and, if found inadequate, to reconstruct it so that it serves as a more effective framework for the evaluation. In 2019, two reports included reconstructed logical frameworks (MEXK54 and MEXY93), one had a reconstructed theory of change (GLO.ACT), and three included a recommendation that such an exercise should be undertaken.

Linkages can also be apparent in well-developed evaluation matrices. In the 2019 reports, six reports included an evaluation matrix and one referred to a matrix being produced for the inception phase but did not otherwise reference or attach it to the report.

Figure 3: Number of reports with clear programme/evaluation frameworks (n=11)



2. **Increasing robustness of methodologies used:** Three of the recommendations made in 2018 focused on methodology; these suggested the use of more diverse evaluation designs as well as the inclusion of more details about sampling strategies, data analysis, and stakeholder engagement processes.

In respect to design, there was some progress seen in in 2019 although over half of the evaluations (6 of 11) relied solely on the more common data collection processes of document review, KIs, and FGDs. The other five evaluations also included surveys and, of these, three

were able to use the data obtained (in one case was the response rate too low and in another, the planned survey could not be administered due to government restrictions). Two evaluations, conducted by the same team lead, went further by collecting data through champions workshops. In at least two cases, observations were mentioned as a form of data collection but there was no indication of this being done in a systematic way with the use of standard protocols such as checklists.

Good practice example of a robust methodology: GLO.ACT stood out for its extensive data collection process. The evaluation team reviewed 1150 documents, visited seven countries, conducted 315 interviews, organized workshops with programme champions, and administered a survey in four languages from which they received 105 full responses.

Four reports had clearly described sampling approaches and five had partial descriptions where either sampling was described just for one type of data source or was only vaguely described. Examples of the latter include one report with a general statement noting that sampling was based on stakeholder analysis, and another report indicating that criteria had been established for determining sites for field visits but did not then provide that criteria.

All evaluations relied heavily on qualitative data and most included at least some quantitative data. Three reports included quantitative data from primary sources (those using surveys), and these along with four other reports drew this type of data from documents. Four reports did not show use of quantitative data except for some financial information on programme costs. In respect to data analysis, five reports included a thorough description of processes used, five gave a partial description, and one did not give any information.

Strong evaluation designs also include involving stakeholders in the development of recommendations. Such involvement was clearly described in three reports. In six reports, there was some reference to stakeholder participation but not specifically in respect to contributing to recommendations, and four reports did not address their involvement.

3. **Reliability of data:** This was another area in the EQA template where there has been continuous improvement. For 2018, 50% of the evaluations had this section rated as Very Good, with the rest being Good. In 2019, improvements continued with seven of the 11 (64%) evaluations being Very Good, with the rest being Good. The main factor in assessing reliability is the connection between findings and the quality of data on which the findings are based. In 2019, the connection was very clear in most evaluations leading to the higher scores.
4. **Recommendations that address any flaws in the project M&E data systems:** This is an area where many evaluations have fallen short in previous years and where there is now some improvement. Evaluations are expected to assess the adequacy of data systems and make recommendations accordingly. In 2019, seven reports included such recommendations while four did not, suggesting an increasing attentiveness on the part of evaluators to intervention monitoring systems.
5. **Practices to increase the functionality and use of reports:** To support evaluation use, a consistent recommendation in these synthesis reports has been about the need to encourage the incorporation of more visual aids to convey key information: and improvements were seen from previous years. In 2019, six reports showed frequent use and diverse types of visual aids to explain evaluation context, processes and results. Exemplary reports in this regard include BHUZ13, EGYZ33, GLO.ACT, IISG, and MEXK54. Figure 4-7 provide illustrative examples. In several reports, while visual aids were included, formatting of some graphics could have been improved, such as where text was too small to easily read, images were blurred or stretched, or labelling was not consistent with UNODC standards.

Use is also promoted by well-structured findings, and past recommendations have encouraged this report section to have evaluation questions and a summary of findings highlighted for each

criterion. This practice has clearly been adopted as all of 2019 reports had findings that included the evaluation questions and all but one also included the summary statements.

Figure 4: Illustrative Example of Good Use of Visual Aids to show evaluation scope (IISG evaluation)

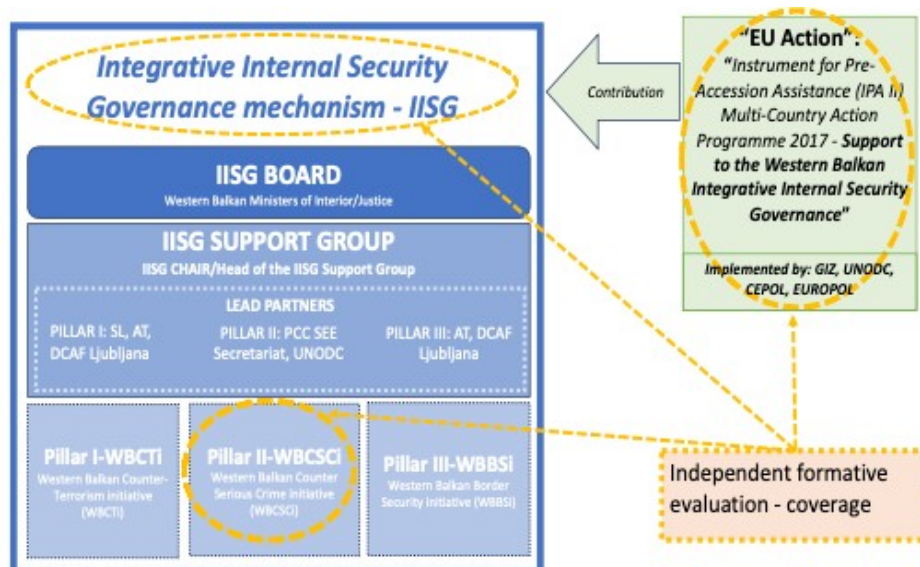


Figure 3. Scope of the current evaluation

Figure 5: Illustrative Example of Good Use of Visual Aids to show programme components (GLO.ACT evaluation)

