

**2019 META-EVALUATION REPORT ON INDEPENDENT EVALUATION
ASSESSMENTS**

**UN WOMEN GLOBAL EVALUATION REPORTS ASSESSMENT AND ANALYSIS
SYSTEM (GERAAS)**

Prepared by

CLAUDIA B. MARCONDES

March 2020

Table of Contents

1. INTRODUCTION 1

2. BACKGROUND 1

3. REVIEW PROCESS AND METHODOLOGY 2

4. THE GERAAS INDEPENDENT ASSESSMENT RESULTS 3

 Overall Rating 4

 Object of Evaluation..... 7

 Evaluation Purpose, Objectives and Scope 8

 Evaluation Methodology 9

 Evaluation Findings 10

 Evaluation Conclusions 10

 Evaluation Recommendations, 11

 Gender, Human Rights and the SDGs 12

 Report Structure 15

 Good practices and trends..... 15

5. ASSESSMENT STRENGTHS AND WEAKNESSES..... 20

6. RECOMMENDATIONS..... 21

ANNEX 1: List of evaluations covered in the 2019/2020 GERAAS and UN-SWAP Assessments 24

1. INTRODUCTION

This document has been prepared at the request of the United Nations Entity for Gender Equality and the Empowerment of Women (UN-Women) to report on the results of the independent assessments carried out in 2019, regarding the Global Evaluation Reports Assessment and Analysis System (GERAAS). The assessment was carried out between December 2019 and January 2020 and covered 38 evaluations undertaken by UN Women in 2019. A list of all evaluations is included in the Annex 1. Each individual evaluation was assessed using the templates provided by UN-Women Independent Evaluation Service (IES) and rated against eight criteria set by the United Nations Evaluation Group (UNEG).

This report summarizes results of this independent assessment against the eight individual parameters/ standards and identifies good practices used by the evaluation reports. It also provides a global analysis of trends showing key weaknesses and strengths identified in the process, and presents some recommendations to improve the GERAAS system as well as the quality of the evaluation reports.

2. BACKGROUND

The United Nations Entity for Gender Equality and the Empowerment of Women (UN-Women) is dedicated to the achievement of gender equality and the empowerment of women. The mandate and functions of UN-Women call for the promotion of organizational and UN system accountability and learning on gender equality, among others, through evaluations.

One of the requirements of the UN Women evaluation policy is that evaluations should be conducted in a systematic manner, applying sound approaches and methods to ensure the quality and credibility of findings, recommendations and lessons generated to inform programme management and contribute to gender equality results. In response to this, the Independent Evaluation Service (IES), since 2013, has rolled-out the Global Evaluation Reports Assessment and Analysis System (GERAAS) to enhance the quality and credibility of both corporate and decentralized evaluations. The system was inspired by similar good practices enforced by other UN entities and consistent with the UNEG Norms and Standards.

The GERAAS is an approach to rating evaluation reports using UN Women, UN Evaluation Group (UNEG) and UN System Wide Action Plan (UN-SWAP) Evaluation Performance Indicator. The system serves as a key instrument to increase the application of sound approaches and methods to continuously improve the quality and credibility of evaluation methods and reports within UN-Women. The UN-SWAP Evaluation Performance Indicator serves both as a reporting tool and a benchmark to help UN entities integrate Gender Equality and the Empowerment of Women (GEEW) into evaluations.

As part of this process, the IES through an external independent consultant assesses the quality of corporate and decentralized evaluation reports and makes available assessment findings to senior managers, programme units, and the UN Women Executive Board.

3. REVIEW PROCESS AND METHODOLOGY

The GERASS assessments used the UNEG Quality Checklist for Evaluation Reports, which establish indicators for high-quality evaluation reports as a basis for the reviews and assessments. The scorecards provided by UN Women, included the following 8 criteria: (1) Object of Evaluation, (2) Evaluation Purpose, Objectives and Scope, (3) Evaluation Methodology, (4) Evaluation Findings, (5) Evaluation Conclusions, (6) Evaluation Recommendations, (7) Gender, Human Rights and the SDGs; and, (8) Report Structure.

The Gender, Human Rights and the SDGs assessments followed the methodology and tools established by the UNEG, including [the UN SWAP Evaluation Scorecard](#), and the [UN SWAP assessment tool](#) and assessed the reports against the three criteria established to capture the integration of gender equality perspectives in the final evaluation reports, including:

- 1) GEWE is integrated in the evaluation scope of analysis and evaluation criteria and questions are designed in a way that ensures GEWE related data will be collected;
- 2) A gender-responsive methodology, methods and tools, and data analysis techniques are selected; and,
- 3) The evaluation findings, conclusions and recommendations reflect a gender analysis.

Each evaluation report was assessed regarding GEEW integration, using a four-point scale (0-3) rating system for each criterion¹. Each evaluation report reviewed was qualified as “*Missing requirement*” (none of the UNEG gender-related norms and standards are met by the evaluation); “*Approaching requirement*” (the evaluation meets some of the UNEG GE-related norms and standards in the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation); or, “*Meeting requirement*” (the evaluation meets UNEG GE-related norms and standards and applies

¹ Each of the scoring levels below corresponds to a numbered score:

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

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

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

For an evaluation to “meet requirements” at least one of the criteria needs to be assessed at “fully integrated (3)”.

the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation during all phases of the evaluation) depending on the overall score obtained.²

As for the rating system used for the assessments of other report sections, this included the following four categories “Very Good”, “Good”, “Fair” and “Unsatisfactory”. The thresholds for each category for both Section ratings and Overall report ratings are presented in below:

Parameter Rating Criteria	Fully	Mostly	Partly	Not at all
UN SWAP (Gender, Human Rights & SDGs)	Fully integrated (3)	Satisfactorily integrated (2)	Partially integrated (1)	Not at all integrated (0)
Section Ratings	Very Good	Good	Fair	Unsatisfactory
Section Rating (Number - Above)	0.7499	0.4999	0.2499	
Overall Rating (Number - Above)	84.99	64.99	49.99	

Each report was rated against the specific parameters for each criterion and individual scores were presented in “Section” ratings (in line with each parameter) in an excel form. The form contained automated calculations based on specific weights pre-determined by UN Women, and automatically provided the ratings per criteria and a total “Overall” rating for the overall quality of each report. The rating explanations to the overall rating for the reports are the following:

Rating Scale	Very Good	Good	Fair	Unsatisfactory
Rating explanation	The report can be used with high level of confidence and is considered a good example.	The report can be used with certain degree of confidence.	Partially meets requirements with some missing elements. The report can be used with caution.	Misses out the minimum quality standards.

