



EVALUATION

INDEPENDENT QUALITY ASSESSMENT OF UNODC EVALUATION REPORTS 2021 SYNTHESIS REPORT



UNODC
United Nations Office on Drugs and Crime

INDEPENDENT EVALUATION SECTION
<https://www.unodc.org/unodc/en/evaluation/> • unodc-ies@un.org

CONTENTS

EXECUTIVE SUMMARY	3
INTRODUCTION	4
METHODOLOGY	4
FINDINGS	6
EQA RATINGS	6
GENERAL STRENGTHS AND IMPROVEMENTS OF THE EVALUATION REPORTS	8
GENERAL WEAKNESSES OF THE EVALUATION REPORTS	11
MAINSTREAMING HUMAN RIGHTS AND GENDER CONSIDERATIONS	15
ASSESSING CONTRIBUTIONS TO SDGs	19
EVALUATION COMPLEXITY, INCLUDING COVID-19	19
PILOT INDEPENDENT PROJECT EVALUATIONS	21
CONCLUSIONS AND RECOMMENDATIONS	23
ANNEX 1 LIST OF EVALUATIONS REVIEWED	
ANNEX 2 UNODC EQA TEMPLATE AS USED IN THE REVIEW	
ANNEX 3 UNODC EQA TEMPLATE FOR PILOT IPES	

EXECUTIVE SUMMARY

UNODC's annual Evaluation Quality Assessment (EQA) Synthesis report provides an overview of the quality of all published independent evaluations commissioned by the agency each year. Each evaluation report is reviewed and assessed using an EQA template that aligns with UNODC and UNEG norms and standards for evaluations. The template is reviewed on an annual basis to improve its functionality and to incorporate new and emerging good evaluation practice.

In 2021, there were 15 evaluations assessed. Four were In-depth Evaluations (IDEs), ten were Independent Project Evaluations (IPEs), and one was a pilot evaluation. The pilot evaluation followed the newly introduced approach to IPEs which calls for a much more concise assessment process and briefer report than has been the case. It was not included in the synthesized results, with the exception of the scores for gender equality, but is discussed in a stand-alone section of this document.

The performance of the 2021 portfolio of evaluations was strong with 43% (6 of the 14) receiving an overall rating of Very Good and the remainder rated as Good. As such, all met the standards for UNODC evaluations and can be used by decision makers with confidence. The overall scores were higher than previous years across five of the ten sections of the template. Reliability of Data, Conclusions, and Lessons Learned were the strongest sections. The most significant improvement was in the integration of gender – UNODC reports already performed well here, with average scores for the UN System Wide Action Plan (UN-SWAP) on Gender Equality and the Empowerment of Women (GEEW) hovering around 7 (on a scale of 1 to 9) for the past four years. This year the average score rose significantly to 7.7. Also notable was that all evaluations used the latest evaluation report template which elevated the level of presentation of the whole portfolio.

The sections that scored lower than the above were the Executive Summary, Design & Methodology, Context/Purpose/Scope, and to a lesser extent, Recommendations. This does not mean that the whole sections had shortcomings, but that there were elements within those sections that could be improved in several reports. The more significant issues were in meeting the expectation for clear descriptions of the context of the intervention, stakeholder involvement in the intervention and the evaluation process, evaluation scope, and some aspects of methodology including reference to ethical considerations that were adhered to in the evaluation process.

All the evaluation reports considered the cross-cutting themes of Human Rights, Gender Equality and Leaving No One Behind (LNOB) and referred to the Sustainable Development Goals (SDGs), although there was variation in the extent of analysis of each theme. There was more difficulty in meeting the requirements of the UN System-wide Disability Action Plan. Only two of the reports could be classified as disability inclusive, the same number as in 2020.

Evaluators and evaluation managers faced multiple challenges in conducting evaluations during the continuing pandemic, including in engaging rightsholders as evaluation participants. Almost all reports had an adequate level of explanation of COVID-19-related methodological restrictions and the mitigation methods used. Despite the limitations, overall, evaluators were able to carry out adequately robust data collection enabling the presentation of evidence-based findings.

This analysis has three recommendations which are explained in more detail at the end of the report.

1. Continue to push the quality of evaluations particularly by ensuring that: background, stakeholder involvement, ethical considerations, and disability inclusion are adequately addressed; evaluation matrices are attached to the reports; the theory of change/results frameworks are used as part of the analysis; and the presentation of Recommendations aligns with requirements.
2. Further consideration be given to the maximum page lengths for each type of report and to the number of evaluation questions to be addressed.
3. Further refinements be made to the review template to ensure the basis by which the evaluation assesses programme/project performance is clear, and that the EQA process adequately reviews disability inclusion.

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.

Since 2014, the IES has commissioned independent evaluation quality assessments of UNODC evaluation reports to help ensure these products are providing credible information to inform planning processes. 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 2021 (the list of reports is provided in Annex 1) and makes comparisons with EQAs since 2014/15.

The reviewers were requested to pay specific attention to the extent to which the cross-cutting themes of human rights, gender equality, leave no one behind and disability inclusion were considered as part of evaluation processes. This report includes a separate section on each theme, as well as on the coverage of SDGs and the effects of COVID-19. It also includes an initial analysis of the extent to which the new approach to IPEs meets evaluation quality standards.

This report makes extensive use of visual aids to convey the results of the analysis. The authors are responsible for all the visual aids presented in this document unless another source is shown.

METHODOLOGY

EQA Template: The first deliverable on the annual evaluation quality assessment review was to assess the EQA template used for the 2020 reports and to propose any needed modifications. The main considerations were to ensure that the template aligned with the new Administrative Instruction on Evaluation in the United Nations Secretariat ([ST/AI/2021/3](#)) and the [accompanying guidelines](#), as well as the new limits on the page lengths for UNODC evaluation reports. The resulting revisions, which were minor compared to the more substantial changes made the previous year, included:

- Intervention budget becoming a requirement as part of Background/Context; timeframe and geographical coverage being required as part of evaluation Scope (to align with Administrative Instructions).
- Environmental considerations and use of right-based terminology will not be required until guidance is developed for each.
- Visual aids require title and source to be specified and limiting the number of recommendations to a

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

maximum of 10 (to align with the expectations set out in the UNODC evaluation report template).

- Maximum report page lengths being reduced for all regular evaluations (does not exceed 10% of the maximum mandatory length of 30 pages/up to 100,000 characters including spaces) excluding the Executive Summary and annexes.
- Limitations required to include constraints encountered due to the pandemic.
- Overall rating of GEEW section aligned with UN-SWAP rating (i.e., the highest rating of Very Good previously required a score of 8 or 9, but now accords to the highest band for UN-SWAP which is 7 to 9).
- Other minor changes for the purposes of clarifying elements and reducing redundancies in the template.

The final version of the template used for the 2021 review is attached as Annex 2.

2021 EQA Process: The reviewers then examined the quality of all the evaluation reports published during 2021. The total number of evaluations for the year was 15. Four were in-depth evaluations, ten were independent project evaluations, and one was a pilot of the new approach to IPEs. One area that was rated more thoroughly this year was at the extent to which evaluations were disability inclusive. This was a new requirement for all evaluations conducted within the UN system in 2020 (as per the UN System-wide Disability Action Plan), and the expectation was for 2021 evaluations to assess more thoroughly the extent to which interventions were disability inclusive and for evaluation processes themselves to be inclusive of people with disabilities.

Limitations: There are two limitations in presenting this analysis, particularly in drawing comparisons with previous years. One is the small number of evaluation reports which enables a single evaluation report to have a significant impact on average ratings. The second is the adjustments made to the template each year which have an impact on comparability. As such, the cumulative ratings of any one year are not necessarily indicative of an overall trend. It is also important to recognize that all the 2021 evaluations were conducted during the COVID-19 pandemic which added to the challenges of conducting robust processes.

Review of Revised Approach to IPEs: The reviewers also developed a revised version of the template for the more concise independent project evaluations that were being piloted in 2021-2022. This template is attached as Annex 3. The one pilot evaluation that was completed in 2021 was reviewed separately from the regular portfolio of evaluations, with the results being discussed in a subsection on page 22.

FINDINGS

EQA Ratings

The portfolio of evaluations commissioned by UNODC in 2021 continued to be of good quality, useful for decision making processes and well aligned with OECD-DAC and UNEG norms for evaluation practice. Figure 1 shows that of the fourteen reports, six (43%) received a Very Good rating and eight reports (57%) were rated as Good. None were rated as Fair or Unsatisfactory, as was the case in 2018 and 2019 (see figure 2).

Figure 3 shows that the percent of Very Good reports has slightly increased from 2020. This continued the trend whereby in most years an increasing percent of reports achieved the highest rating, apart from a spike in quality in 2019. Improvements have happened even though regular changes to the template raise the bar for quality. There were substantial changes made in 2020 which was a factor in the drop in the percent of Very Good reports after 2019.

Figure 1: Overall rating of 2021 reports (# of reports)



Figure 2: Percent of combined Very Good and Good reports by year

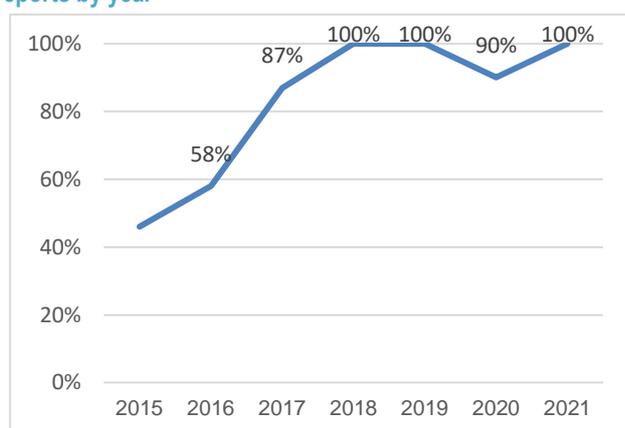


Figure 3: Percent of Very Good reports by year

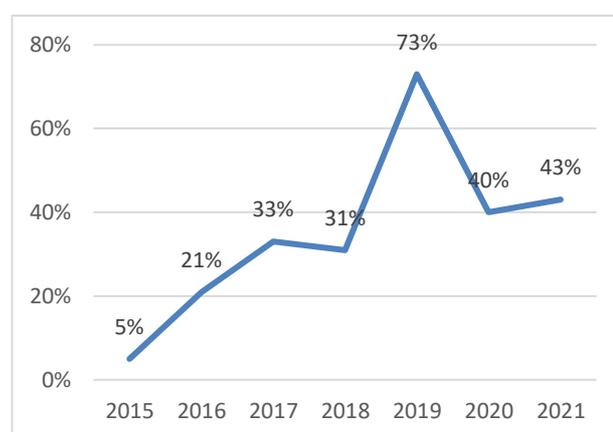


Table 1 shows the number of reports that have been commissioned annually since 2015 and their ratings. The 14 regular evaluation reports in 2021 show an increase compared to the previous two years. When the one pilot IPE from 2021 (rated as Good) is also taken into account, the total is close to the average number of 16 reports commissioned in the period between 2016 – 2018. The 2015 batch was larger as it included some reports from 2014 as well.