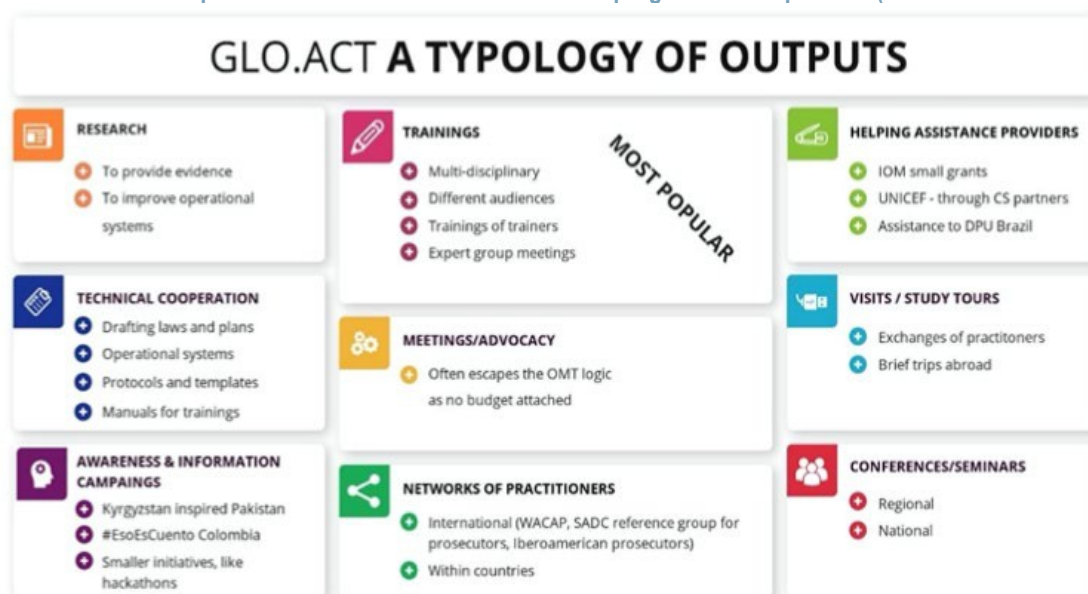


Figure 6: Illustrative Example of Good Use of Visual Aids to summarize evaluation conclusions (IISG)

Table 1. SWOT Analysis for the WBCSCi/ IISG Pillar II

Strengths	Weaknesses
'Proof of concept'	Underdeveloped systems and processes
Ability to bring decision-makers on the table	Uneven stakeholder involvement
High visibility and presence	Weak results-based monitoring systems
Operating on shoestring budget	Inadequate attention to HRGE issues
Opportunities	Threats/ Challenges
High donor interest and funding in the region	Easy funding opportunities for beneficiaries
Visible need for better coordination	Stakeholder impatience for quick results
Operational coordination	Organizational location: Limiting to stakeholders?
Independent organisational structure	Unclear and conflicting stakeholder expectations

Figure 7: Illustrative Example of Good Use of Visual Aids to show programme expenditure by result area (EGYZ33 evaluation)⁵

Outcome	Budgeted	Spent	Discrepancy	Comment
1. VAW legislation	Campaigns \$431,232	Campaigns \$157,833	Campaigns \$273,399	Campaign material of the NCW hotline and not for legislation
2. Governmental coordination	Coordination mechanism \$98,008	Coordination mechanism \$0.00	Coordination mechanism \$98,008	Reallocated to NCW campaign and the FMA equipment
3. General capacity development for MoI	Training Police \$176,413 Equipment \$198,014	Training Police \$148,446 Equipment \$0.00	Training Police \$27,967 Equipment \$198,014	Details for reallocation not traceable
4. Capacity for the VAW office at MoI	Training staff \$158,800 Strategy \$278,400	Training staff \$0.00 Strategy \$0.00	Training staff \$158,800 Strategy \$178,400	Details for reallocation not traceable
5. General capacity development for FMA	Overall \$158,820	Training staff \$95,246 Clinics \$454,957	Overall -\$393,383	Extra funds came to FMA from cancelled activities
6. General capacity development for PPO	Trainings \$158,811 VAW Courts \$78,406	Trainings \$204,821 VAW Courts \$0.00	Trainings -\$48,010 VAW Courts \$78,406	
7. Data management for PPO and NCW	PPO db \$176,427 NCW hotline \$158,816	PPO database \$412,535 NCW training \$21,122 NCW hotline \$85,370	PPO database -\$236,108 NCW \$60,325	Most reallocated money invested in this outcome
	\$1,960,147	\$1,570,329	\$389,818	

6. Inclusion of standard information required: This category includes the need for a brief description of core evaluation team members and their suitability for the assignment. Improvements were also seen here as all but one report had this information. Another frequently missing element is the dates/timeframe of the evaluation process. In 2019, this information was specified in five of the 11 reports, less than half.

3.3 General Weaknesses of the Evaluation Reports

Recommendations: As noted above, while better than in 2018, this was still the weakest section with three of the 11 reports receiving a Fair rating and only two receiving a Very Good rating. The most common issues were that (a) recommendations were not clustered or prioritized, (b) there

⁵ This image is somewhat blurred but is included as an illustrative example for its effective presentation of content.

were no recommendations related to SDGs, and (c) there was no indication that stakeholders were consulted as part of their development.

Evidence of stakeholder consultation in developing conclusions and recommendations: Good practices call for stakeholder involvement in providing feedback on preliminary findings and in the formulation of conclusions and recommendations - UNODC ToRs are generally explicit about the need for engaging Core Learning Partners (CLPs) throughout the process. Although most all UNODC evaluations include consultation with a range of relevant stakeholders as part of data collection, few reports articulate how stakeholders may have been further involved. In fact, only three of the 11 evaluations received a Yes on *“High degree of participation of internal and external stakeholders, including the Core Learning Partners, throughout the evaluation process is planned for and made explicit”* in the scope and methodology section, slightly lower than in 2018. And, in terms of recommendations, only four evaluations received a Yes on *“Reflect stakeholders’ consultations whilst remaining balanced and impartial”* while two received a No.

Consideration of SDGs: As was the case in 2018, in all but two of the reports the connection and contribution of the initiative to the SDG as goals is described. However, only two evaluations had mention of specific SDGs targets. Section 4 provides more analysis and guidance for how SDGs could potentially be addressed.

3.4 Mainstreaming Human Rights and Gender Considerations

All evaluation reports are rated on the extent to which they meet 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. Each criterion is rated on a scale of 0 - 3 with 0 being awarded when there is no integration of gender, 1 when gender issues are partially integrated, 2 when it is satisfactorily integrated, and 3 when gender is fully integrated.

Table 5 shows that the average scores for considering GEEW have continued to improve. As with the previous two years, all UNODC evaluation reports have included specific sub-sections on Human Rights & Gender or mainstreamed GEEW into multiple criteria under Findings. Although full mainstreaming of gender may be held up as the gold standard, the provision of a sub-section on GEEW has been a factor to ensure that the issue is always considered in UNODC evaluation reports and, in many cases, it appears that GEEW has been more fully assessed when it is done in a dedicated section.

Table 5: Average scores for the integration of GEEW (UN-SWAP)⁶

Quality Assessment Criteria	Average Score (0-3)				
	2015	2016	2017	2018	2019
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	1.93	2.44	2.36
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	2.2	N/A	N/A
c. Gender-responsive evaluation methodology, methods and tools, and data analysis techniques are selected.	.875	1.37	1.87	2.25	2.09
d. Evaluation findings, conclusions and recommendations reflect a gender analysis.	1.875	1.84	2.2	2.31	2.73

⁶ In 2018 the number of UN-SWAP criteria were reduced from four to three.