4. THE GERAAS INDEPENDENT ASSESSMENT RESULTS

The universe of 38 evaluations included in this assessment covered 23 countries in all regions. The Table 1 shows the geographical distribution of the evaluations per region.

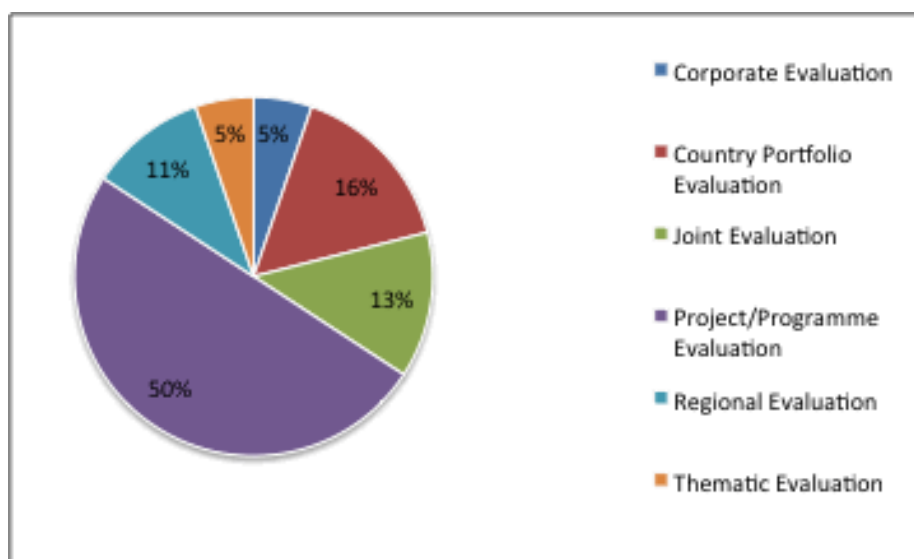
² Scale and their meaning are the following: *Misses requirement* (0-3,49 points); *Approaches requirement* (3,50-6,49 points); and *Meets requirement* (6,50-9,0 points).

Regions	N of reports	%
Americas and the Caribbean	4	11%
Arab States	9	24%
Asia and the Pacific	5	13%
Corporate	2	5%
East and Southern Africa	6	16%
Europe and Central Asia	6	16%
West and Central Africa	6	16%
Grand Total	38	100%

Source: UN-SWAP Assessment Dec 2019-Jan 2020

The evaluations represented a variety of evaluation types, including project/programme evaluations, evaluation of joint programmes, regional/thematic as well as corporate evaluations. The Graph 2 shows the evaluations per type:

Graph 2: Distribution of evaluations per type



Source: UN-SWAP Assessment Dec 2019-Jan 2020

Overall Rating

Overall, as it can be seen in the Table below, the results of the assessments show that the large majority of the UN-Women evaluations (about 68% of the total) obtained an overall rating of “Very good” and/or “Good”. About 26% of them were rated “Fair”. About 5% of them were assessed as “Unsatisfactory” and unsuitable for use since they were missing

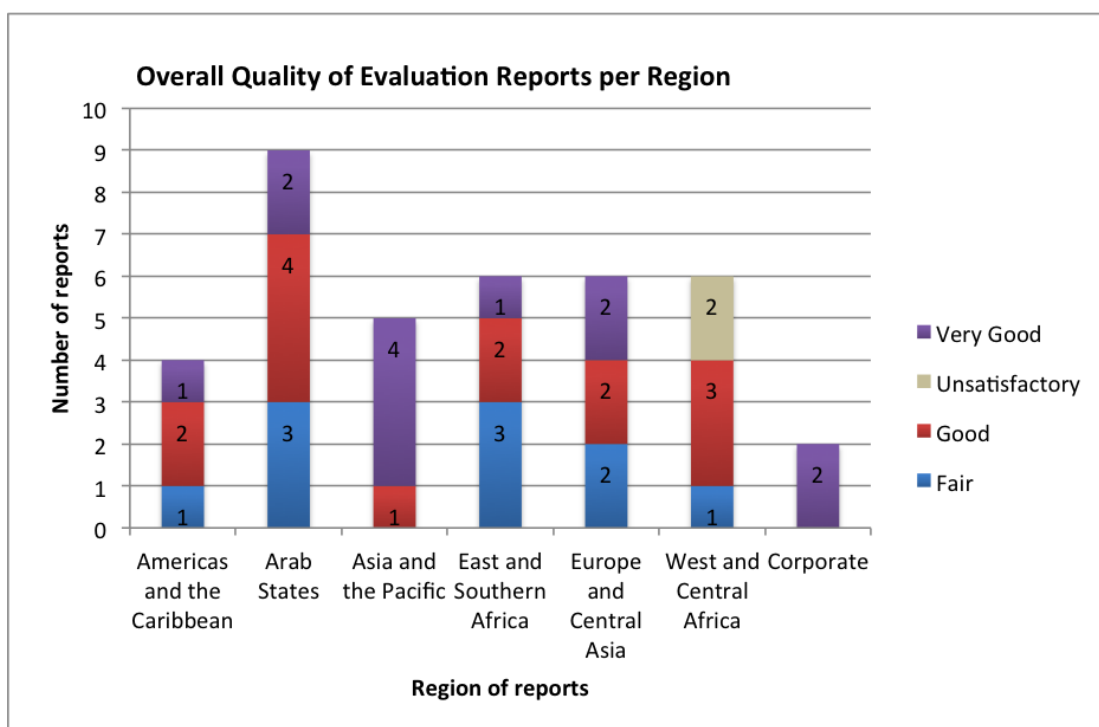
out the minimum quality standards. More information and details about these is included in the pages that follow.