Table 1: Overall rating of 2021 reports compared with previous EQA cycles

	Unsatisfactory	Fair	Good	Very Good	Total
# of Reports – 2021	0	0	8	6	14
% of Reports – 2021			57%	43%	
# of Reports – 2020	0	1	6	4	11
% of Reports – 2020		9%	55%	36%	
# of Reports – 2019	0	0	3	8	11
% of Reports – 2019		0%	27%	73%	

# of Reports – 2018	0	0	11	5	16
% of Reports – 2018		0%	69%	31%	
# of Reports – 2017	0	2	8	5	15
% of Reports – 2017		13%	54%	33%	
# of Reports – 2016	0	8	7	4	19
% of Reports – 2016		42%	37%	21%	
# of Reports – 2015 ²	0	12	9	1	22
% of Reports – 2015		53%	41%	5%	

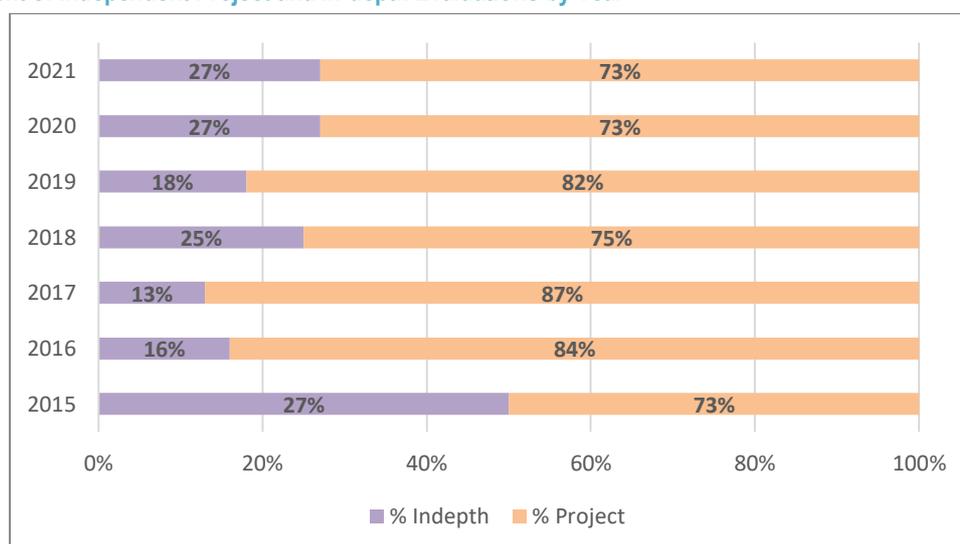
Comparability with other UN entities: Direct comparisons in evaluation quality are difficult to make as each entity has its own criteria in addition to the standards set out by UNEG and OECD-DAC, as well as different weighting and scoring processes. Thresholds for overall rating categories also differ. For example, UNODC has a relatively high bar for evaluations rated above Good. While UNODC requires a total percent score of 90+ for Very Good, the same threshold for another entity is 85%. In a third case, evaluations are rated as Highly Satisfactory with a total score of 87.5% +, and there is an additional category of Exceptional for those evaluations that score 96% and higher. In 2021, of the eight UNODC evaluations rated as Good, four of them scored in the band between 85-89%.

Table 2 shows that there is a clear difference in quality in respect to the different types of reports - Independent Project Evaluations (these were all final evaluations) and programmatic In-Depth Evaluations (these included by mid-term and final evaluations). All of the four in-depth evaluations (2 mid-term and 2 final) were rated as Good, whereas six out of 10 independent project evaluations were rated as Very Good. The IDEs all had one section rated as Fair – either Executive Summary or Design & Methodology. Only one of the four reports had a Very Good rating for Findings, the highest weighted section, and this was primarily because SDGs and the full set of cross-cutting issues were not fully addressed.

Table 2: Rating by type of evaluation in 2021

Type of evaluation	Good	Very Good	Grand Total
In-Depth Evaluation Final	2		2
In-Depth Evaluation Mid-term	2		2
IPE Final	4	6	10
Pilot IPE	1		1
Grand Total	9	6	15

Figure 4: Percent of Independent Project and In-depth Evaluations by Year



The number of each type of evaluation carried out each year in 2020 and 2021 remains same with 27% in-depth evaluations and 73% independent project evaluations (figure 4 above).

The type of evaluation is one factor in assigning the budget for the evaluation. In 2021, the cost of the evaluation ranged from USD 23,000 to 170,000 for interventions that ranged from USD 619,000 to almost USD 185,039,803. As was the case in 2020 when this analysis was first conducted, there was not a clear connection between the evaluation budget and the quality rating of the evaluation.

Table 3 shows the scores by thematic area. There were evaluations carried out of interventions in seven areas. Two evaluations covered multiple themes: [UNODC Programming in West and Central Asia](#) and [GLOX34 Global Firearms](#). The thematic areas of Countering Corruption and Justice each had the most reports at three. However, again, there were too few reports in any of the areas to draw conclusive findings about which had the strongest reports.

Table 3: Rating by theme of project in 2021

	Good	Very good	Grand Total
Countering corruption	1	2	3
Countering the world drug problem	1		1
Countering transnational organized crime	1		1
Justice	1	2	3
Multiple areas	2		2
Policy Support		1	1
Research, trend analysis, and forensics		1	1
Terrorism prevention	2		2
Grand Total	8	6	14

General Strengths and Improvements of the Evaluation Reports

Overall, the strengths found in the past two years have carried over to 2021. These include Reliability of Data, Conclusions, and Lessons Learned especially, but also Findings and consideration of GEEW. Table 4 shows the full range of ratings for each criterium in the 2021 reports. Six of the 14 reports (43%) rated Good or Very Good in all sections. Nine of the 14 reports (64%) rated Good or Very Good in eight of the ten sections. However, this shows a slight drop from the 2020 reports when 89% were rated as Good or Very Good in eight of the ten sections.

Table 4: Criteria ratings for 2021 reports

	Unsatisfactory	Fair	Good	Very Good
Structure		0	10	4
Executive Summary		3	9	2
Context, Purpose, Scope		2	10	2
Design and Methodology		4	10	0
Reliability of Data		0	1	13
Findings & Analysis		0	5	9
Conclusions		0	2	12
Recommendations		0	14	0
Lessons Learned		0	4	10
Consideration of GEEW		0	5	9

Table 5 looks at the ratings by criteria in 2020 and 2021. It shows that in 2021 there were further improvements in all the five criteria mentioned above that were already strong. The GEEW scores appear to have the most significant increase, however this is primarily due to the change in how scores are assigned for this criterion in the template.

Table 5: Report rating by criteria, 2020 and 2021

	Unsatisfactory		Fair		Good		Very Good	
	2020	2021	2020	2021	2020	2021	2020	2021
Structure	0	0	2	0	6	10	3	4
Executive Summary	1	0	1	3	5	9	4	2
Context, Purpose, Scope	0	0	4	2	3	10	4	2
Design & Methodology	1	0	2	4	8	10	0	0
Reliability of Data	0	0	0	0	3	1	8	13
Findings & Analysis	0	0	1	0	4	5	6	9
Conclusions	0	0	1	0	3	2	7	12
Recommendations	0	0	1	0	9	14	1	0
Lessons Learned	0	0	2	0	1	4	8	10
Consideration of GEEW	0	0	1	0	8	5	2	9

The remainder of this section briefly looks at the elements that stood out in the highest performing sections, and the reports that exemplified good practice.

Reliability of Data: 13 of the 14 reports (93%) commissioned in 2021 were rated as Very Good for this section, compared to 72% in 2020. Mixed methods were consistently used, as were both qualitative and quantitative data sources, which are important for enabling triangulation. [XACZ61 Prevention of Radicalization Kyrgyz](#) and [GLOU68 Civil Society Africa](#) were particularly notable for strong use of multiple methods and for triangulation being clearly evident in the presentation of Findings.

Conclusions: This section showed similar increase in performance and was consistently clearly presented. This is in large part due to all reports having Conclusions organized by criteria. There was just one report where this practice wasn't consistently followed. This was an evaluation of an intervention funded by the Peace Building Fund (PBF); the OECD-DAC criteria were all reflected in the conclusions, however the PBF-specific criteria, although covered in Findings, were not addressed in Conclusions.

Lessons Learned & Good Practices: Most evaluations continued to provide lessons and practices that appeared to be useful for informing future programming. Those provided in [XACZ61 Prevention of Radicalization Kyrgyz](#) stand out especially for being instructive. For reports where Partials were given, this was usually because better formatting – use of numbering and headings - would make them more easily identifiable. The [GLOZ31 Wildlife in Asia](#) report is notable for the way in which the topics for Lessons Learned and Good Practices are highlighted.

Design & Methodology: Although the overall scores here were lower than other sections, there were also some notable improvements seen in several of the criteria. Data analysis and sampling processes were both better covered than previously with Table 6 showing the difference in these rating compared to 2020. For data analysis the percent of reports rated as Yes increased to 79% from 55%, while adequacy of explanations for sampling increased to 71% from 64%.

Table 6: Scores for data analysis and sampling in 2020 and 2021

	Year	Yes	Partial	No
Methods of analysis for all types of data are appropriate and clearly described . . .	2020	55%	27%	18%
	2021	79%	21%	0%
Sampling frame is clearly described . . .	2020	64%	36%	0%
	2021	71%	7%	21%

Source: includes data from UNODC 2020 EQA Synthesis Report

Despite the limitations on conducting evaluations during the pandemic and the need for remote process, most evaluators were able to carry out adequately robust data collection, which enabled the presentation of evidence-based and triangulated findings. Some of the evaluations with solid methodological approaches are included in the box below.