Average Overall Score (out of 12)	5.5	6.53	8.2	N/A	N/A
Average Overall Score (out of 9)	N/A	N/A	N/A	6.94	7.09

Table 6 shows that evaluators are consistently carrying the gender-related findings into the conclusions and recommendations.

Table 6: Number of reports achieving scoring requirements for each criterion

	A. Scope/Indicator	B. Methodology	C. Findings/Conclusions/Recommend
Not at all integrated	0	0	0
Partially integrated	0	2	0
Satisfactorily integrated	8	6	3
Fully Integrated	3	3	8

Table 7 provides the overall EQA rating and the GEEW category scores for each report. It shows that all of 2019 reports were rated as Good/Met Requirements both in their overall rating and in their GEEW rating. The overall GEEW score for the 2019 reports is 7.1 which shows that UNODC evaluations are Meeting Requirements (which is the highest category). The average score for all of the three criteria was 2.4. The table also includes the average scores (at the individual evaluation level and not rounded) for the respective criteria for the 2017 and 2018 reports.

Table 7: GEEW scores by evaluation report

Project Number	Overall EQA Rating	Scores for each criterium			GEEW Total Score
		A	B	C	
MEXK54	Very Good	2	1	2	5
MEXZ93	Very Good	2	3	3	8
PSEY13	Very Good	2	2	3	7
EGYZ23	Very Good	3	2	3	8
GLO.ACT	Very Good	3	2	3	8
GLOZ99	Very Good	2	2	3	7
XCAX75	Very Good	2	2	3	7
BHUZ13	Good	2	3	2	7
ISSG	Very Good	3	3	3	9
XASV23	Good	2	2	3	7
XSP/Z91	Good	2	1	2	5
Average score 2019		2.27	2.09	2.73	7.09
Average score 2018		2.44	2.25	2.31	6.94
Average score 2017		1.93	1.87	2.2	N/A

Observations about the way that each criterium was addressed in the 2019 evaluations are as follows:

Scope and Indicators: All evaluation teams made a clear effort to incorporate gender into the evaluation design. Most included gender as part of the scope and/or objectives of the evaluation, and over half identified the process as being gender responsive. In addition to increasing awareness amongst evaluators, the success here likely reflects the emphasis being placed on gender in evaluation ToRs and suggests that evaluation managers are communicating the need

for a gender lens at the inception stage. All evaluations included criteria and questions pertaining to gender. However, it was not always possible to see if gender-related indicators were specified, given that logical frameworks or evaluation matrixes were not consistently included in the reports.

Methodology: As seen in previous years, scores in this sub-section tend to be somewhat lower than the other GEEW sub-sections. Although most evaluators included language about the methodology being gender responsive, they did not include an explicit description of how this was done. This contrasts with the 2018 evaluations where the strongest GEEW-scoring reports clearly described how gender was to be analyzed and how the methodology was designed to be gender responsive. These also tended to be evaluations that included a gender expert on the evaluation team and, in one case, a HRGE evaluability assessment was carried out during the inception phase of the evaluation. On the other hand, the average scores for 2019 were still higher than previous years suggesting an increasing awareness across evaluation teams of the need to attend to gender. In accordance with UN-SWAP guidance, mixed-methods appropriate to evaluating GEEW were used, a diverse range of data sources and processes were employed to enable triangulation, and an appropriate diversity of stakeholder affected by the intervention were consulted. In most cases, stakeholders consulted as part of the evaluation were gender disaggregated.

Less clear is the extent of alignment with the SWAP guidance on ethical standards being considered throughout the evaluation process. Several of the 2019 evaluations attended to the need for confidentiality but none went further, for example by explaining how differences in power relations amongst evaluation participants were addressed. An indicative EQA comment being *“The methodology was noted as being participatory and gender-disaggregated data was collected. However, the evaluators were not specific about steps taken to ensure the data collection was gender sensitive. There were two women in a mainly male focus group and it is not specified what steps, if any, were taken to ensure their voices were heard or to ensure they felt comfortable participating.”*

Findings, Conclusions and Recommendations: All reports included gender-related findings. The level of inquiry varied with some reports briefly considering the extent to which gender was integrated into the design and implementation of the evaluation object. Others provided a more thorough analysis aligned with SWAP guidance in referencing relevant normative instruments and policies related to HRGE. Several evaluations considered gender within multiple criteria (with gender being considered in all criteria in XASV23). Nine of the 11 reports had both gender-related conclusions and recommendations.

Good practice example of a gender-responsive evaluation process: The PSEY13 evaluation team constructed a theory of change highlighting the project's role in reducing gender inequalities. The inception phase included a stakeholder analysis to ensure the diversity of stakeholders were included in the study. The sampling criteria for field visits then took into account villages where the 'presence of women is stronger' – approximately 40% of people consulted were women. The background section of the report addressed how gender was integrated into the project. Gender analysis is included under multiple criteria - Design, Relevance, Partnership, Effectiveness, and HRGE.

Good practice example of substantial analysis of HRGE: GLOZ99 is notable for having a thorough gender analysis as part of findings. The report's section on Human Rights, Gender Equality and Leaving No One Behind references Human Rights Council resolutions, UNODC strategies and guidance, and relevant regional declarations in assessing the project's related accomplishments and shortcomings.

Good practice example of gender mainstreaming: The IISG report is explicit about GEEW being a key part of the evaluation scope. The evaluation includes three evaluation questions, with corresponding indicators, specific to HRGE - under Design, Relevance and Effectiveness. Interview and survey respondents are disaggregated by gender, and interviewees are also gender disaggregated by stakeholder group. Part of the substantial analysis of gender in the findings includes a rating of the intervention on the Gender Results Effectiveness Scale (UNDP/Gender@Work).

Good practice example of highlighting use of ethical principles: The BHUZ13 stands out for having a sub-section on evaluation ethics in the methodology section that notes the evaluation was conducted in accordance with UNODC Evaluation Policy and UNEG Ethical Guidelines and Code of Conduct for Evaluation in the UN System. The evaluators explicitly describe how they then attended to each of the five elements of the Code. The analysis looks at how human rights issues underpin the project and are reflected in project activities. There is also a section specific on No One Left Behind that further assesses the extent that vulnerable groups were included in the project design and implementation.

3.5 Assessing Contributions to SDGs

The role of evaluation in the review process of the Sustainable Development Goals continues to be important, especially in relation to Goal 16, to promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels. The review process is organized around Voluntary National Reviews (VNRs) that are considered by the High-Level Political Forum (HLPF) of the Economic and Social Council and, once every fourth year, to the General Assembly. Goal 16 was considered at the 2019 Session of the HLPF. It concluded that:⁷

Advances in ending violence, promoting the rule of law, strengthening institutions and increasing access to justice are uneven and continue to deprive millions of their security, rights and opportunities and undermine the delivery of public services and broader economic development. Attacks on civil society are also holding back development progress. Renewed efforts are essential to move towards the achievement of Sustainable Development Goal 16.