There were variations among these results across the regions. The graph below shows that the Arab states region had the most evaluation reports (9) and also the largest number of “Fair” reports (3) along with the East and Southern Africa region (also with 3 reports rated “Fair”). The West and Central Africa region had 2 reports considered “Unsatisfactory”. The Americas had 1 “Fair” report (and a total of 4 reports). There were also 2 Corporate reports. The Asia and the Pacific region had the largest number of reports rated “Very Good” (4 reports).

Overall Evaluation Rating	# of Evaluations	%
Very Good	12	32%
Good	14	37%
Fair	10	26%
Unsatisfactory	2	5%
Total	38	100%

Source: UN-SWAP Assessment Dec 2019-Jan 2020

More details explaining these ratings are presented below in each of the report sections categories analyzed.

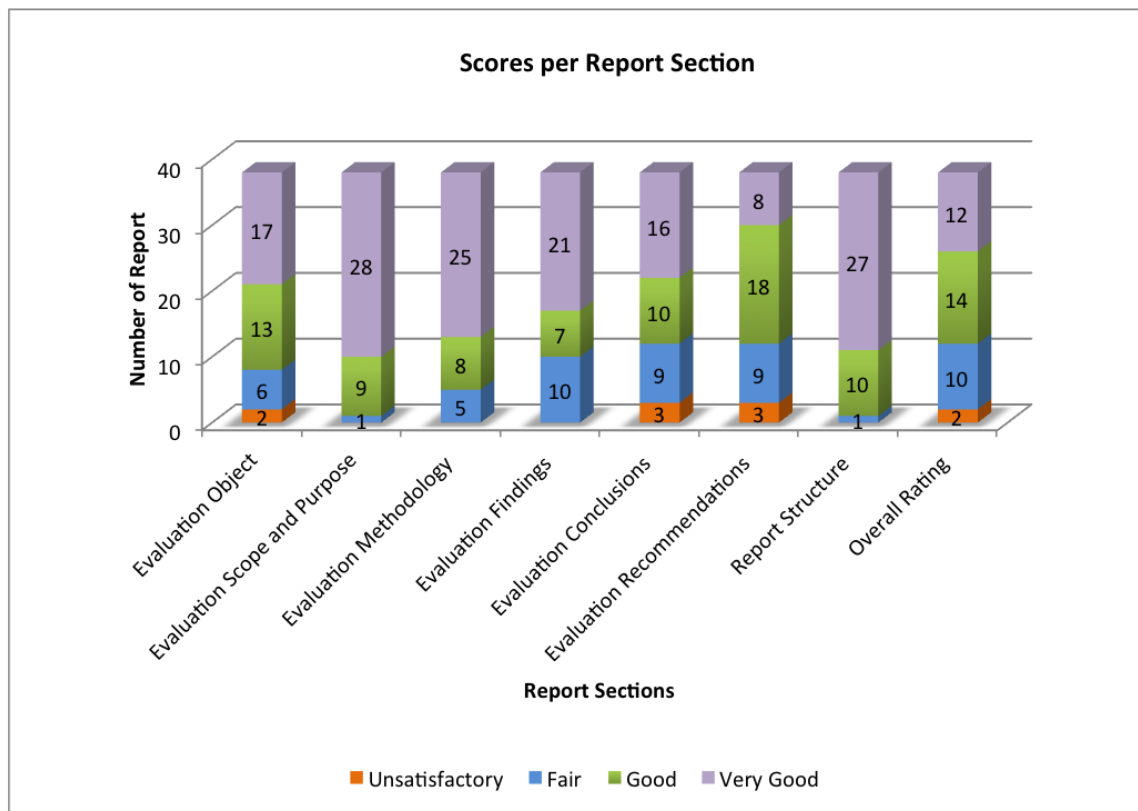


Source: UN-SWAP Assessment Dec 2019-Jan 2020

As per the scores obtained per each of the 8 section rating criterion, while overall the reports scored well in sections such as “Evaluation Scope and Purpose”, “Report Structure”, “Evaluation Methodology”, others sections were more challenging and scores were lower, such as “Evaluation Recommendations” and “Conclusions”. The ratings in

these last sections seemed to have contributed to push their “Overall rating” scores downward. The graph below shows the number of reports rated “Unsatisfactory”, “Fair”, “Good” and/or “Very Good” in each one of them and in the category “Overall rating”.

These individual ratings were the following:



Source: UN-SWAP Assessment Dec 2019-Jan 2020

Object of Evaluation

Box: 2 Evaluation Object Parameters

Does the report present a clear and full description of the 'object' of the evaluation?
1.1 The report clearly specify the object of the evaluation, and provides clear and complete description of the intervention's logic or theory of change, intended beneficiaries by type and by geographic location(s) as well as resources from all sources including humans and budgets, and modalities.
1.2 The context includes factors that have a direct bearing on the object of the evaluation: social, political, economic, demographic, and institutional. This also includes explanation of the contextual gender equality and human rights issues, roles, attitudes and relations.
1.3 The key stakeholders involved in the implementation, including the implementing agency(s) and partners, other stakeholders and their roles are described.
1.4 The report identifies the implementation status of the object, including its phase of implementation and any significant changes (e.g. plans, strategies, logical frameworks) that have occurred over time and explains the implications of those changes for the evaluation.