Good practice examples of methodological approaches:

1819U Legal Aid Western Africa: This is a good example of a theory-based evaluation. The approach included a workshop to develop a theory of change that was then used to guide the evaluation processes. Even though data collection was conducted remotely, a good range of stakeholders were involved, including rightsholders. The methodology was robust and carefully explained. The findings were well evidenced and led to a useful set of recommendations, lessons and good practices. The report is notable for the inclusion of a theory of change, logical framework and evaluation matrix.

Programming in West and Central Asia: This was one of the most complex in-depth evaluations due to the number of programmes and countries covered. Despite a number of challenges, the team managed to obtain input from 250+ stakeholders via interviews, an email questionnaire and a staff survey. There were four national case studies shown separately.

Counter Terrorism Maldives: The overall design and methods are explained. Over 200 documents were reviewed, and the evaluators managed to consult 57 people from the range of stakeholder groups, well over a third of which were from government (rightsholders). The analysis was able to show a number of achievements in line with the programme logic, and where contextual factors hindered reaching other result.

GLOU68 Civil Society Africa: The evaluators used a range of methods to collect qualitative and quantitative data from all stakeholder groups. A multi-staged process was used whereby an adapted version of Most Significant Change informed the selection of cases for further examination. The basis for triangulated Findings is clearly shown, with the evaluation being notable for presenting the perspectives of different stakeholders. The evaluation is also exemplary for using well-formatted and sourced visual aids, and for its solid analysis of human rights, gender equality, leave no one behind, and disability inclusion.

A new criterion incorporated into the template in 2020 was use of innovative practice in the evaluation and/or reporting process. Innovation is broadly defined, and the box below highlights some of the ways in which it was evident in the 2021 evaluations.

Good examples of innovative evaluation practice:

BOLZ68 Cultivos de Coca: One innovation was the use of change stories ("historias de cambio") as a mitigation strategy for not being able to engage directly with rightsholders during the pandemic. These are narrations written by duty-bearers who were part of the project, and which describe the changes that the project has given to the protagonists. Six of these were included in the report.

XACZ61 Prevention of Radicalization Kyrgyz: Innovation was seen in terms of the amount of translation that needed to be done (70% of documents were in Russian and translated as feasible using DeepL) and the ability of the team to engage 10 probation clients via interviews on WhatsApp.

GLO/R35 Legal Regime Against Terrorism: Innovation was evident in terms of the number of different types of data collection (in addition to document review, KIIs and FGDs, there were multiple online surveys, event observations and case studies). The use of Nvivo to analyze qualitative data also points to a more systematic level of analysis than is seen in some other evaluations.

General Weaknesses of the Evaluation Reports

The lower scoring sections tended to be the same as in previous years: the Executive Summary, Design & Methodology, Context/Purpose/Scope, and to a lesser extent in Recommendations. For learning purposes, this section takes a closer look at performance within each of these areas by providing tables showing the respective questions from the EQA template and the number of reports that scored as Yes, Partial and No for each. In some cases, the questions have been abbreviated to save space and these are indicated by “. . .”; the full questions can be found in **Annex 2**.

Executive Summary: This is arguably the most important part of the evaluation report as it is often the only section that is read in full. There was a slight drop in the quality of the 2021 Executive Summaries with just two reports rated as Very Good compared to four in 2020. As shown in table 7, Summaries were consistently clearly presented (Q a) and all but four were concisely written and within the maximum page length for the narrative portion (Q d). However, as with previous years, the lower marks had to do with structure (Q b), specifically the intended audience not being specified. Although this may not seem important, it is useful information for conveying the contexts in which the reports are being read. Another issue was seen in the summary matrix (Q c) whereby the text in the data sources column was often generically stated as "interviews, desk review, field observation", and with the same information appearing for each recommendation. For that column to be meaningful, the sources should be more tailored, for example, by specific stakeholder group and type of documents. A report that does well in differentiating data sources is [GLOZ31 Wildlife in Asia](#). A further concern seen in several reports was that the matrices extended over more than three pages (in one case, 6 pages in landscape view) which counters the intent to have concise Executive Summaries.

Table 7: Cumulative Ratings for Executive Summary Questions

Executive Summary	Yes	Partial	No
a. Clearly presented and serves as a stand-alone section that provides an overview of the evaluation and its main results.	13	1	0
b. Generally follows the structure of: i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology); iv) Main Findings/Conclusions; . . .	4	10	0
c. Summary Matrix presents only the key and most important recommendations from evaluation report. Findings, sources and recommendations are clear and cohesive, and specify the stakeholder to whom they are addressed.	8	6	0
d. Maximum length of 4 pages, excluding the Summary Matrix.	10	2	2

Context, Purpose, Scope: As can be seen in table 8 below, one issue was that the expected chain of results (Q b) – which could have been done in form of a narrative description, theory of change or logical model – was not always presented even though it appeared that most of the evaluated interventions did have these in place. In some cases, the evaluators provided a critique of the programme logic and/or suggested improvements but did not include the logic in the report. Ideally, evaluators should ensure that at least a basic results chain is established to provide a framework for measuring the success of the intervention. [GLOU68 Civil Society Africa](#) and [XACZ61 Prevention of Radicalization Kyrgyz](#) were both notable for presenting the intended results and then systematically assessing the logframe achievements under the Effectiveness criterion.

Only six reports were assessed as having an adequate description of the context (Q d), which should include the most relevant social, economic, institutional factors that would help the reader understand the need for the intervention and the environment in which it is implemented. Only three reports had a sufficient description of the range of stakeholders involved (Q e). Usually, the UNODC staff structure and government partners (main duty bearers) were identified but there was less clarity about other actors such as civil society groups and on rightsholders, and even less information on the roles of different stakeholders. The inclusion of some version of a stakeholder map would enable a judgement to be made on the representativeness of the sampling process. At least one report noted that a stakeholder mapping process had been conducted during the Inception Phase, but the results were not included/annexed to the evaluation report. The evaluation scope (Q g) was another area where more information was often needed. The trend of only providing a sparse amount of background information may be due to the restricted page lengths for reports but there are good examples of concise

presentations such as [Counter Terrorism Maldives](#). Links to further information, particularly on context, is one way of enabling the reader to access more information when space is limited.

Table 8: Cumulative Ratings for Context, Purpose and Scope Questions

Evaluation Context, Purpose and Scope	Yes	Partial	No
a. Clear description of the project evaluated is presented. Project status is described including its phase of implementation, budget, and any significant changes (e.g., to strategies, logical frameworks) that have occurred.	12	2	0
b. Logic model and/or the expected results chain and /or programme theory (that at a minimum identifies and links objectives, outcomes and indicators of the project) is clearly presented. If included in the annex, it is also referred to in the main body of the report.	9	4	1
c. Connection with Sustainable Development Goals is clear.	10	3	1
d. Context of key cultural, gender-related, social, political, economic, demographic, environmental and institutional factors are described, as relevant for the object of the evaluation.	6	8	0
e. The main stakeholders - intended rightsholders (beneficiaries), duty bearers (state and non-state actors with responsibilities regarding the intervention), and financial supporters - are clearly described. Their interests, roles and linkages are identified.	3	11	0
f. Purpose and objectives of evaluation are clearly defined, including why it was needed at that point in time, its intended use, the key intended audience, and what it seeks to achieve.	11	3	0
g. Evaluation scope is clearly explained including time frame, geographical coverage, the main evaluation criteria, questions and justification of what the evaluation did and did not cover.	9	4	1

Good practice example of presenting the context, purpose and scope of the evaluation:

1819U Legal Aid Western Africa – This report clearly specifies the evaluation scope. It is also notable for including the programme logframe, theory of change and an evaluation matrix.

Counter Terrorism Maldives – This report is exemplary for its concise and well-sourced presentation of the project and context, as well as for providing a solid description of the evaluation purpose, scope and stakeholder involvement.

Design and Methodology: The scores are shown in table 9. As in 2020, while many reports were strong in multiple aspects of this section, none had a Very Good rating, and in 2021, four only achieved a rating of Fair. The shortcoming may in part also be attributable to page length constraints. Most evaluators presented this section in two to three pages, providing only minimal details on key elements of the evaluation process, and few taking the opportunity to present additional information in the annexes. However, there were examples of reports that mostly did well in providing a solid and precise explanation of the methodology used (Q a). [GLO565 INCB Precursor Control](#) is one of these. Its methodology section spanned 4 pages with visual aids taking up the equivalent of one page. Notable aspects were that the representativeness of the sample (in terms of stakeholder group, gender and geographic location) is clearly shown in charts, and it has a good description of data sources, including of documents reviewed. Its only shortcoming is that ethical considerations were not adequately covered.

There was somewhat of a drop in scores for methods being appropriate for analysing cross-cutting issues (Q b). This was due to half of the reports not addressing disability inclusion. The reports that had lower ratings for data collection methods and adequacy for measuring outcomes (Q c) were those that did not include a framework or other basis for gauging success. The most frequently seen issue was for high degree of participation of stakeholders (Q g); the shortcoming here being that there was often no mention of the involvement of stakeholders, including of Core Learning Partners (CLPs), in the design and review aspects of the evaluation. Although CLPs are engaged to varying degrees in all UNODC evaluations, that cannot always be seen from the evaluation reports – and in one case where involvement of stakeholders was not apparent, this may have contributed to a problem noted in limitations whereby some stakeholders were not accepting of the ToR.

Most reports failed to fully address ethical considerations (Q h). While there was usually mention of maintaining confidentiality and anonymity, the discussion generally did not go further into issues such as informed consent, ensuring comfort levels, data security and how evaluations adhered to the UNEG Principles for Evaluators. The [Counter Terrorism Maldives](#) evaluation demonstrated good practice in how it presented ethical considerations. [PERW77 Forest Crime](#) and [XACZ61 Prevention of Radicalization Kyrgyz](#) were also strong in this regard although both would have been stronger if evaluators had cited the UNEG Principles.