As currently designed, the reviews are supposed to be based on indicators, whose determination takes place via the UN Statistical Commission. There are three levels of indicators, called Tier I, II and III. Tier I indicators are those for which there is a consensus and reasonable assurance that they could be measured if the data were collected by the country. Tier II Indicators are conceptually clear, with established methodology and standards available but data are not regularly produced by countries. Tier III Indicators are those for which there are no established methodology and standards or methodology/standards are being developed/tested. Six of the goals and 44 indicators are specifically a concern of UNODC. UNODC is a custodian (or co-custodian) of 13 indicators, meaning that it is responsible for developing, maintaining and encouraging data collection on the indicators. Ten of these are Tier II or Tier III indicating that more work needs to be done to make them useable for the review process.

One problem that is evident from a review of the indicators is that for most of those for which

⁷ United Nations, Sustainable Development Goals Knowledge Platform, [Sustainable Goal 16](#), 2019

UNODC is the principal, especially in Goal 16, there are no real indicators of targets that have been agreed that measure how national institutions function. The exceptions are 3.3.1 on HIV incidence, 3.8.1 on coverage of health services which would include substance abuse treatment, 16.1 on homicides and 16.4 on money-laundering. For the others, there are few indicators for which data are currently being collected by national statistical systems, or where the indicators really measure whether the target has been achieved. As a result, reporting on these has been limited.

This can be seen in the first stage of the review process that has been the presentation of voluntary national reviews (VNRs) at the High-Level Segments in 2016, 2017, 2018 and 2019. One hundred forty-two countries presented VNRs, some of them twice. While there are annual appraisals of the VNRs, they do not provide target-based data and they varied in content and approach. The analysis of which of the UNODC indicators the VNRs considered in 2017, shown in Table 8 and presented in the 2017 review of UNODC evaluations, shows significant variation and confirms the finding that many indicators are not really available. The target with the most references in the VNRs was 5.2 on violence against women (35 percent), largely because there are five organizations that collect this data systematically in many countries. The next largest number of references is to 3.3.1 on HIV/AIDS (30 percent) because WHO collects this information in many countries. Only two other indicators had references greater than 20 percent, and those are 3.5.1 dealing with health services (a Tier III on which UNODC is working with WHO) and 16.1 on homicides (a Tier I indicator).

As can be seen from Table 8, the analysis from 2017 also shows that the number of references to Goal 16 is very limited.

Table 8. Goal 16 Indicators mentioned in VNRs in 2017

Indicator	Whether Mentioned		
	No	Yes	Total
16.1 - Violence	32	11	43
16.3 - Prosecution	35	8	43
16.4 - Illicit flows	37	6	43
16.5 - Corruption	35	8	43
16.6 - Transparent government	36	7	43
16.7 - Participation in Decision-making	39	4	43
16.8 - Transparent public services	40	3	43
16.10 - Human rights violations	36	7	43
16.a - Independent Human Rights institutions	39	4	43
16.b - Harassment	40	3	43

The table suggests that a special effort will be needed to ensure that States report on progress, including by assisting them to collect data, both by determining what to measure and when that is clear, how to collect the data. Indeed, in 2019 only four of the 47 countries presenting VNRs mentioned crime in their “main messages” to the High-Level Task Force.⁸

In this context, evaluation, which is part of the SDGs in paragraph 74 (g) and (h) of General Assembly resolution 70/1 that adopted the SDGs, should be a major element in reporting, especially where Tier I indicators do not exist. Because evaluations collect data, especially at the country-level, about what has happened in subject areas and why, they can be an important tool in SDG reporting. To test this, SDG targets that were covered by UNODC evaluations in 2019 were examined as shown in Table 9. All but two of the evaluations made specific reference to the SDGs.

⁸ United Nations, Economic and Social Council, Compilation of main messages for the 2019 voluntary national reviews, (E/HLPF/2019/5), 24 May 2019.

As in past years, most (8) were connected with Goal 16, with the largest connected with Target 16.2 on ending abuse, exploitation, trafficking and all forms of violence against and torture of children. Six evaluations addressed target 5.2 on eliminating violence against women. The number of targets noted is much smaller than in 2018, primarily because most evaluations did not indicate the specific targets within the goal which were covered. However only two reports went further by specifying the actual targets addressed.

Table 9. SDG Targets Covered by UNODC Evaluations, by Type, 2019

SDG Targets	In-Depth Evaluations	Independent Project Evaluations	Grand Total
1.4 Equal rights to economic resources			0
2.1 Access to safe, nutritious and sufficient food		1	1
2.3 Double the agricultural productivity and incomes of small-scale food producers			0
2.4 Ensure sustainable food production systems			0
3.5 Strengthen the prevention and treatment of substance abuse		1	1
4.7 All learners acquire the knowledge and skills needed to promote sustainable development		1	1
5.2 Eliminate all forms of violence against all women and girls	1	5	6
5.a Undertake reforms to give women equal rights to economic resources			0
8.2 Achieve higher levels of economic productivity			0
8.3 Promote development-oriented policies that support productive activities			0
8.5 Achieve full and productive employment and decent work for all women and men			0
8.7 Eradicate forced labour, end modern slavery and human trafficking		2	2
10.1 Achieve and sustain income growth of the bottom 40 per cent of the population		1	1
10.2 Empower and promote the social, economic and political inclusion of all,			0
11.7 Provide universal access to safe, inclusive and accessible, green and public spaces		1	1
15.1 Ensure the sustainable use of terrestrial and inland freshwater ecosystem			0
15.2 Promote the implementation of sustainable management of all types of forests			0
15.3 Combat desertification,			0
15.4 Ensure the conservation of mountain ecosystems			0
15.7 Take urgent action to end poaching		1	1
15.9 Integrate ecosystem and biodiversity values into national and local planning			0
16.1 Significantly reduce all forms of violence	1	1	2
16.2 End abuse, exploitation, trafficking and all forms of violence against and torture of children	1	5	6
16.3 Promote the rule of law at the national and international levels and ensure equal access to justice for all	1	4	5
16.4 Significantly reduce illicit financial and arms flows and combat all forms of organized crime		1	1
16.5 Substantially reduce corruption and bribery		3	3
16.6 Develop effective, accountable and transparent institutions		1	1
16.7 Ensure responsive, inclusive, participatory and representative decision-making		0	0
16.10 Ensure public access to information and protect		1	1

fundamental freedoms, in accordance with national legislation and international agreements			
16.a Strengthen relevant national institutions to prevent violence and combat terrorism and crime	1		1
16.b Promote and enforce non-discriminatory laws and policies			0
Grand Total	4	16	34

This suggests that, if organized and applied, UNODC evaluations, like those of other organizations, could be used in the SDG review process. How this could be done, however, is something that is still being discussed at the UN system level. Partly it is because the main instrument for review, formerly called the UNDAF but now called the UN Sustainable Development Cooperation Framework, does not have a consistent evaluation process. However, one could be developed in the context of the reform of country-level activities. In practice, in countries where it has projects, UNODC is on the country teams and as such has a role in the UNSDCF and should participate in their evaluations. In fact, evaluations have been done in 19 countries since 2013 as well as 14 evaluations of regional and sub-regional projects, and eleven global projects. Forty-six of these were focused on Goal 16 subjects. This shows that UNODC can contribute to UNSDCF evaluations where its projects exist.