Most reports rated “Good” and “Very Good” in this parameter. They contained a clear and full description of the evaluation “object” as specified in the criterion above. Only 2 reports rated “Unsatisfactory” and these include the reports also rated “Unsatisfactory” in the overall rating (Cote d’Ivoire) and/or “Fair” (Mozambique). These reports did not include full descriptions of the object but only limited high-level overviews. Six reports rated “Fair” and the rationale for these scores varied. Some reports (specially evaluations of portfolio of projects) provided little information about the portfolio and/or included limited information about the interventions, with information about activities, resources and/or key stakeholders sometimes not included.

One of the evaluations in this category was the Knowledge Management (KM) Evaluation (# 27), which is a “study of the KM practices and inquiry into the KM organizational culture/practice”. Perhaps due to the nature of the evaluation, none of key aspects of the evaluation object were delineated (e.g. ToC, beneficiaries, resources, timeframe of implementation).

Some evaluations also lacked information about the implementation status and, some reports provided very little descriptions of stakeholders (sometimes only listing them). For the reports with lower scores, contextual information was also sometimes limited to national level statistics and little information was provided about the context of capacity needs, which the projects/programmes were designed to address.

Evaluation Purpose, Objectives and Scope

Box: 2 Evaluation Purposes Parameters

Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation?

2.1 Purpose, objectives and use of evaluation: The evaluation report provides a clear explanation of the purpose and the objectives of the evaluation, including the intended use and users of the evaluation and how the information will be used.

2.2 Evaluation Scope: The evaluation report provides a clear description of the scope of the evaluation, including justification of what the evaluation covers and did not cover (thematically, geographically etc) as well as the reasons for this scope (eg. specifications by the ToR, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).

The large majority of reports rated “Good” or “Very Good” regarding this parameter (see Box 2). They provided clear information about the evaluation’s purpose, objectives and scope was also well described, with complete information about the evaluation coverage. This is one of the report sections with better scores and the good ratings reached by most reports in this category may be explained by the fact that most evaluation ToRs include fairly good descriptions of these aspects, which in turn are used in the evaluation reports.

No evaluation reports were rated “Unsatisfactory”. One report rated “Fair” explained by the fact that the evaluation purpose, objectives and scope were not described in the body of the report but were only outlined in the Executive Summary.

Evaluation Methodology

Box 3: Evaluation Methodology Parameters

Is the methodology used for the evaluation clearly described and appropriate, and the rationale for the methodological choice justified?

3.1 Methodology: The report specifies and provides a complete description of the relevant design and sets of methods including the chosen evaluation criteria, questions, and performance standards. The methods employed are appropriate for analysing gender and human rights issues identified in the evaluation scope.

3.2 Data collection, analysis and sampling: The report clearly describes the methods for the data sources, rationale for their selection, data collection and analysis methods. The report includes discussion of how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limitations.

3.3 Stakeholders Consultation: The evaluation report gives a complete description of the stakeholder consultation process in the evaluation, including the rationale for selecting the particular level and activities for consultation.

3.4 Limitations: The report presents a clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias.

3.5 Ethics: The evaluation report includes a discussion of the extent to which the evaluation design included ethical safeguards, mechanisms and measures that were implemented to ensure that the evaluation process conformed to relevant ethical standards including but not limited to, informed consent of participants, confidentiality and avoidance of harm considerations.

Similarly, as for the evaluation methodology, the large majority of evaluation reports (33 out of the 38 reports) rated “Good” or “Very Good” in the methodological criteria specified above. They contained good overview of the methodology with details of the sampling strategy used, descriptions of data collection, analysis and triangulation were included, along with identification of a few limitations. Most reports also provided detailed information in the Annex showing that the methods they used were appropriate for the triangulation, analysis and assessments of GE and HR specific results. Some Annexes included full Inception Reports and consisted of long (200+ pages) separate volumes.

Only 5 reports were rated “Fair” (4). In certain cases, the methodological weaknesses were linked to limited or no information provided about the methodology. Often this was also associated with the fact that there were a limited number of people interviewed/consulted for the evaluation, which did not provide a clear sense of the appropriateness and/or soundness of the methodological approach adopted. Some of these reports did not include evaluation questions and/or criteria. Also, details about the sampling strategy, tools and/or data analysis were limited. In most cases the evaluations referred to ethics and/or to the adoption of the UNEG principles and/or UN Women Evaluation Policy.

Evaluation Findings

Box 4: Evaluation Findings Parameters

Are the findings clearly presented, relevant and based on evidence?
4.1 The evaluation report findings provide sufficient levels of high quality evidence to systematically address all of the evaluation questions and criteria.
4.2 Findings are clearly supported by and respond to the evidence presented, reflecting systematic and appropriate analysis and interpretation of the data; they are free from subjective judgements.
4.3 The causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified.
4.4 Findings are presented with clarity, logic and coherence (e.g. avoid ambiguities).

The large majority of evaluations also scored “Very Good” (21) and/or “Good” (7) regarding the criteria established for the evaluation findings (as described in the box below). Findings systematically addressed the evaluation questions and criteria. They were presented with clarity and logic, and are properly grounded on evidence. In most cases, outputs and indicators are assessed individually, supported by detailed information and explanations of causal factors. Overall, these evaluations findings were well substantiated, and showed that their analysis was grounded on multiple lines of evidence.

Ten reports scored “Fair” (10)³ regarding their findings. Most issues found in reports that scored poorly in this category relate to “descriptions” as opposed to clear “assessments”, limited explanations of causal factors and/or poor or no substantiation of the findings.