Table 9: Cumulative Ratings for Design & Methodology Questions – 2021

Design and Methodology	Yes	Partial	No
a. The design and set of methods are clearly described . . .	14	0	0
b. Methods are appropriate for analysing cross-cutting issues (human rights, gender equality, leave no-one behind, disability inclusion) . . .	7	7	0
c. Data collection methods and data sources are carefully described . . . Reference indicators and benchmarks are included where relevant.	8	6	0
d. Methods of analysis for all types of data are appropriate and clearly described . . .	11	3	0
e. Sampling frame is clearly described . . .	10	1	3
f. Innovative practice is used to improve the quality of evaluation process . . .	8	3	3
g. High degree of participation of internal and external stakeholders, including the Core Learning Partners . . .	2	12	0
h. Ethical considerations are addressed, including adherence to UNEG Ethical Principles for evaluators . . .	3	10	1

Good practice example of presenting ethical considerations:

Counter Terrorism Maldives - *The evaluation team was responsible for safeguarding and ensuring ethics at all stages of the evaluation. All team members were cognizant of the UNEG Code of Conduct for Evaluation in the UN System and Ethical Guidelines for Evaluation. Two key requirements are always to safeguard the independence of the team and to safeguard the rights and interests of its informants. This includes, but is not limited to, ensuring informed consent, protecting privacy, confidentiality and anonymity of participants, ensuring cultural sensitivity, respecting the autonomy of participants, ensuring fair recruitment of participants (including women) and ensuring that the evaluation results do no harm to participants. The evaluation team recognized UNODC’s strict policy of zero tolerance concerning unethical, unprofessional or fraudulent acts. (page 7 in the report)*

Recommendations: This section was rated as Good in all the 2021 reports. Table 10 shows there are two questions where shortcomings were more apparent. The primary one is regarding alignment with evaluation purpose and the subjects to be covered (Q b) – the main issue here was that there were no recommendations related to SDGs and/or to all the cross-cutting themes. [GLO565 INCB Precursor Control](#) and the [Maldives Counter Terrorism](#) evaluation are two reports where this was well done. There were several evaluations where the recommendations were silent on the cross-cutting themes even though shortcomings had been highlighted in Findings. There were also some problems with recommendations not being clustered and prioritized, and clearly identifying who is responsible and timeframe for action (Q d). Although the responsible party was usually noted, there was less consistency in reports addressing clustering and/or prioritizing, and timeframe. Another contributing factor to lower scores was that most evaluations did not reflect stakeholders’ consultations (Q e). It is important for the evaluators to describe how consultations, particularly on draft recommendations, took place. This requires more than stating that there is a CLP by also describing how the group was engaged.

Table 10: Cumulative Ratings for Recommendations Questions

8. Recommendations	Yes	Partial	No
a. Are clearly formulated, based on the conclusions, and substantiated by evidence collected.	14	0	0

b. Align with the evaluation purpose. They include recommendations related to the SDGs covered by the evaluation and emerging cross-cutting issues . . .	6	8	0
c. Are specific, realistic, indicate a timeframe, are actionable, and of a manageable number (maximum of 10).	12	2	0
d. Are prioritized and clustered (if appropriate), and clearly identify who is responsible for action	10	4	0
e. Reflect stakeholders' consultations whilst remaining balanced and impartial.	0	14	0

Good practice example of presenting recommendations:

XACZ61 Prevention of Radicalization Kyrgyz provided a clear structure for this section showing the conclusion(s) on which each recommendation was based, the priority level, the main recommendation highlighted in bold, and explanatory text. The responsible party was identified in the text but could instead have been highlighted as one of the check marked bullet points. UNODC may again want to consider a recommendation from the 2020 Synthesis Report that the evaluation report template provide a structure for the presentation of Recommendations to help ensure all required elements are included. [Note that only part of the recommendation is copied below.]

RECOMMENDATION 1 – PROJECT DESIGN

- ✓ Based on Conclusion 5
- ✓ Priority: high

For future programming, UNODC project management should elaborate a more detailed Theory of Change to build a strong intervention logic, in collaboration with partner agencies, experts and counterparts.

Taking into account existing guidance for example in the form of the UNODC Handbook on Results-based

Structure, Completeness and Clarity of Report: Although this section was generally well done, there were two criteria where scores tended to be lower. One was that annexes did not always include a logic model, theory of change or evaluation matrix (Q b). All should be included but this was only done in [1819U Legal Aid Western Africa](#). Several reports indicated that the evaluation matrix was included in the Inception Report, however good practice is for the matrix to also be attached to the evaluation report. The use of visual aids (Q g) scored lower in 2021 because many reports did not provide the respective source; this was a new requirement in the 2021 EQA template.

Table 11: Cumulative Ratings for Structure, Completeness and Clarity of Report

Structure, Completeness and Clarity of Report	Yes	Partial	No
a. Format (structure, headings, font) accords to IES Guidelines and Templates for Evaluation Reports. Structure accords to the following logical sequence: List of acronyms; Executive Summary . . .	14	0	0
b. Annexes include at a minimum: evaluation terms of reference; logic model and/or evaluation matrix; list of persons interviewed, and sites visited (if applicable); list of documents consulted; evaluation data collection tools used. They are formatted in the same style as the report.	9	5	0
c. Objectives stated in the terms of reference are adequately addressed.	14	0	0
d. Composition of Evaluation Team is described and includes subject matter, gender and geographic expertise. Preferably is gender balanced & includes professionals from countries/regions concerned.	11	3	0
e. Report is easy to read and understand . . . It is generally free from grammar, spelling . . . errors.	13	1	0
f. Language is empowering and inclusive avoiding gender, gender identity, age, cultural and religious bias, among others.	14	0	0
g. Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labeled (title and source), and referenced in text	3	11	0
h. Main body of the report (excluding executive summary, summary matrix, annexes) does not exceed 10% of the maximum mandatory length . . . (up to 30 pages . . .)	8	4	2

Mainstreaming Human Rights and Gender Considerations

All the evaluation reports included a section within Findings on Human Rights, Gender Equality and Leaving No One Behind (LNOB). This is the first year that LNOB was consistently included in UNODC evaluations as a cross-cutting theme. There was variation seen in the extent of analysis of each theme with human rights and gender equality generally getting more attention. Some reports also mainstreamed the cross-cutting themes into other criteria; [Counter Terrorism Maldives](#) is one example whereby the themes, primarily gender, were explored under multiple criteria. In another case, this section of the report included a thorough analysis of gender equality but made only passing mention of the other themes.

Human Rights

Several reports covered human rights at a general level with evaluators discussing the ways in which the topic was evident in project documents, progress reports, training curriculum, etc. and briefly explaining the connection between the intervention and the notion of human rights. [XACZ61 Prevention of Radicalization Kyrgyz](#) is one that provided a more in-depth analysis including on alignment of the intervention with international conventions and rules, use of relevant UNODC guidance, and the application of the principle of do no harm.

Gender Equality

All evaluation reports are rated specifically on the extent to which they meet the UN-SWAP criteria. The SWAP tool assesses how gender equality and the empowerment of women (GEEW) is integrated into evaluation processes. Each of the three 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 11 shows that the average cumulative scores for considering GEEW have increased significantly from 7.0 in 2020 to 7.8 in 2021. As with the previous four years, all UNODC evaluation reports included a specific sub-section on GEEW.

Table 11: Average scores for the integration of GEEW (UN-SWAP)³

Quality Assessment Criteria	Average Score (0-3)						
	2015	2016	2017	2018	2019	2020	2021
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	2.58	2.67
b. Gender-responsive evaluation methodology, methods and tools, and data analysis techniques are selected.	.875	1.37	1.87	2.25	2.09	2.17	2.53
c. Evaluation findings, conclusions and recommendations reflect a gender analysis.	1.875	1.84	2.2	2.31	2.73	2.25	2.53
Average Overall Score (out of 12)	5.5	6.53	8.2	N/A	N/A	N/A	N/A
Average Overall Score (out of 9)	N/A	N/A	N/A	6.94	7.09	7.00	7.73

³ In 2018 the number of UN-SWAP criteria were reduced from four to three.

Figure 5 shows that 87% of the 2021 portfolio Met Requirements for UN-SWAP (achieving a total score of 7 to 9) with 13% Approaching Requirements (achieving a total score of 4 to 6). None fell into the lowest band of Misses Requirements. Once again, most of the evaluations integrated GEEW into the evaluation scope (or objectives), as well as the questions and indicators (table 12). There has been a slight improvement from the 2020 reports when 64% of the 11 reports were rated as Fully Integrated compared to 67% in 2021. Methodology and findings/conclusions criteria also have markedly improved from 2020 with majority of the reports rated as Fully Integrated.

Figure 5: Overall Performance for UN-SWAP

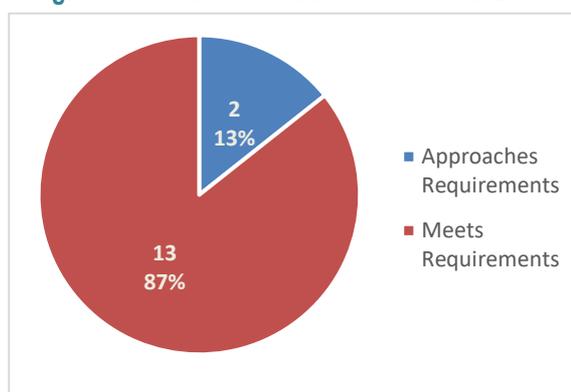


Table 12: Number of reports achieving scoring requirements for each criterion - 2021

	A. Scope/Indicator	B. Methodology	C. Findings/Conclusions/Recommend
Not at all integrated	0	0	0
Partially integrated	0	1	1
Satisfactorily integrated	5	5	5
Fully Integrated	10	9	9

Table 13 provides the overall EQA rating as well as the GEEW category and total scores for each report. It shows that there is not a consistent link between them. Of the four reports that had a total GEEW score of 9, only one had an overall EQA rating of Very Good, although the two reports that had the lowest GEEW score of 6 both had overall ratings of Good.

Table 13: GEEW scores by evaluation report - 2021

Project Number	Overall EQA Rating	Scores for each criterium			GEEW Total Score
		A	B	C	
1617AL	Very good	3	2	2	7
1819U	Very good	3	2	3	8
BOLZ68	Good	3	2	2	7
BOLW 35	Very good	3	3	2	8
GLOR35, GLOW63, MDVAB9	Good	3	3	3	9
GLO565	Very good	3	3	2	8
GLOU68	Very good	3	3	3	9
GLOX34	Good	3	3	3	9
GLOZ31	Good	2	3	1	6
GLOZ 99	Good	2	2	3	7
PERW77	Good	2	1	3	6
UNODC Programming in West and Central Asia	Good	3	3	2	8
XACZ61	Very good	2	3	3	8
GLO/R35	Good	3	3	3	9
SLKAB8	Good	2	2	3	7
Average score 2021		2.67	2.53	2.53	7.73
Average score 2020		2.6	2.2	2.3	7.00

Average score 2019	2.27	2.09	2.73	7.09
Average score 2018	2.44	2.25	2.31	6.94

Observations about the way that each criterion was addressed in the 2021 evaluations are as follows:

Design (Scope and Indicators): Average scores for this criterion increased from 2.6 in 2020 to 2.67 in 2021. All evaluation teams incorporated GEEW into the evaluation design by including gender-related questions and dedicating a section of the findings to HRGE issues. To achieve Fully Integrated, gender needed to be reflected in the objectives and/or scope of the evaluation (this was done in 10 out of 15 reports) and show that gender-related data was collected. Evidence of the latter includes the use of gender-related indicators in logical frameworks or evaluation matrices; however, these elements were not consistently included in the reports.