One essential element is to ensure that an appropriate reference to the relevant SDG and target for which the evaluation is relevant is made in all project documents and communications material (and, in fact, projects need to specify this in their design). The IES should continue to capture and catalogue data on SDG results in its Unite Evaluations Platform as a matter of routine, and have these results reflected on the UNODC website for external audiences. The current SDG [webpage](#) frames the discussion as the goals that UNODC will support but does not communicate the contributions that have and are being made. Clicking on the individual goal icons takes the reader to text-heavy pages that could more clearly articulate UNODC's important work on achieving SDG targets and goals. The UNICEF SDG [webpage](#) provides a model to consider.

Good practice example of assessing contributions to SDGs: GLO.ACT also stood out for clearly specifying the relevant goals and targets in table format and then discussing the linkages between SDG targets and intervention activities. This excerpt from the report text shows how the evaluators also took care to explain the relevance:

Targets 5.2 and 16.2 focus on addressing trafficking and exploitation of women and children which requires actors to use a gender and age-sensitive lens when addressing human trafficking. Alongside the measures highlighted above in relation to target 8.7, the focus on certain kinds of trafficking to which women, girls and boys may be particularly vulnerable, including for example GLO.ACT's activities to raise awareness of bride kidnapping, forced marriages or the recruitment of children for armed groups all point towards positive action towards achieving this target. (p. 16)

3.6 Evaluation Follow Up

One way to assess evaluation use is to look at how recommendations emerging from evaluation exercises are dealt with. For 2019, ten of the eleven evaluations had data on follow-up to the recommendations made. As can be seen from Table 11, well over half of the applicable recommendations in nine of the evaluations were accepted and if the 20% percent that were partially accepted are included, acceptance was nearly universal.

Just three recommendations were rejected. For one evaluation, GLOZ67, there was no management response and, as a result, no data are available about acceptance for these. The IISG evaluation had 11 recommendations but these were recorded as being Not Applicable.

Table 11. Status of Acceptance of recommendations, by project

	Accepted	Partially Accepted	Rejected	No Management Response	Not Applicable	Grand Total
BHUZ13	3	5	1	-	-	9
EGYZ33	3	3	-	-	-	6
GLOZ67	-	-	-	7	-	7
GLOZ99	8	3	-	-	-	11
MEXK54	3	3	2	-	-	8
MEXZ93	5	-	-	-	-	5
IISG	-	-	-	-	11	11
PSEY13	6	1	-	-	-	7
XASV23	10	-	-	-	-	10
XCAX75	10	-	-	-	-	10
XSPZ91	6	1	-	-	-	7
Grand Total	57	16	3	7	11	91
Percent	59%	18%	3%	8%	12%	100%

Two of the rejected recommendations were in the project MEXK54 on illicit growing of crops; these were methodological recommendations to the field management. However, a review of the evaluation report shows that there are 10 recommendations in the project but three are shown in the EFP tracking table as rejected, but without a reason, one with the notation “N/A” or not applicable.

The other rejection was in the project BHUZ13 on Promoting rights-based multi-sectoral responses to prevent trafficking of persons in Bhutan - Phase II. The recommendation that was rejected called for development of a strategy for efficient management, especially for more timely decision-making and disbursement of funds. It stated that this may require ROSA to delegate certain authority to POBTN to make decisions, and to provide petty cash/imprest money to cover expenses in any future project. The management response noted that “This was one-time case during the introduction of Umoja when disbursements were delayed” and for that reason was not a realistic recommendation.

The overall rate of acceptance of recommendations suggests that managers see the evaluations as being useful for programme improvement. However, a more accurate assessment would require further follow up to determine whether the recommendations had been implemented.

3.7 Evaluation Complexity

This review also looked at the level of complexity of UNODC evaluation processes. It is suggested that complexity generally involves one or more of the following factors:

- The intervention itself is complex. For example, it might have multiple and interacting components (such as training, advisory support, intergovernmental aspects), be implemented in multiple countries and levels, or the results chain may be multi-faceted or unclear.
- The composition of stakeholders creates complexity. Development partners, other duty bearers and rights-holders may have different agendas, requirements, expectations, levels of trust, or understandings of the intervention that have to be included in the evaluation, taking up additional time and requiring more complex analysis. Similarly, the number, distribution and languages of beneficiaries may create issues for sampling that need to be addressed in an environment in which resources for evaluation are limited.
- The environment in which the intervention takes place involves challenges and risks that add complexity to, and possibly disrupt or curtail, the evaluation process. For example, the dynamic

and adaptive nature of the issues UNODC is mandated to address often necessitate revisions to the design of interventions which the evaluation then has to take into account. As well, there are cases where interventions are implemented over an extended timeframe, and in environments with frequent staff rotations, making it difficult to access key informants. In some cases, there may be a need for extensive mitigation to address the probability of risks in data collection in places where the safety of stakeholders and/or evaluators is a concern.

In reviewing the 2019 evaluations, one or more of these factors were highlighted in six of the reports as contributing to the complexity of the evaluation process. These are shown in table 10⁹.

Table 10: Factors Noted as Increasing the Complexity of 2019 Evaluations

	Intervention	Stakeholders	Environment	EQA Rating
GLO.ACT	X	X		Very Good
GLOZ99	X			Very good
IISG	X	X		Very Good
MEXK54			X	Very Good
MEXZ93			X	Very Good
XCAX75	X		X	Very Good

Of these, the IISG evaluation is notable for having been particularly challenging to undertake. Among the complexity factors in this undertaking were the high number of implementors (50), the complexity of the topic itself (there was not a shared understanding amongst all stakeholders of how it worked and should work), the number of countries involved, as well as having multiple donors and their different requirements¹⁰. The graphic in figure 9, from the background section of that report, helps to illustrate stakeholder complexity.