Evaluation Conclusions

Box 5: Evaluation Conclusions Parameters

Are the conclusions clearly presented based on findings and substantiated by evidence?
5.1 Conclusions are well substantiated by the evidence presented and are logically connected to evaluation findings.
5.2 The conclusions reflect reasonable evaluative judgments that add insight and analysis beyond the findings
5.3 Conclusions present the strengths and weaknesses of the object (policy, programmes, projects or other intervention) being evaluated, based on the evidence presented and taking due account of the views of a diverse cross-section of stakeholders.
5.4 Lessons Learned: When presented, the lessons learned section stems logically from the findings, presents an analysis of how they can be applied to different contexts and/or different sectors, and takes into account evidential limitations such as generalizing from single point observations.

As for Conclusions, while 26 evaluations were rated “Very Good” or “Good”, 9 evaluations were rated “Fair” and 3 reports were rated “Unsatisfactory”.⁴ The Conclusions in those reports rated “Very Good” and/or “Good” were usually well formulated, directly connected to the findings, and provided clear responses to the evaluation questions. They also contained a balance of identified strengths and weaknesses and were broad, going beyond the findings, providing more explanations and

³ Algeria, Bosnia & Herzegovina, Ecuador, Kenya, Liberia, Malawi, Morocco, Mozambique and Cote d’Ivoire (2).

⁴ Algeria, Georgia, Malawi, Mali, Morocco and Regional Reports (2) and Cote d’Ivoire rated “Fair” and Kenya, Sierra Leone and Zimbabwe rated “Unsatisfactory”.

insights to them. They also contained good sets of lessons learned well articulated and generalized.

As for those rated “Fair” and “Unsatisfactory”, in most cases, these evaluations also scored poorly in the evaluation findings section and these poor ratings persisted throughout the evaluation conclusions, since conclusions were connected to the findings in most cases. As such, conclusions also reflected the same issues identified earlier, including limited substantiation, limited analysis and/or limited evidence. Most of the time, the conclusions of these reports that rated poorly, did not add insights to the findings, and the “analysis” presented only a “summary of the findings”. Sometimes they were presented in a really summarized way (e.g. point form). Other times, they were disproportional regarding the level of issues identified earlier in the findings.

Issues with lessons learned also had a key effect regarding pushing these ratings downward. In many cases, lessons learned were either not included or they were not well articulated as “lessons” per se. They were not generalized and did not clearly articulated what worked well or not in the project experience with a view to use the lesson in other projects/initiatives.

Evaluation Recommendations,

Box 6: Evaluation Recommendations Parameters

Are the recommendations relevant, useful, actionable and clearly presented in a priority order?

6.1 Recommendations are logically derived from the findings and/or conclusions.

6.2 The report describes the process followed in developing the recommendations including consultation with stakeholders.

6.3 Recommendations are clear, realistic (e.g. reflect an understanding of the subject's potential constraints to follow-up) and actionable.

6.4 Clear prioritization and/or classification of recommendations to support use.

Ratings for recommendations followed similar patterns as Conclusions ratings. While the majority of evaluations (26 out of the 38 evaluations) were rated “Very Good” and “Good”, a larger number of evaluations were also rated “Unsatisfactory” (3) or “Fair” (9) regarding the criteria set up in the box above.

Recommendations of those reports that rated “Good” or “Very Good” derived from findings and conclusions, were well articulated, actionable, contained prioritization and/or classification information and were directed to specific agencies for implementation.

As for those reports rating “Unsatisfactory” or “Fair”, some recommendations did not reflect an understanding of the constraints, or were logically connected with arguments presented earlier. Some were not articulated as “recommendations” and others did not contain enough details to enable implementation. Various evaluations only included these “partially”.

Criteria related to prioritization (6.4) and process of formulation of recommendations (6.2) also contributed to pushing the ratings downward generally. Many evaluations did

not refer to the process followed to develop the recommendations and/or include prioritization.

Gender, Human Rights and the SDGs⁵

Box 7: Gender, Human Rights and SDGs Parameters

Does the evaluation meet UN SWAP evaluation performance indicators? Note: this section will be rated according to UN SWAP standards.

7.1 GEWE is integrated in the evaluation scope of analysis and evaluation criteria and questions are designed in a way that ensures GEWE related data will be collected.

7.2 A gender-responsive methodology, methods and tools, and data analysis techniques are selected.

7.3 The evaluation findings, conclusions and recommendation reflect a gender analysis.

Gender, Human Rights and the SDGs assessments followed the criteria set up by the UN SWAP evaluation performance indicators, as highlighted in the box below. The rating system qualified each element as “Misses requirement” “Approaches requirement” or, “Meet requirement”, as explained earlier based on the ratings obtained in each of the criteria specified in the Box 7.

The table below show the overall results. The large majority of the evaluations assessed (72%) “meets the requirements”. About 22% of evaluations “approached the requirements”, with some parameters met, and only 5% (or 2 evaluations) “missed the requirements”.

Table: Evaluation scores in overall GEEW integration

Individual Evaluation Scores	# of Evaluations	%
Meets requirements	28	72%
Approaches requirements	8	22%
Miss Requirements	2	5%
Total	38	100%