Methodology: Scores for this criterion showed the biggest increase from the previous year – from 2.17 to 2.53. To fully align with UN-SWAP guidance, it should be clear that the evaluation process was HRGE-responsive; for example, a mixed-methods approach appropriate to evaluating GEEW was used, a diverse range of data sources and processes were employed to enable triangulation, an appropriate diversity of stakeholders affected by the intervention were consulted, and ethical practices were followed to facilitate safe participation of all.

Most evaluations used a mixed methods approach, consulted an adequate range of stakeholders given the challenges of the pandemic, and provided a gender-disaggregated breakdown of evaluation participants. Most evaluators also highlighted the use of a gender responsive methodology, although not all of these went further by explaining how this was done. [GLOZ31 Wildlife in Asia](#) is one of the reports that did provide a good discussion of gender responsiveness and is also notable because the evaluators made a concerted effort to incorporate a gender sensitive approach in a male-dominated project. In addition to including questions related to gender equality in the data collection protocols and disaggregating evaluation participants by both stakeholder group and gender, they noted "*making a strenuous effort to interview as many females as possible*" and indicated that the high percent of male forest rangers presented a limitation to the evaluation process. [GLO/R35 Legal Regime Against Terrorism](#) is another good example as illustrated by the respective EQA comment below:

The evaluators highlight their use of a mixed methods approach that is gender-sensitive and inclusive, with attention given to gathering sex-disaggregated stats and gender-related information. Ethical practices used in collecting data are indicated and appropriate for an online process. The number of evaluation participants is gender-disaggregated by stakeholder group. A gender expert was part of the evaluation team and the gender-balance of the team is highlighted.

Findings, Conclusions and Recommendations: Improvement from last year was also notable here with the average for this criterion increasing from 2.25 to 2.53. To fully align with UN-SWAP guidance, there should be reference to GEEW dimensions of the context of the intervention, there should be a solid analysis of the extent to which gender was taken up by the intervention within the findings section (and women’s perspectives should be evident), and the issue raised should be carried through to the conclusions and recommendations.

All the evaluation reports included gender-related findings, with most providing an adequate level of analysis. Reports that were noted in the EQAs to be strong in this regard included [1819U Legal Aid Western Africa](#), [BOLZ68 Cultivos de Coca](#), [Counter Terrorism Maldives](#), [GLO565 INCB Precursor Control](#), [GLOZ31 Wildlife in Asia](#), [GLOZ99 UNCAC and SDG 16, Programming in West and Central Asia](#), and [GLO/R35 Legal Regime Against Terrorism](#). Where some of these and other reports fell short in this criterion, it was due to the gender-related issues raised in Findings not being carried through to both Conclusions and Recommendations.

Good Practice Example of Incorporating GEEW:

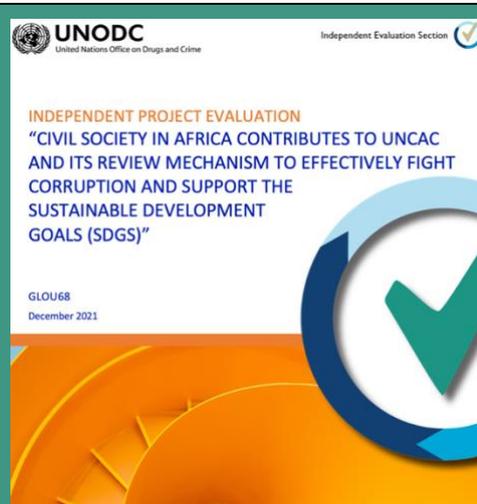
GLOU68 was exemplary for integrating all the cross-cutting issues, including disability, and was particularly thorough in its analysis of gender.

The evaluation covers gender in design as a cross cutting criteria and was one of the only reports to include gender issues in the discussion of context. The methodology section describes how the evaluation processes were gender sensitive. The findings included both qualitative and quantitative data on gender, and the voices of evaluation participants are reflected. Gender perspectives are also taken up in conclusions, recommendations, lessons learned and best practices. Below are excerpts of how GEEW is reflected in different sections.

Findings – *“The project literature positively reports that almost three quarters of CSOs who responded to the monitoring survey have since involved women in their activities or were women-led, and in fact 58% of the CSO beneficiaries who took part in the evaluation survey work on gender issues. Some attendees at the MSW explained during interviews that they were not aware of the gender-focused aspect of the project, so more explicit reference to this during the workshops and the importance of gender-inclusivity for the review process would aid in bringing this to the attendees’ attention”.*

Recommendations – *“The CSU should build on the national framework assessments undertaken at the start of this project, and incorporate a Political Economy Analysis, [including of] intersectionality issues affecting engagement of women, and a baseline assessment of skills and knowledge of attendees and likelihood to be involved in the review process (such as via a KAP survey)”*

Lessons – *“ . . . Responding to recommendations from the previous evaluation to put more focus on gender, human rights and leaving no-one behind, the CSU stepped up by inviting a wider range of CSOs including those working on youth, gender and HR issues. The selection of grantees for the small grants also reflected this drive for a broader inclusivity. This inclusivity brought more diverse voices to the table and enriched the discussions. . . .”*



Leave No One Behind

There was a reasonably solid analysis of LNOB in six of the 15 reports. [GLOU68 Civil Society Africa](#) is again a notable example, as is [XACZ61 Prevention of Radicalization Kyrgyz](#), which both looked more closely at the extent to which project activities reached more marginalized groups. Three reports briefly discussed LNOB, covering the theme in one dedicated paragraph. Another three reports only made brief mention, including the pilot IPE, within one to two sentences. As with human rights, a common treatment of LNOB was to look at the extent to which the theme appeared in documents.

Disability Inclusion

Only two of the reports (highlighted in the box below) could be classified as disability inclusive, the same number as in 2020. In these cases, disability was addressed in either the evaluation scope or questions, the findings included an analysis of disability inclusion, and the shortcomings found were taken up in the recommendations. Two other reports indicated that disability was to be addressed (either in the scope or questions) but it was not actually attended to anywhere in the analysis. In six reports, disability issues were not mentioned even though the ToRs explicitly requested that disability inclusion be covered. It should be noted that, unlike the topics of human rights and gender equality, there was no clear direction available on mainstreaming disability inclusion into evaluations until January 2022 when UNEG released [Guidance on Integrating Disability Inclusion in](#)

Good practice examples for disability inclusion:

PERW77 Forest Crime had a specific question under cross-cutting criteria that focused on people with disabilities. In addition, one of the recommendations was to *“Incorporate women’s, human rights, and indigenous peoples’ organizations, as well as organizations of people with disabilities, into the Platform for Forest Crimes in the months 1 to 6 of 2022, so that rights are incorporated into guides, documents, awareness and communication materials, as well as other elements”*.

GLOU68 Civil Society Africa covered disability in findings, best practices and recommendations. One of the recommendations included *“In pre-project planning, the CSU should deepen its assessment of the local country contexts, including intersectionality issues affecting women and those with disabilities . . .”*

Assessing Contributions to SDGs

UNODC activities are expected to contribute to the broader context of the Sustainable Development Goals. The EQA process assesses whether the connection to the relevant SDGs is made clear in the evaluation reports, and tracks which SDGs or SDG targets were identified. All 2021 reports referred to the SDGs; 11 evaluations identified the goals that were the most relevant and three reports just mentioned SDGs in general terms. Table 10 shows the SDG targets that were identified. All dealt with SDG 16, for which UNODC is the UN system lead. The second most frequently mentioned was SDG 5 which deals with gender equality. In terms of evaluation type, the IDEs addressed the biggest range of SDGs.

Table 14. SDG Targets Covered by UNODC Evaluations, by Type, 2021

SDG Targets	In-Depth Evaluations	IPEs	Total
3 Good Health and Wellbeing	1	1	2
4 Education	1		1
5 Gender Equality	1	4	5
5.2 Eliminate all forms of violence against all women and girls	1		1
11 Sustainable Cities and Communities	1		1
12 Responsible consumption and production	1		1
15 Life on land	1	1	2
16 Peace, Justice, and Strong Institutions	2	6	8
16.1 Significantly reduce all forms of violence and related death rates	1		1
16.3 Promote the rule of law at the national and international levels and ensure equal access to justice for all	1		1
16.4.1 Total value of inward and outward illicit financial flows		1	1
16.6 Develop effective, accountable, and transparent institutions . . .	1		1
16.A Strengthen relevant national institutions . . . to prevent violence and combat terrorism and crime	1		1
17 Partnerships for the Goals	1	1	2
Total	17	11	28

Evaluation Complexity, Including COVID-19

The EQA template includes a basic checklist to capture whether one or more of the following factors may have made the evaluation process particularly challenging to be undertaken. The categories are:

1. **The intervention itself is complex.** For example, it may 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 also adds to this complexity. Development partners, other duty bearers and rights-holders may have different agendas, requirements, expectations, levels of trust, or understandings of the intervention that must 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.
2. **The COVID-19 pandemic added complexity** usually because of travel restrictions for collecting in-person data and accessing the diversity of stakeholders and observing the context in which the intervention(s) operate.
3. **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.

All the 2021 evaluations were conducted during the pandemic and were therefore all complex in that regard. Five evaluations were seen to have at least one additional complexity factor, and the BOLZ68 Cultivos de Coca evaluation was noted as being complex due to all three factors. Partly this had to do with the need to collect data from beneficiaries which was difficult. It is probable that there were fewer cases of environmental complexity in both 2020 and 2021 due to travel restrictions.

Table 15 identifies the complexity factors and the overall EQA rating for each report. As was the case in 2020, this (limited) sample does not suggest a link between complexity and the quality of evaluations.