The experience of the IISG evaluation and others, such as the evaluation of the Global Programme on Money Laundering (2012), underscore the imperative of advance planning and resource allocation. A key lesson is that resources for complexity responsive evaluation processes include not only budget, but also time, particularly for IES staff. Consideration needs to be given to pre-planning or preparation stages when substantial efforts can be put into initial discussions, including with different organizational units and development partners, to ensure a shared understanding of the intervention, and to flush out and negotiate forms of collaboration and evaluation management. Although this stage can often be labour intensive, it is generally not captured or recognized in work planning processes. More complex evaluations require significantly more time investments from IES throughout evaluation implementation phase as well.

Consideration also has to be given to the implications of complexity on evaluation quality and usefulness, particularly when the scope of the evaluation is extensive. This is a risk when multiple components are bundled into one evaluation process. In such cases, there is strong potential for learning opportunities from evaluations to be compromised, particularly when the evaluators have to narrow the scope or decrease the depth of the inquiry in order for the evaluation process to be realistic and feasible. For the 2019 reports flagged for having complexity factors, their quality does not appear to have been compromised as all were rated as Very Good.

⁹ It should be noted that other evaluations may have also had to deal with complexity issues, but that these were not apparent in the evaluation reports.

¹⁰ UNODC was commissioned by the EU to conduct this evaluation process. It was clear from discussions with IES that this evaluation required an extensive time commitment from them, particularly during the evaluation planning stages as many stakeholders were not familiar with independent evaluations and had diverse expectations and requirements of the process that had to be managed.

Figure 9: Graphic depiction of IISG stakeholders

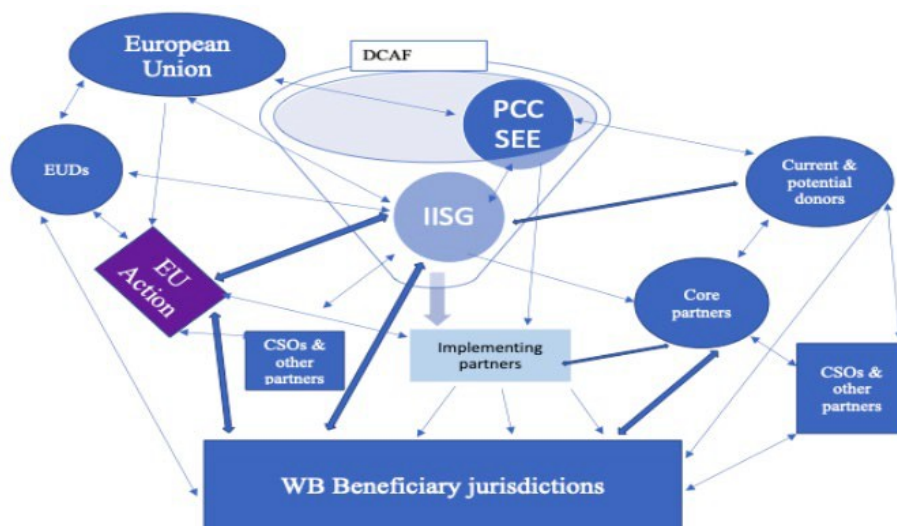


Figure 2. Regional Cooperation System in Western Balkans: A Simplified Version

Drawing from the work done on complexity and the learning from evaluations commissioned over the past few years, a basic classification and rating framework is proposed by the reviewers to guide planning, management and accountability for complexity responsive evaluations within UNODC. Being cognizant of the need for an easy to use system relevant for UNODC's work, four general categories are offered and within them are supporting questions to inform the analysis, and potentially a rating, for that category. This preliminary framework is included as Annex 3.

4. RECOMMENDATIONS

The main recommendations for the IES to consider are:

1. Continue to encourage uptake of practices raised in previous synthesis reports even though there has been considerable progress shown:
 - a. Ensure that cause-effect links are clearly reflected in the evaluations. At the institutional level, IES should continue to stress the need for all UNODC programmes and projects to have clear and up-to-date logical frameworks and theories of change. It is also recommended that the adequacy of both be assessed, and revised if necessary, during the evaluation inception phase by the evaluators in collaboration with IEU. These then need to be the basis for the evaluation findings.
 - b. Continue to place more emphasis on robust methodologies: It is important that the methodology chosen adequately facilitates answers to the evaluation questions. Evaluators should be encouraged to include more quantitative 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 more explicit about how they analyzed data from each source. Additionally, they should be encouraged to routinely disaggregate responses by stakeholder group as well as by gender - it is unlikely that all stakeholders hold the same views and the varying perspectives should be illuminated in the Findings, including through use of illustrative quotes and stories.
 - c. Encourage more explicit descriptions of gender-responsive practices: As part of the inception stage, it is suggested that IES review the relevant sections of the SWAP

scoring system with each evaluation team. Evaluators should subsequently be reminded of the need to clearly articulate, in the methodology section of the report, the steps they have taken to ensure that the evaluation process used was gender-responsive.

- d. Encourage more explicit descriptions of stakeholder participation: Good practice calls for stakeholder involvement in evaluation processes and this should be built into the evaluation design. Evaluators should also be reminded about the need to be explicit about how the Core Learning Partners (and other stakeholders as relevant) have been involved in the evaluation, including their contribution to the development of the report's conclusions and recommendations. Stakeholder mapping processes during the Inception Phase should also be encouraged, with the results being reflected in the evaluation report.

2. Addressing other ongoing issues that affect EQA scores:

- a. Visual aids: Although evaluators are increasingly using data visualization techniques, more could be done in this regard, including to sharpen the quality and communication value of the graphic images.
 - b. Dates/timeframe of the evaluation process: This information is frequently not provided and should be.
3. Improve assessment of SDGs: UNODC should continue to take a lead among members of UNEG in showing how evaluation can improve the assessment of the SDGs. This can involve working to improve indicators that are currently Tier II or III, or where the current indicators do not measure what the target should achieve. IES should also work with country-teams to ensure that evaluations are consistent and that SDGs for which UNODC is responsible are included. A re-design of the UNODC SDG webpages - to more succinctly communicate agency contributions to each relevant goal and highlight examples of interventions - may make this work more visible and increase awareness of its importance.
4. Inclusion of Ethical Considerations and Evaluation Principles: An element that is not captured in UNODC's template, but is in other agency quality assurance processes, is reference to UNEG Ethical Guidelines and Code of Conduct for Evaluation in the UN System and to the agency's own evaluation policy and guidance. While these are set out in evaluation ToRs, evaluators may need to be reminded that although the guidelines and code are imperative for work with vulnerable populations, they apply to all evaluations. It is suggested that within the methodology section evaluators confirm, and articulate how, they have abided by these guidelines and the obligations (independence, impartiality, credibility, conflicts of interest, accountability), as well as describe the ethical safeguards followed in collecting data from/with evaluation participants. These considerations should also be incorporated into the EQA template.
5. Follow up on the implementation of recommendations: The tracking of management responses is important since this is one way to capture use of evaluations. To the extent feasible, it is suggested that there be further follow up to look at the percent of recommendations that were implemented from previous year's evaluations. If programme managers are aware that recommendations will be tracked and the results reflected in the EQA Synthesis Report, more attention may be given to this process.
6. Further refinement of the EQA template:
- a. Add Coherence as part of required evaluation criteria;
 - b. Adjust the sections to separate Scope from Methodology in order to align with other UN agency EQA processes which have Methodology as its own section (Scope should then be included with Context & Purpose);

- c. Adjust the question on evaluation team members to include subject matter expertise;
- d. Separate the questions for data collection and data analysis to better highlight the need for clearly described analysis processes;
- e. Incorporate reference to UNEG ethical guidelines and principles as discussed above.