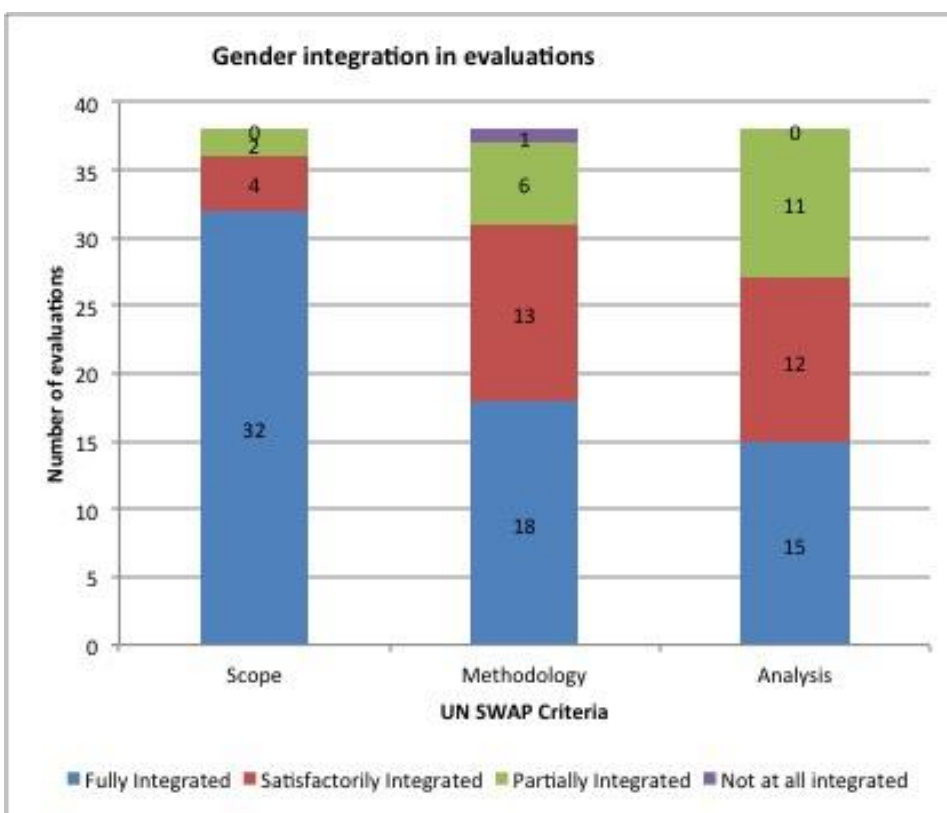
Source: UN-SWAP Assessment Dec 2019-Jan 2020

The 2 evaluations that missed the requirements were the evaluations from Cote d’Ivoire (# 26 and #32) as mentioned, which were written by the same consultant. They contained little evidence of gender integration in their scope. There was also little evidence the methods were appropriate for the triangulation, analysis and assessments project results, let alone gender-related specific results. They contained some gender-related analysis, but their overall scores were very low.

Regarding analysis of scores obtained in each individual criterion of assessments of GEEW integration (1- scope; 2- methodology and 3- analysis), generally, all 38 UN Women evaluations scored well in these individually, with the exception of those that missed the requirements,

⁵ This section contains specific more detailed explanations of the ratings in each subcriteria, extracted form a separate report on the UN SWAP assessment, prepared for UN Women.

As illustrated in the graph below, regarding integration of gender in the evaluations scope, the large majority of evaluations (95% of all evaluations assessed) integrated GEEW fully or partially in the evaluation scope, and contained evaluation criteria and questions designed to assess GEEW related outcomes. Only 2 reports integrated GEEW partially in their scope as explained and those include those 2 reports that missed the requirements.



As for gender integration in the methodology, while the majority of evaluations integrated gender, there were variations regarding the depth and breadth of integration across the evaluations assessed. While almost all evaluations referred to the adoption of UNEG and/or UN-Women Evaluation policy and guidance for gender-responsive evaluations, just about half of the evaluations (18) demonstrated full integration by describing how that was done, showing the methods and tools used, and presenting sampling frames disaggregated by sex.

There were also some evaluations (13) that referred to the adoption of the same principles but did not provide detailed explanations of the appropriateness of methodology. Yet, they presented a sampling frame disaggregated by sex, and/or described ethical standards and/or gender/inclusion data collection processes. Several variations were found regarding these various parameters and these were rated “satisfactorily integrated”. Those rated as “partially” integrating gender, (6 evaluations) did not include additional

information and/or only described methodology too generally, yet they contained references to gender-responsive methodology and/or standards. Only one report (which also missed the requirements) did not show evidence of integration of gender in methodology at all.

All evaluations integrated gender in their analysis. However the rate of integration was varied, mostly regarding the extent to which they integrated in all aspects (findings, conclusions, recommendations and background sections) or only a few. Also, the assessment took into consideration the quality of analysis provided, even though this was not a specific criterion delineated in the individual scoring tool.

About 15 evaluations fully integrated GEEW in all aspects with solid in-depth analysis. Eleven evaluations satisfactorily integrated GEEW in findings but the extent to which they also integrated GEEW in their conclusions and/or recommendations were varied. Only 11 evaluations partially integrated GEEW in their analysis, with issues ranging from weaknesses and/or focus of the evaluations reports (as identified above, in line with evaluations missing requirements) and/or poor quality of analysis.

Report Structure

Box 8: Report Structure Parameters

Is the report well structured, written in accessible language and well presented?
8.1 Report is logically structured, well written and presented with clarity and coherence (e.g. the structure and presentation is easy to identify and navigate (numbered sections, clear titles and subtitles, context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations) and is written in accessible language with minimal grammatical, spelling or punctuation errors.
8.2 The title page and opening pages provide key basic information on the name of evaluand, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents including, as relevant: tables, graphs, figures, annexes-; list of acronyms/abbreviations, page numbers.
8.3 The Executive Summary is a stand-alone section that includes an overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key findings, conclusions and recommendations. The Executive summary should be reasonably concise.
8.4 Annexes should include, when not present in the body of the report: ToR, evaluation matrix, list of interviewees, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence. Other appropriate annexes could include: additional details on methodology, copy of the results chain, information about the evaluator(s).

Almost all reports scored well in this criterion (see box above). The large majority of the reports assessed (37 out of the 38) were rated “Very Good” and “Good”, depending on the completeness of the information provided. Some contained good graphs and figures, to present complex information, as highlighted below as “good practices”.