Table 15: Factors Noted as Increasing the Complexity of 2021 Evaluations

Project/Programme Number and Name	Intervention	COVID	Environment	EQA Rating
1617AL Developing Indicator on Illicit Financial Flows and Monitoring Them in Latin America		X		Very good
1819U Improving Access to Legal Aid for Women in Western Africa		X		Very good
BOLZ68 Evaluación Final Independiente Programa de Apoyo de la UNODC a la Implementación del Plan de Acción de la Estrategia de Lucha Contra el Narcotráfico y Reducción de Cultivos Excedentarios de Coca, 2011-2015 (ELCN-RCEC) del Estado Plurinacional de Bolivia (EPB)		X		Good
BOLW 35 Transparencia Institucional Y Participación Ciudadana Para La Gobernabilidad Municipal	X	X	X	Very good
GLOR35, GLOW63, MDVAB9 Support to Maldives on Counter Terrorism		X		Good
GLO565 INCB Databank for Precursor Control		X		Very good
GLOU68 Civil Society in Africa Contributes to UNCAC And Its Review Mechanism to Effectively Fight Corruption and Support the Sustainable Development Goals (SDGS)		X		Very good
GLOX34 - Global Firearms Programme: Countering Illicit Arms Trafficking and its Links to Transnational Organized Crime and Terrorism		X	X	Good
GLOZ31 Law Enforcement and Demand Management of Wildlife in Asia Project	X	X		Good

GLOZ 99 Global Programme to prevent and combat corruption through effective implementation of the UNCAC in support of Sustainable Development Goal 16		X	Good
PERW77 Criminal Justice Responses to Forest Crime in Peru		X	Good
UNODC Programming in West and Central Asia	X	X	Good
XACZ61 Support to the Prevention of Radicalization to Violence in Prisons and Probation Settings in the Kyrgyz Republic		X	Very good
GLO/R35 Independent In-Depth Evaluation Strengthening the Legal Regime Against Terrorism	X	X	Good

COVID-19: All reports indicated that the pandemic presented limitations to the evaluation process and provided one or more mitigation strategies. The level of explanation varied but all highlighted that the need to rely on remote data collection was the main challenge. Nine of the reports went into further details citing issues such as limitations for engaging with a broad range of stakeholders and in particular rights holders (including that some stakeholders were less available, institutions faced unexpected emergencies and were focused on responding to the emergency) and that it was more difficult for evaluators to understand the context for the intervention in the absence of field missions.

The most commonly mentioned mitigation strategy was to increase reliance on other forms of data collection such as conducting more in-depth document review, using online surveys, and reaching out to more duty bearers than would normally be the case. [GLO/R35 Legal Regime Against Terrorism](#) is notable for the number of methods used including event observations and case studies. Some evaluations, such as [Programming in West and Central Asia](#), relied on teams of local evaluators. [BOLZ68 Cultivos de Coca](#) used proxy measures to obtain rightsholders' perspectives by having field staff who worked most closely with communities write change stories about individuals who had benefited from the project (and noting the limitations to this approach).

[GLO565 INCB Precursor Control](#) is a good example of a report that provided a solid explanation of the limitations and how these were handled, stating that *"When interviews were interrupted by internet connections or stakeholders' unfamiliarity with remote communication platforms, the evaluation team was able to keep the conversation on the phone. In addition, the online survey was designed in a simple and clear way to make it easy for respondents to answer the questions and provide feedback."*

In addition to discussing COVID-19 in the Limitations section, [GLO/R35 Legal Regime Against Terrorism](#) provided a broader analysis of the pandemic. In the Background section, evaluators discussed the effect that the pandemic had on the threat posed by terrorism. One of the mini-case studies presented in the report looked at the intervention's use of on-line tools during the pandemic, and the findings from this followed through to one of the Conclusions.

Although the pandemic placed limitations on the extent to which evaluators could employ the most robust methodological approaches, the 2021 evaluations were all able to conduct adequate processes given the unprecedented constraints and uncertainty faced.

Pilot Independent Project Evaluations

One innovation during the year was to develop what were termed pilot independent project evaluations (IPE) that would be shorter in implementation time and length than standard evaluations. The argument for them was that many evaluations were too long and took too much time to prepare or read and that this reduced their value. The expected length of these reports, excluding the Executive Summaries and annexes, is 15 pages. While it was originally thought that there would be several of these finalized in 2021, in practice there was only one completed, the [Final Independent Project Evaluation of Promoting the Use of Non-custodial measures in Sri Lanka \(SLKAB8\)](#). Another was finalized in early 2022 and will be assessed in the next EQA cycle. SLKAB8 was reviewed under a briefer version of the standard EQA template. The pilot IPE EQA template has approximately 12 fewer questions and does not have the same expectations for issues such as innovative practice being

evident, a discussion on relevant SDGs, and as in-depth an analysis of contextual factors and cross-cutting issues as regular evaluations. A summary matrix for recommendations is also not a requirement of the Executive Summary.

The review of SLKAB8 suggests that it is feasible to produce an evaluation report that aligns with evaluation quality standards with the new approach. Its overall EQA total was Good with a weighted score of 88.75, just below the threshold for Very Good which is 90. The EQA summary states:

This is a pilot project to increase the use of non-custodial measures in Sri Lanka (with a more general applicability) through research and training. The evaluation shows that much of what the project intended was achieved, but that other outcomes would require more time to see the results (such as a reduction in the number of prisoners in prisons). Evaluators notably reconstructed the theory of change for the project and used this within the analysis. The methodology was based on documents and interviews and benefited from the use of an on-line survey of the selected interviewees before the in-depth interviews were done. It worked around the problems created by COVID-19. The findings are succinct but adequately address all criteria and questions, including (within the constraints of space) for the cross-cutting themes of human rights, gender equality, disability inclusion and leave no one behind. Given the page restrictions, the evaluators did well in presenting a succinct yet evidence-based report and fulfilling the requirements of the ToR (which was half the permitted length of the evaluation). The evaluation's recommendations appear useful for decision makers, and the set of lessons learned and best practices have a broader applicability.

SLKAB8 achieved a rating of Very Good in six of the nine sections of the review template: Structure, Completeness & Clarity; Context, Purpose & Scope; Design & Methodology; Reliability of Data; Lessons Learned & Good Practices; and GEEW. A noticeable difference from regular evaluations was the relatively small sample of respondents (15 stakeholders) which was considered acceptable for this more concise process. The remaining four sections were scored as Good. The shortcomings, mostly minor, were of issues commonly seen in regular evaluations – intended audience not identified in the Executive Summary, more clarity needed on the role of community-level stakeholders in project implementation, analysis process not specified for each type of data collected, data sources not sufficiently differentiated by stakeholder group, disability inclusion only briefly taken up, recommendations not prioritized, and minimal use of visual aids.

The evaluation was 18 pages in length and included most of the material expected of an evaluation. It addressed the purpose and all criteria expressed in the ToR. Partly this was because the annexes, totaling 24 pages, included much of the required supporting text. The main reason that such a concise report received a Good overall rating was that it was a very specific project with very clear expected results and causal connections with output that could be measured.

The length of evaluations is largely determined by the complexity of the project and that is generally reflected in the number of questions and sub-questions that need to be answered. The pilot IPE [EQA] template is limited to six evaluation questions, which is approximately 12 fewer questions as compared to the previous IPE template. In the case of SLKAB8, there were six criteria, eight main questions and a total of fifteen sub-questions, shown in an annex. Fewer criteria and questions would enable a more in-depth analysis of each area of enquiry. One conclusion is that the permitted length of an IPE should be determined by the complexity of the project being evaluated. This is related to whether it is in a single country, the number and scope of implementers of the project and the complexity of obtaining data given the context of the evaluation. The Sri Lanka project was one-country, a very specific program with clear but limited numbers of implementation stakeholders, where there were good accessible statistics and where there were few constraints like conflicts, weather, elections, or other factors that affected data collection. This provides a useful model for determining the length of future IPEs. There is a second pilot that was completed in February 2022 that was a regional project in Latin America that met the page limit. It will be assessed next year and may provide more information on how well pilots align with the EQA requirements.

CONCLUSIONS and RECOMMENDATIONS

The overall quality of UNODC evaluations continues to be high with 40% of the 2021 portfolio (six of the 15 reports) being rated as Very Good and the remainder as Good (including the pilot IPE). The overall scores were higher than previous years across five of the ten sections of the template, with the most significant improvement being in the Integration of GEEW. The performance in Reliability of Data, Conclusions, and Lessons Learned were the strongest sections. The evaluation reports were all clearly presented. The use of the new report template has elevated the overall presentation and ensured a clear structure that can be easily navigated. The lowest scoring sections were Executive Summary, Design & Methodology, Context/Purpose/Scope, and to a lesser extent in Recommendations.

Evaluators have mostly managed to conduct thorough assessments despite the significant limitations posed by the pandemic and the resultant need to move to online methods of data collection. All evaluators discussed the strategies they had implemented to mitigate the constraints faced.

Although the portfolio aligned well with UNEG and UNODC evaluation standards, there continues to be some room for improvement. In addition to some persistent issues from previous years, it was observed that there are varying levels of reference made to ethical considerations, which may be related to the predominant use of remote data collection practices. This may be an area for specific guidance as issues around comfort level, confidentiality, and data security still need to be thoughtfully addressed and discussed. Best practice within this sub-criterion includes having a dedicated section under methodology that provides an explanation of (i) how evaluators upheld each of the UNEG Ethical Principles (including signing the Ethical Pledge by all team members), and (ii) ethical considerations for collecting and storing data (such as ensuring that interview and focus group locations and times were appropriate and easily accessible to all participants; and steps to safeguard data).

Another area where most evaluations fell short was in meeting the standards set out in the UN System-wide Disability Action Plan for disability inclusion. As it may not be obvious to see how disability-related issues can be addressed in some types of UNODC programming, this may also be an area where specific guidance is needed for evaluators and evaluation managers. [UNFPA guidance](#) may be a useful resource to draw from as it establishes a set of practices that should be applied during six phases of the evaluation process.

UNODC evaluation reports have not typically included a stakeholder analysis. In addition to being a useful component of the evaluation process, stakeholder mapping (including of more secondary stakeholders which are often civil society groups) is helpful background information for the reader/audience; as well, without clarity on the range of actors involved in the intervention, the representativeness of the evaluation sample cannot be assessed. To help support this, a template for a basic stakeholder map/table may be a useful addition to the annexes in the evaluation report template. Stakeholder involvement in the evaluation process is also important to have and to explain. If it is not necessary or feasible to have this, then the rationale should be specified in the report. Similarly, a specific format for the presentation of Recommendations may be useful to assist evaluators in aligning with the quality expectations for this section.