7. Addressing complexity:

- a. More attention should be given to IES capacity to effectively support and maintain the quality of evaluations as the complexity of evaluation processes increases;
- b. Consider expanding the scope of the EQA Synthesis Report to look further at the impact of complex evaluation processes, including how learning opportunities may be compromised.

Annex 1. List of evaluations reviewed

All reports are published on UNODC website at www.unodc.org/unodc/en/evaluation/reports.html

Project Number	Project Title
MEXK54	Evaluación independiente inicial del proyecto MEXK54 “Sistema de Monitoreo de Cultivos Ilícitos en el Territorio Mexicano”
MEXZ93	Fortalecimiento para la Seguridad de Grupos en situación de vulnerabilidad
PSEY13	Supporting the establishment of evidence-based drug dependence treatment and rehabilitation system for the Palestine National Rehabilitation Centre
EGYZ23	Improving the Criminal Justice Response to Violence against Women in Egypt
GLO.ACT	Global Action against Trafficking in Persons and Smuggling of Migrants
GLOZ99	Global Programme segment - Asia-Pacific Joint Action Towards a Global Regime against Corruption (2016-2020)
XCAX75	Proyecto: Apoyo al A.B.1. Coordinación Interinstitucional y Regional para la Seguridad Fronteriza en Centroamérica
BHUZ13	Enhance Government and Civil Society Responses to Counter Trafficking in Persons in Bhutan
ISSG	Western Balkans Counter-Serious Crime Initiative (WBCSCi) in the context of the Integrative Internal Security Governance (IISG) mechanism including the European Union action
XASV23	Support Project for the SADC-UNODC Regional Programme on Making the SADC Region Safer from Drugs and Crime, with the specific focus on Violence against Women and Children
XSP/Z91	Joint Mid-Term Independent Project Evaluation of the United Nations Pacific Regional Anti-Corruption Project (UN-PRAC)

Annex 2. Proposed Complexity Assessment Framework for Evaluation Planning

1. Level of ambition of intervention outcomes - how complex is the intervention in terms of what it is trying to achieve and the number of expected results that have to be measured? Are there a number of components (such as multiple substantive areas, a combination of strategy and project-level initiatives) bundled into the evaluation? Is the intervention expected to effect change at different levels (i.e. global, regional, national levels)? Is the intervention and its intended results well understood? In other words, will it be difficult to measure how it works, and therefore what is the likelihood that a more robust evaluation design required given the number of expected results?

high – above average – average for UNODC – less than average - low

2. Nature of stakeholder involvement - how complex is the intervention in terms of the number and diversity of participants? Are there multiple development partners, and will they have different expectations or requirements of the evaluation process? Are there a large number of implementation partners involved? Are there logistical constraints to their participation in the evaluation? Are there extraordinary issues requiring trust building with stakeholders or of confidentiality related to UNODC's mandate? Are there key partners with little or no experience with independent evaluation processes? Is a large sample of respondents needed, and to what extent are they dispersed?

high – above average – average for UNODC – less than average - low

3. Level of risk - How complex is the evaluation environment? Are evaluation activities likely to be carried out in locations with safety and security concerns? Will technology tools be needed to help overcome issues of access? Are there significant political, health, social, cultural, organizational or other sensitivities/dynamics to be taken into account that might otherwise disrupt the evaluation process? Is the evaluation budget commensurate with the scope of the assignment? What is the likelihood that adequate data will not be available for the level of assessment required?

high – above average – average for UNODC – less than average - low

4. Extent of IES support - What level of effort is required from IES to adequately support the evaluation process? Based on the preceding factors, what are the human resource and other investments needed for ensuring a quality evaluation?

high – above average – average for UNODC – less than average - low

Annex 3. UNODC EQA Template as Used in the Review

UNODC Evaluation Quality Assessment (EQA)

version: December 2019

General Project Information			
Project/Programme Number and Name			
Thematic Area			
Geographic Area (Region, Country)			
Relevant SDG(s)			
Approved Project/Programme Budget at the Time of the Evaluation (USD)			
Type of Evaluation (In-Depth/Independent Project; Final/Midterm; Other)			
Cost of Evaluation (USD)			
Evaluation Team (# of men/# of women; names of team members)			
Evaluation Complexity Factors (if any)			
Date of Evaluation (from MM/YYYY to MM/YYYY)			
Date of Evaluation Report (MM/YYYY)			
Quality Assessment Conducted On/By			
EQA Provided for Draft Report (Y/N; if Y, indicate rating)			

OVERALL QUALITY RATING:	
SUMMARY:	

Quality Assessment Criteria	Yes No Partial	Assessment Levels: Very Good - Good - Fair - Unsatisfactory Meets Criteria: Y = Yes N = No P = Partially N/A- Not Applicable	
1. Structure, Completeness And Clarity Of Report		RATING:	Very good
a. Format (headings, font) accords to IEU Guidelines and Templates for Evaluation Reports.	Yes		
b. Structure accords to IEU Guidelines for Evaluation Reports with the following logical sequence: List of acronyms; Executive Summary; Summary Matrix of Findings, Evidence and Recommendations; Introduction (Background and Context, Evaluation Scope and Methodology, Limitations to the Evaluation); Findings (Relevance, Efficiency, Partnership and Cooperation, Effectiveness, Impact, Sustainability, Human Rights and Gender Equality/Mainstreaming, as well as Design and Innovation if in ToR); Conclusions; Recommendations; Lessons Learned.	Yes		
c. Objectives stated in the terms of reference are adequately addressed.	Yes		
d. Report is easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience).	Yes		
e. Language is empowering and inclusive avoiding gender, heterosexual, age, cultural and religious bias, among others.	Yes		
f. Report is generally free from grammar, spelling, or punctuation errors.	Yes		
g. Visual aids, such as maps and graphs, are used to convey key information.	Yes		
h. 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.	Partial		