Overall, most reports were well written and contained appropriate structure. They were also well presented and contained a good level of information both in the Annex and in the Executive Summary. Consistent with the scores obtained earlier, one reports was rated “Fair”⁶ due to limited and/or incomplete information provided.

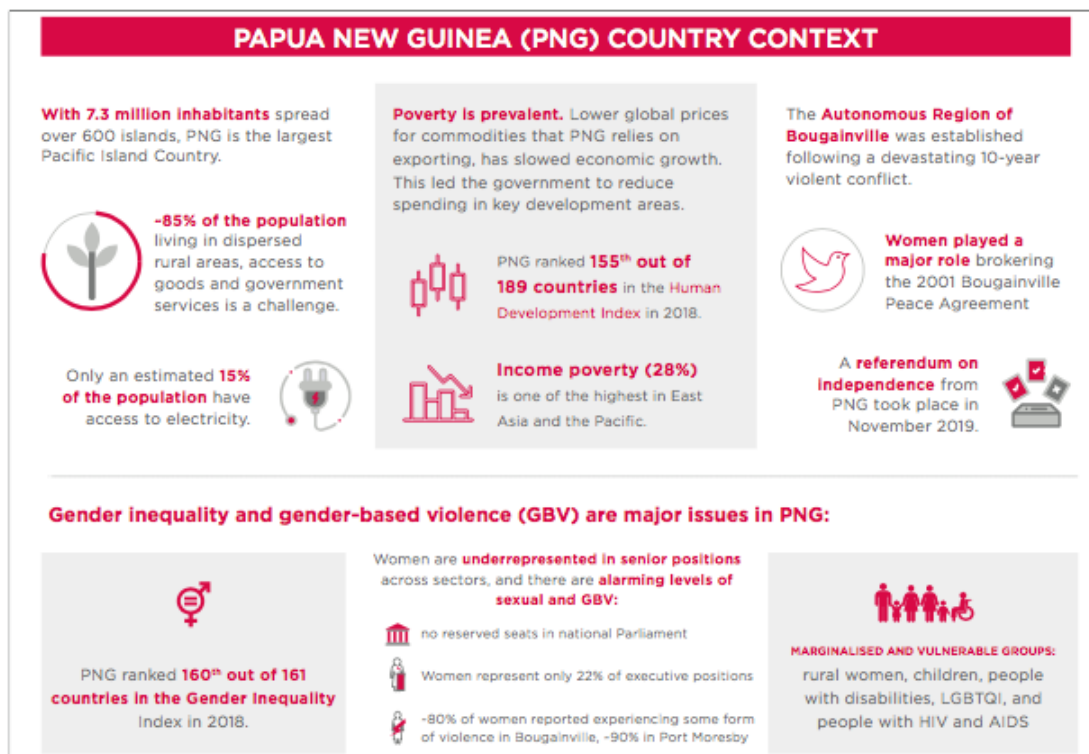
Good practices and trends

There were several good practices and trends identified in the evaluation reports assessed.

1. **Extremely well conceptualized graphs and figures are used to provide a good synthesis, yet comprehensive presentation of information.** Technology and software developments have enable this trend to emerge with more reports using graphs, infographs and figures to synthesize complex information. This could be more encouraged through dissemination of examples to show the possibilities and broaden their use by the UN-Women evaluation community.

For example, the UN Women Country Portfolio Evaluation Papua New Guinea (#38) uses very good infographs to summarize the background, the evaluation object, methodology, presenting a range of information in a snapshot. See example below.

⁶ Morocco.



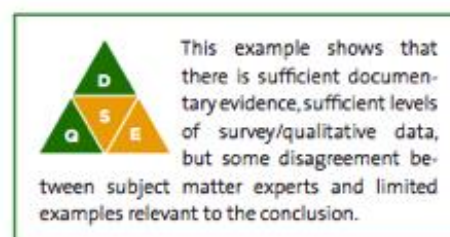
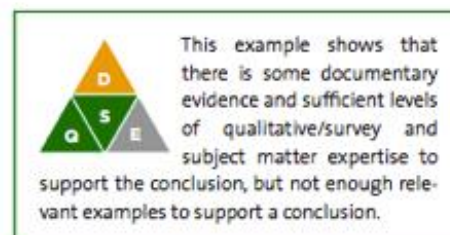
Similarly, the Evaluation of the Guatemala Strategic Note 2015-2019 (#35) presents the results of analysis of the indicators (output, outcome and impact level) using an interesting matrix with results per thematic area. While it is not possible for this QA to validate the assessment (as it was based on individual analysis of each indicator (and the report does not show such analysis), the format and categories used are a great way of presenting the data.

Other examples worth citing are the Evaluation of Strategic Note 2014-2019 in Mexico (#19), which contains an excellent graph showing relationships between findings, conclusions and recommendations, the Evaluation of the 'Port Moresby: A Safe City for Women and Girls Programme (#22) with very good ways to present information.

2. **The evaluations contain excellent analysis, grounded on solid methodological approach, regarding issues and/or opportunities to empower of women as well as lessons on gender-related programming. Some evaluations also developed gender-related tools for their evaluation assessments.** Most evaluations rated "Very Good" are in this category. As an example of gender-related evaluation tool, the same of the Guatemala Strategic Note 2015-2019 (#35) lists, as a result of the analysis a series of "unintended" results related to empowerment of women at the individual, family, organizational and community level which seems to be an excellent "road map" upon which other evaluations can base their data collection and assessments (see pp. 45).