Recommendation 1 UNODC should continue to push the quality of evaluations. Areas for improvement include paying further attention to:

- a) More substantive discussion of the intervention, its context, and stakeholders involved, as well as of the scope of the evaluation. Ideally the evaluation process should include a stakeholder analysis with the results being provided in, or attached to, the evaluation report.
- b) Inclusion and assessment of the theory of change and/or results framework in the evaluation report; proposing revisions based on the assessment; and using the ToC and/or results framework in the analysis of the results of the intervention.
- c) Inclusion of the evaluation matrices in the evaluation report; and ensuring that these clearly define how success is to be measured by the evaluation.
- d) Engagement of key stakeholders in the validation of Findings and the development/validation of Recommendations.
- e) Discussing ethical considerations and the application of UNEG Evaluation Principles for Evaluators.
- f) Addressing disability inclusion and extracting direction for this from the recently released UNEG Guidance on Integrating Disability Inclusion.
- g) Providing a structure for the Recommendations section to help ensure evaluators attend to all the quality requirements for this section. This could be done in the form of a template that has a specified space for timeframe, priority, to whom directed, link to relevant conclusions, operational suggestions, etc.

UNODC has been diligent in its efforts to increase the accessibility and usefulness of evaluations by restricting the number of pages for these reports. The new 30-page limit for regular evaluations is significantly shorter than what is set by other UN entities with which the reviewers are familiar. Although several IPE reports met this requirement, the IDE reports were all considerably longer. The IES may want to consider a slight increase in the page limits for the IDEs given that EQAs noted more substantial discussions were needed in the introductory and methodology sections of many reports. The limit for the new IPE reports is significantly more restrictive with a 15-page limit. The one pilot evaluation came close to achieving this at 18 pages and was able to meet most of the EQA requirements. It will be useful to review a few more examples of the pilot evaluations before making a more definitive assessment of this template.

At the same time, it must be recognized that if an in-depth exploration of evaluation questions is desired, the number of questions should be commensurate with the desired length of the report, otherwise they risk being addressed in a superficial manner. The ToR for one of the 2021 evaluations included 27 questions, providing an illustrative example of the challenge facing evaluators in delivering a 30-page report. This is another area where guidance may be useful to help evaluation managers to prioritize and consolidate areas of enquiry, and making clear which questions are expected to have a more substantive and triangulated analysis.

Recommendation 2 IES should consider increasing the maximum page lengths to help ensure that all UNEG and UNODC evaluation standards and the requirements of the EQA template can be met. It is suggested that:

- a) IDEs have a maximum length of approximately 50 pages.
- b) Pilot IPEs have a maximum length of approximately 20 pages.
- c) The number of questions that can be included in the ToRs for evaluations be restricted; and in cases where more questions are needed or the complexity of the intervention requires more extensive analysis, the ToR should set a higher page limit for that report.

Although the UNODC EQA template appears to function well in providing a solid assessment of the quality of evaluations, there are a few minor adjustments that should be considered to further encourage improvements in areas where shortcomings are more persistent.

Recommendation 3 IES should consider having the following issues reflected in the template, some of which could be incorporated into existing questions:

- a) The evaluation matrix to include indicators, benchmarks, assumptions and/or other processes from which the analysis can be based, and conclusions drawn.
- b) Explicit use of the intervention's results framework/ToC in the formulation of the findings.
- c) Findings to be based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.
- d) Questions, or a set of questions, to enable a more structured assessment of disability inclusion.

Annex 1

List of evaluations reviewed

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

Project Number	Project Title
1617AL	Developing Indicator on Illicit Financial Flows and Monitoring Them in Latin America
1819U	Improving Access to Legal Aid for Women in Western Africa
BOLZ68	Evaluación Final Independiente Programa de Apoyo de la UNODC a la Implementación del Plan de Acción de la Estrategia de Lucha Contra el Narcotráfico y Reducción de Cultivos Excedentarios de Coca, 2011-2015 (ELCN-RCEC) del Estado Plurinacional de Bolivia (EPB)
BOLW 35	Transparencia Institucional y Participación Ciudadana para la Gobernabilidad Municipal
	GLOR35, GLOW63, MDVAB9 Support to Maldives on Counter Terrorism
GLO565	INCB Databank for Precursor Control
GLOU68	Civil Society in Africa Contributes to UNCAC and its Review Mechanism to Effectively Fight Corruption and Support the Sustainable Development Goals (SDGS)
GLOX34	- Global Firearms Programme: Countering Illicit Arms Trafficking and its Links to Transnational Organized Crime and Terrorism
GLOZ31	Law Enforcement and Demand Management of Wildlife in Asia Project
GLOZ 99	Global Programme to prevent and combat corruption through effective implementation of the United Nations Convention against Corruption in support of Sustainable Development Goal 16
PERW77	Criminal Justice Responses to Forest Crime in Peru
	UNODC Programming in West and Central Asia
XACZ61	Support to the Prevention of Radicalization to Violence in Prisons and Probation Settings in the Kyrgyz Republic
GLO/R35	Independent In-Depth Evaluation Strengthening the Legal Regime Against Terrorism
Pilot: SLKAB8	Final Independent Project Evaluation Promoting the Use of Non-Custodial Measures in Sri Lanka

Annex 2

UNODC EQA Template as Used in the Review

UNODC Evaluation Quality Assessment (EQA)

version: December 2021

In-Depth/Independent Project/Strategic Evaluations

General Project Information	
Project/Programme Number and Name	
Thematic Areas	
Countering transnational organized crime	
Countering world drug problem	
Countering corruption	
Terrorism prevention	
Justice	
Research, trend analysis and forensics	
Policy support	
Technical cooperation and field support	
Support to international bodies	
Geographic Area (Region, Country)	
Relevant SDG(s)	
Approved Project/Programme Budget at the Time of the Evaluation (USD)	
Type of Evaluation	
In-Depth/Independent Strategic	
Final/Midterm	
Other characteristics (i.e., Joint, Cluster)	
Cost of Evaluation (USD)	
Evaluation Team (# of men/# of women; names of team members)	
Evaluation Complexity Factors (Intervention, Stakeholders, Environment, COVID-19, Other, Not applicable)	
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:	#DIV/0!
Summary:	

Quality Assessment Criteria	Meets Criteria?	RATING:	Assessment Levels: Very Good - Good - Fair - Unsatisfactory
1. Structure, Completeness And Clarity Of Report			#DIV/0!
a. Format (structure, headings, font) accords to IES Guidelines and Templates for Evaluation Reports. Structure accords to the following logical sequence: List of acronyms; Executive Summary (also in English for Spanish reports); Summary Matrix of Findings, Evidence and Recommendations; Introduction (Background and Context, Evaluation Scope and Methodology, Limitations to the Evaluation); Findings presented by relevant criteria (Relevance, Efficiency, Partnership and Coherence, Effectiveness, Impact, Sustainability, Human Rights/Gender Equality/Leaving No One Behind, as well as Design and Innovation if in ToR); Conclusions; Recommendations; Lessons Learned & Best Practice.			
b. Annexes include at a minimum: evaluation terms of reference; logic model and/or evaluation matrix; list of persons interviewed and sites visited (if applicable); list of documents consulted; evaluation data collection tools used. They are formatted in the same style as the report.			
c. Objectives stated in the terms of reference are adequately addressed.			
d. Composition of Evaluation Team is described and includes subject matter, gender and geographic expertise. Preferably it is gender balanced and includes professionals from countries or regions concerned.			

e. Report is easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience). It is generally free from grammar, spelling, or punctuation errors.		
f. Language is empowering and inclusive avoiding gender, gender identity, age, cultural and religious bias, among others.		
g. Visual aids, such as maps and graphs, are used to convey key information. Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labeled (title and source), and referenced in text		
h. Main body of the report (excluding executive summary, summary matrix, annexes) does not exceed 10% of the maximum mandatory length established in the evaluation report template (up to 30 pages, or 100,000 characters, including spaces)		
2. Executive Summary		RATING: #DIV/0!
a. Clearly presented and serves as a stand-alone section that provides an overview of the evaluation and 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 Findings/Conclusions; v) Recommendations; vi) Main Lessons/Best Practices.		
c. Summary Matrix presents only the key and most important recommendations from evaluation report. Findings, sources and recommendations are clear and cohesive, and specify the stakeholder to whom they are addressed.		
d. Maximum length of 4 pages, excluding the Summary Matrix.		
3. Evaluation Context, Purpose and Scope		RATING: #DIV/0!
a. Clear description of the project evaluated is presented. Project status is described including its phase of implementation, budget, and any significant changes (e.g. to strategies, logical frameworks) that have occurred.		
b. Logic model and/or the expected results chain and /or programme theory (that at a minimum identifies and links objectives, outcomes and indicators of the project) is clearly presented. If included in the annex, it is also referred to in the main body of the report.		
c. Connection with Sustainable Development Goals is clear.		
d. Context of key cultural, gender-related, social, political, economic, demographic, environmental and institutional factors are described, as relevant for the object of the evaluation.		
e. The main stakeholders - intended rightsholders (beneficiaries), duty bearers (state and non-state actors with responsibilities regarding the intervention), and financial supporters - are clearly described. Their interests, roles and linkages are identified.		
f. Purpose and objectives of evaluation are clearly defined, including why it was needed at that point in time, its intended use, the key intended audience, and what it seeks to achieve.		
g. Evaluation scope is clearly explained including time frame, geographical coverage, the main evaluation criteria, questions and justification of what the evaluation did and did not cover.		
4. Design And Methodology		RATING: #DIV/0!
a. The design and set of methods are clearly described. They appear appropriate and adequately robust (including by having an adequate sample size) for addressing the evaluation purpose, objectives, questions and criteria, as well as for drawing causal connections between outputs and expected outcomes.		
b. Methods are appropriate for analysing cross-cutting issues (human rights, gender equality, leave no-one behind, disability inclusion) as required and relevant to evaluation object; methodology takes into account power relations during an evaluation process; the process is inclusive and participatory.		
c. Data collection methods and data sources are carefully described, as is the rationale for selecting them. Processes are adequate for measuring outcomes set out in logic model. Reference indicators and benchmarks are included where relevant.		