i. Annexes include at a minimum: evaluation terms of reference; logic model and/or evaluation matrix; list of persons interviewed and sites visited; list of documents consulted; evaluation tools used.	Yes	
2. Executive Summary		RATING: #DIV/0!
a. Written as a stand-alone section that provides an overview of the evaluation and presents its main results.	<Select one>	
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.	<Select one>	
c. Summary Matrix presents only the key and most important recommendations from evaluation report.	<Select one>	
d. Findings, sources and recommendations in the Summary Matrix are clear and cohesive, and specify the stakeholder to whom they are addressed.	<Select one>	
e. Maximum length 4 pages, excluding the Summary Matrix.	<Select one>	
3. Evaluation Context And Purpose		RATING: #DIV/0!
a. Clear description of the project evaluated is presented.	<Select one>	
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.	<Select one>	
c. Connection with Sustainable Development Goals is clear.	<Select one>	
d. 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.	<Select one>	

e. Project status is described including its phase of implementation and any significant changes (e.g. to strategies, logical frameworks) that have occurred.	<Select one>	
f. Purpose of evaluation is clearly defined, including why it was needed at that point in time, what information is needed, and the target audience.	<Select one>	
4. Scope And Methodology		<div>RATING:</div> <div>#DIV/0!</div>
a. Evaluation scope is clearly explained including the main evaluation criteria, questions and justification of what the evaluation did and did not cover.	<Select one>	
b. Transparent description presented of methodology applied, including how it was designed to address the evaluation purpose, objectives, questions and criteria.	<Select one>	
c. Methodology allows for drawing causal connections between outputs and expected outcomes.	<Select one>	
d. Methods are appropriate for analysing gender equality/mainstreaming and human rights issues identified in evaluation scope; methodology takes into account power relations during an evaluation process; is inclusive and participatory.	<Select one>	
e. Data collection methods and analysis, and data sources are carefully described, as are the rationale for selecting them. Processes are adequate for measuring outcomes set out in logic model. Reference indicators and benchmarks are included where relevant.	<Select one>	
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.	<Select one>	

g. High degree of participation of internal and external stakeholders, including the Core Learning Partners, throughout the evaluation process is planned for and made explicit, particularly in the development of the evaluation design, conclusions and recommendations.	<Select one>	
5. Reliability of Data <i>To ensure quality of data and robust data collection processes</i>		#DIV/0!
a. Triangulation principles (using multiple sources of data and methods) were applied to validate findings.	<Select one>	
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.	<Select one>	
c. Limitations that emerged in primary and secondary data sources and collection processes (bias, data gaps, etc.) are identified and, actions taken to mitigate such issues are explained.	<Select one>	
d. Evidence provided of how data was collected with a sensitivity to issues of discrimination and other ethical considerations.	<Select one>	
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.	<Select one>	
6. FINDINGS AND ANALYSIS <i>To ensure sound analysis and credible findings</i>		RATING: #DIV/0!
<i>Findings</i>		-
a. Are clearly formulated and presented	<Select one>	
b. Are based on rigorous analysis of the data collected; take into account any identified benchmarks.	<Select one>	

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.	<Select one>	
d. Address any limitations or gaps in the evidence and discuss any impacts on responding to evaluation questions raised in ToR.	<Select one>	
e. Include findings related to the SDGs covered by the evaluation.	<Select one>	
f. Discuss any variances between planned and actual results of the project (in terms of objectives, outcomes, outputs).	<Select one>	
<i>Analysis</i>		
a. Interpretations are based on carefully described assumptions.	<Select one>	
b. Contextual factors are identified (including reasons for accomplishments and failures, and continuing constraints).	<Select one>	
c. Cause and effect links between an intervention and its end results (including unintended results) are explained.	<Select one>	
d. Includes substantive analysis of human rights issues.	<Select one>	
7. CONCLUSIONS		RATING:
		#DIV/0!
a. Take into consideration all evaluation criteria and questions, including human rights and gender equality/mainstreaming criteria.	<Select one>	
b. Have been formulated clearly, are based on findings and substantiated by evidence collected and go beyond the findings and provide a thorough understanding of the underlying issues of the project and add value to the findings.	<Select one>	
c. Convey the evaluators' unbiased judgement of the intervention.	<Select one>	

d. Present a comprehensive picture of both the strengths and weaknesses of the project.	<Select one>	
8. RECOMMENDATIONS		RATING: #DIV/0!
a. Are clearly formulated, based on the conclusions, and substantiated by evidence collected.	<Select one>	
b. Address flaws, if any, in project's data acquisition processes.	<Select one>	
c. Are specific, realistic, indicate a time, are actionable, and of a manageable number.	<Select one>	
d. Are clustered and prioritized.	<Select one>	
e. Include recommendations related to the SDGs covered by the evaluation.	<Select one>	
f. Reflect stakeholders' consultations whilst remaining balanced and impartial	<Select one>	
g. Clearly identify who is responsible for action.	<Select one>	
9. LESSONS LEARNED		RATING: #DIV/0!
a. Are clearly identified, innovative and add value to common knowledge.	<Select one>	
b. Are based on specific evidence and analysis drawn from the evaluation.	<Select one>	
c. Have wider applicability and relevance to the specific subject and context.	<Select one>	
10. ASSESSMENT OF THE INTEGRATION OF GENDER EQUALITY AND EMPOWERMENT OF WOMEN (GEEW) for UN-SWAP		RATING: #DIV/0!
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.	<Select one>	

b. Gender-responsive evaluation methodology, methods and tools, and data analysis techniques are selected.	<Select one>	
c. Evaluation findings, conclusions and recommendations reflect a gender analysis.	<Select one>	
Overall Score	#DIV/0!	
Overall Rating	#DIV/0!	

SCORING

Assessment Levels: Very Good (90%+) Good (70-89%) Fair (50-69%) Unsatisfactory (<50%)

Element Of The Evaluation	Points Per Category	Average score	Weighted score
Presentation And Completeness	10	0.94	9.44
Executive Summary	5	#DIV/0!	#DIV/0!
Evaluation Context And Purpose	5	#DIV/0!	#DIV/0!
Evaluation Scope And Methodology	10	#DIV/0!	#DIV/0!
Reliability Of Data	5	#DIV/0!	#DIV/0!
Findings And Analysis	30	#DIV/0!	#DIV/0!
Conclusions	10	#DIV/0!	#DIV/0!
Recommendations	15	#DIV/0!	#DIV/0!
Lessons Learned	5	#DIV/0!	#DIV/0!
Gender Equality and Empowerment of Women	5	#DIV/0!	#DIV/0!
Total Maximum Score	100		#DIV/0!
GEEW scoring: 0 - 3 points = Misses requirements / 4 - 6 points = Approaches requirements / 7 - 9 points = Meets requirements			

Independent Evaluation Section
United Nations Office on Drugs and Crime
Vienna International Centre
P.O. Box 500
1400 Vienna, Austria
Telephone: (+43-1) 26060-0
Email: unodc-ies@un.org
Website: www.unodc.org/unodc/en/evaluation/index.html