d. Methods of analysis for all types of data are appropriate and clearly described. In addition, the type of analysis (such as descriptive &/or analytical statistics) and process for analysing data (such as excel database for coding, or any specific software used) is identified.		
e. Sampling frame is 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.		
f. Innovative practice is used to improve the quality of evaluation process. This could be evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong human rights focus), approaches suitable for unsafe environments, or ways of sharing of evaluation results.		
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.		
h. Ethical considerations are addressed, including adherence to UNEG Ethical Principles for evaluators. Ethical safeguards and data security relevant to each data source/type of stakeholder are described.		
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.		
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. Methodological challenges and 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. This should include any constraints due to the pandemic.		
d. Categories of relevant stakeholders as for data collection are adequately described and justified in the methodology section of the report; and disaggregation of such categories is taken into account, as appropriate, in evaluation findings, conclusions, and recommendations.		
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. They are logical and coherent.		
b. Are objective and based on rigorous analysis of the data collected; take into account any identified benchmarks.		
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, coherence/partnership and cooperation, innovation, and the cross-cutting themes of human rights and gender.		
d. Include findings related to the SDGs covered by the evaluation.		
e. Discuss alignment and any variances between planned and actual results of the project (in terms of objectives, outcomes, outputs).		
<i>Analysis</i>		
f. Contextual factors are identified (including reasons for accomplishments and failures, and continuing constraints), as are opportunities to build upon.		
g. Cause and effect links between an intervention and its end results (including unintended results) are explained.		
h. Includes substantive analysis of human rights and other cross-cutting issues as relevant to the evaluation.		
7. CONCLUSIONS		RATING: #DIV/0!
a. Take into consideration all questions and evaluation criteria, including human rights and gender equality/mainstreaming.		

b. Have been formulated clearly, are based on findings and substantiated by evidence collected, and go beyond the findings to provide a thorough understanding of the underlying issues of the project and add value to the findings.			
c. Convey the evaluators' unbiased judgement of the intervention.			
d. Present a comprehensive picture of both the strengths and weaknesses of the project.			
8. RECOMMENDATIONS		RATING:	#DIV/0!
a. Are clearly formulated, based on the conclusions, and substantiated by evidence collected.			
b. Align with the evaluation purpose. They include recommendations related to the SDGs covered by the evaluation and emerging cross-cutting issues, and also address flaws found, if any, in project's data acquisition processes.			
c. Are specific, realistic, indicate a timeframe, are actionable, and of a manageable number (maximum of 10).			
d. Are prioritized and clustered (if appropriate), and clearly identify who is responsible for action			
e. Reflect stakeholders' consultations whilst remaining balanced and impartial.			
9. LESSONS LEARNED & GOOD PRACTICES		RATING:	#DIV/0!
a. Are clearly 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, and have the potential to improve future actions..			
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.			
b. Gender-responsive evaluation methodology, methods and tools, and data analysis techniques are selected.			
c. Evaluation findings, conclusions and recommendations reflect a gender analysis.			
Overall Score	#DIV/0!		
Overall Rating	#DIV/0!		
Cross-cutting Themes Adequately Addressed			
Leave No One Behind			
Gender Equality			
Disability Inclusion			

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	#DIV/0!	#DIV/0!
Executive Summary	7	#DIV/0!	#DIV/0!
Evaluation Context, Purpose and Scope	5	#DIV/0!	#DIV/0!
Evaluation and Methodology	15	#DIV/0!	#DIV/0!
Reliability Of Data	5	#DIV/0!	#DIV/0!
Findings And Analysis	25	#DIV/0!	#DIV/0!
Conclusions	10	#DIV/0!	#DIV/0!
Recommendations	15	#DIV/0!	#DIV/0!
Lessons Learned and Good Practice	3	#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			

Annex 3

UNODC EQA Template for Pilot IPEs

UNODC Evaluation Quality Assessment (EQA)

version: December 2021

Independent Project Evaluations (IPEs) - Pilot

General Project Information	
Project/Programme Number and Name	
Thematic Areas	
Countering transnational organized crime	Yes
Countering world drug problem	
Countering corruption	
Terrorism prevention	
Justice	
Research, trend analysis and forensics	
Policy support	
Technical cooperation and field support	
Support to international bodies	
Geographic Area (Region, Country)	
Relevant SDG(s)	
Approved Project/Programme Budget at time of the Evaluation (USD)	
Type of Evaluation (Final/Midterm; Cluster; UNDA funded; Other characteristics)	
Cost of Evaluation (USD)	
Evaluation Team (# of men/# of women; names of team members)	
Evaluation Complexity Factors (Intervention, Stakeholders, Environment, COVID-19, Other, Not applicable)	COMPLEX FACTORS HERE (if selecting multiple options use comma with no space (i.e., Intervention,COVID-19)
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	Meets Criteria?	Assessment Levels: Very Good - Good - Fair - Unsatisfactory	
1. Structure, Completeness And Clarity Of Report		RATING:	#DIV/0!
a. Format (structure, headings, font) accords to IES Guidelines and Templates for IPE Reports. Structure accords to the following logical sequence: List of acronyms; Executive Summary (also in English for Spanish reports); Introduction (Background and Context, Evaluation Scope and Methodology, Limitations to the Evaluation); Findings presented by relevant criteria specified in ToR); Conclusions; Recommendations; Lessons Learned.	<select one>		
b. Annexes include at a minimum: evaluation terms of reference; evaluation tools; desk review list; stakeholders consulted. They are formatted in the same style as the report. UNDA-funded projects should also include an evaluation matrix and results framework.	<select one>		
c. Objectives stated in the terms of reference are adequately addressed.	<select one>		
d. Composition of Evaluation Team is described and includes subject matter, gender and geographic expertise. Preferably it is gender balanced and includes professionals from countries or regions concerned.	<select one>		
e. Report is easy to read and understand (i.e. written in an accessible non-technical language appropriate for the intended audience). Report is generally free from grammar, spelling, or punctuation errors.	<select one>		
f. Language is empowering and inclusive avoiding gender, gender identity, age, cultural and religious bias, among others.	<select one>		
g. Visual aids, such as maps and graphs, are used to convey key information. Adequate use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labeled (title and source), and referenced in text	<select one>		
h. Report is of reasonable length (maximum of approximately 15 pages excluding Executive Summary and annexes).	<select one>		
2. Executive Summary		RATING:	#DIV/0!
a. Clearly presented and serves as a stand-alone section that provides an overview of the evaluation and its main results.	<select one>		

b. Generally follows the structure of: i) Brief description of intervention including objectives; ii) Purpose, scope, intended audience(s) of evaluation; iii) Methodology; iv) Composition of evaluation team; v) Main Findings/Conclusions; vi) Key Recommendations; vii) Main Lessons/Good Practice	<select one>	
c. Maximum length of approximately 3 pages, excluding the Summary Matrix.	<select one>	
3. Evaluation Context, Purpose and Scope		RATING: #DIV/0!
a. Clear description of the project evaluated is presented, including project status, duration, and financial and human resources.	<select one>	
b. Expected chain of results is provided (a brief narrative, logic model or theory of change that identifies and links objectives, outcomes and outputs of the project or parts of the project under review).	<select one>	
c. Connection with Sustainable Development Goals is clear.	<select one>	
d. Context of key cultural, gender-related, social, political, economic, demographic, environmental and institutional factors are described, as relevant for the object of the evaluation.	<select one>	
e. The main stakeholders - intended rightsholders (beneficiaries), duty bearers (state and non-state actors with responsibilities regarding the intervention), and financial supporters - are identified.	<select one>	
f. Purpose (objectives) and scope of evaluation are clearly defined, as is the key intended audience.	<select one>	
4. Design And Methodology		RATING: #DIV/0!
a. The design/approach is clearly described. This appears appropriate and adequately robust for addressing the evaluation purpose, questions and criteria, as well as for drawing causal connections between outputs and expected outcomes.	<select one>	
b. Explanation provided for how methods are appropriate for analysing gender and human rights issues, and for how methodology takes into account power relations during the evaluation process.	<select one>	
c. Data collection methods and data sources are described, as is the rationale for selecting them.	<select one>	
d. Methods of analysis for all types of data are described.	<select one>	
e. Sampling process is described.	<select one>	
f. Ethical considerations are addressed, including adherence to UNEG Ethical Principles for evaluators. Ethical safeguards and data security relevant to each data source/type of stakeholder are described.	<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. Methodological challenges and 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. This should include any constraints due to the pandemic.	<select one>	
6. FINDINGS AND ANALYSIS <i>To ensure sound analysis and credible findings</i>		RATING: #DIV/0!
a. Findings are clearly formulated and presented. They are logical and coherent.	<select one>	
b. Findings are objective and evidence-based.	<select one>	
c. Findings address all relevant evaluation criteria and questions raised in the ToR as well as the cross-cutting themes of human rights, gender equality, disability inclusion and leaving no one behind.	<select one>	
d. Cause and effect links between an intervention and its end results (including unintended results) are explained, including reasons for accomplishments and failures as well as any variance between planned and actual results of the project	<select one>	
7. CONCLUSIONS		RATING: #DIV/0!
a. Take into consideration all questions and evaluation criteria, including human rights and gender equality, and disability inclusion.	<select one>	
b. Have been formulated clearly, are based on findings and substantiated by evidence collected, and go beyond the findings to provide a thorough understanding of the underlying issues of the project and add value to the findings.	<select one>	
c. Present a comprehensive picture of both the strengths and weaknesses of the project.	<select one>	
8. RECOMMENDATIONS		RATING: #DIV/0!

a. Are clearly formulated (with the topic of each indicated with a sub-heading), based on the conclusions, and substantiated by evidence collected.	<select one>	
b. Are specific, realistic, indicate a timeframe, are actionable, and of a manageable number (preferably 4-6).	<select one>	
c. Are clustered and prioritized, and clearly identify who is responsible for action.	<select one>	
9. LESSONS LEARNED AND GOOD PRACTICES		RATING: #DIV/0!
a. Lessons and Good Practices are both presented. They 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, and have the potential to improve future actions.	<select one>	
Overall Score	#DIV/0!	
Overall Rating	#DIV/0!	
Cross Cutting Issue Adequately Addressed		
Leave No One Behind		
Gender Equality		
Disability Inclusion		

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	#DIV/0!	#DIV/0!
Executive Summary	10	#DIV/0!	#DIV/0!
Evaluation Context, Purpose and Scope	5	#DIV/0!	#DIV/0!
Evaluation and Methodology	15	#DIV/0!	#DIV/0!
Reliability Of Data	5	#DIV/0!	#DIV/0!
Findings And Analysis	25	#DIV/0!	#DIV/0!
Conclusions	10	#DIV/0!	#DIV/0!
Recommendations	15	#DIV/0!	#DIV/0!
Lessons Learned and Good Practices	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			

This report was prepared by Ann Sutherland and John Mathiason. Both have extensive experience in conducting evaluations and meta-evaluations for international organizations. They are Director and Board Chair, respectively, for DeftEdge Corporation.

March 2022



INDEPENDENT EVALUATION SECTION

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/evaluation