3. **Presenting triangulation of lines of evidence at a more granular level seems to be a trend.** Considering that not all findings and conclusions are always based on the same level of convergence of information, it makes sense that evaluations should contain some indication of such level of convergence between different sources of information for assessments, showing the completeness (or lack of thereof) of data/information for each area upon which the judgments are made. The evaluation #8 - *Corporate Evaluation of UN-Women's Contribution to Humanitarian Action* used a graph to measure the level of “triangulation” used in each conclusion put forward, and included primary qualitative/survey data (Q) from the evaluation, documentary evidence (D), examples of UN Women’s work (E) and the subject matter expertise of the evaluation team (S). See example on the side.

This provides a shorthand of how and if all data sources were used. These are summarized as being either Grey (no data), Amber (some data), or Green (sufficient data) as below:



Source: Corporate Evaluation of UN-Women's Contribution to Humanitarian Action

4. **Simple ways are used to show triangulation of evidence:** One evaluation - the Evaluation of UN Women's Contribution to Women's Economic Empowerment in Nepal (#36) used footnotes throughout the report to back statements made through Key Informant Interviews. Because data collected was codified and anonymised to respect confidentiality, and KII were associated with a specific numbered, it was ready for use as references, without compromising confidentiality. The report referred to these statements using their numbers (KII 1, KII 3) and showed triangulation by referencing to 2 or more KIIs and other sources.
5. **Contribution Scores” and “Evidence Scores” used to qualify each “contribution” identified by the evaluation and to provide assurance of the evidence associated with it.** The Evaluation of the ‘Port Moresby: A Safe City for Women and Girls Programme’ (#22) uses simple 4-point scales to demonstrate the level of importance of the contributions in line with the project Theory of Change – “contribution score” as well as a score of confidence in the supporting findings, depending on the relative strength of the evidence.

Women's/girls' empowerment and opportunities due to their access and use of safe markets³

Contribution Score ★ ★ ★

Evidence Score ★ ★ ★

Through the collective association into Market Vendor Associations (MVAs), previously marginalized women vendors developed a collective identity. Through the establishment of multi-stakeholder Market Management Committees, for the first time, they also obtained a voice in market management. This contributed to a decline in ethnic violence and VAWG inside the markets, and an increase in women's sense of collective power and ability to exert influence and negotiate their position and interests.

Source: Evaluation of the 'Port Moresby: A Safe City for Women and Girls Programme'

6. **Complex regional programme level findings well articulated and properly substantiated in synthesized and concise way, yet with a good level of explanations and country level findings.** The assessments of complex regional programmes involving interventions in multiple countries is very challenging and many evaluations fail to convey comprehensive high level and detailed information. The Evaluation of the 3rd Phase of the regional GRB programme (#11) used an interesting approach to overcome the challenge by presenting regional-level information in the body of the report, complemented by county level data presented (as separate reports) in the Annex. Findings are articulated around the evaluation criteria at the regional/programme level in the body of the report, complemented with details of programme impact, strength and weaknesses at the country level in the Annex. Country-level responses to all evaluation questions are also included in the Annex along with detailed assessments of results against outputs and indicators. The findings are presented with clarity and logic and information in the Annex show they are supported by evidence.
7. **Linkages between the Theory of Change and the evaluation assessments are clearly demonstrated.** Often a challenge for evaluations, the Corporate Evaluation of UN-Women's Contribution to Governance and National Planning provides an illustration of an effective and simple way to show this. The evaluation used a simple table, describing the ToC component, the evaluation questions, the underlying nature of causal connections and main assumptions effectively showing the relationships and comprehensiveness of the assessments made (See Table below)

TABLE 2
Key assumptions underpinning the Theory of Change

ToC Component	Evaluation Question	Main Assumptions	Nature of Causal Connections
Mandates to Strategies	1, 4, 6, 8, 9	<ul style="list-style-type: none"> Normative, coordination and operational roles complement and reinforce one another 	Attribution (direct causal linkage)
Strategies to output	1, 4, 6, 8, 9	<ul style="list-style-type: none"> Availability of financial and human resources within UN Women and its key partners Active participation of the target groups in activities 	Attribution (direct causal linkage)
Outputs to outcomes	2, 3, 4, 5, 7	<ul style="list-style-type: none"> Willingness and resources among key stakeholders to use new or strengthened skills, knowledge, data and tools Political will to support commitments towards gender equality and HIV/AIDS in national development priorities, economic policies, legislation & governance structures among national governments, civil society and donors Political consensus can be reached through data and evidence Gender advocates and organizations, and networks of WLWHIV have the capacity to influence national financing decisions Adequate capacity of the various actors to plan, budget, deliver and monitor performance in terms of gender equality and HIV/AIDS obligations Planning and budgeting systems and overarching macro-economic frameworks are coherent with gender equality objectives 	Significant/plausible contribution (main contributing factor)
Outcomes to impact	5, 8, 9, 11	<ul style="list-style-type: none"> Adequate funds are allocated to implement gender equality commitments as evidenced by increased benefit to women and girls from public services and resources Social norms and gender roles allow women and girls to use services, and to participate in economic decision-making activities Supportive political, cultural, social and economic contexts 	Remote contribution (influencing factor)

8. **Evaluations used key principles for Ethical and VAWG Responsive Research to guide evaluation work.** Several reports contain excellent descriptions of such principles. Some refer to data collection approach guided by the context sensitivity and Do No Harm principles, with consideration to data security and safeguarding. Others include the analysis of structural gender inequalities that prevent women from meaningfully engaging in politics and peacebuilding (e.g. Syria) and taking them into consideration in the consultation exercises, along with a strict ethical protocol. Some also used approval of the tools by ethical review boards of the respective UN organizations and the adoption of ethical standards for interviewing survivors or GBV and children.⁷

⁷ Other examples include: the Evaluation of the ‘Port Moresby: A Safe City for Women and Girls Programme’ (#22), the Evaluation of the UN Women’s Supporting Syrian Women’s Engagement in the Syrian Political Process (#21), the Mali Country Programme Evaluation (#17); the Final evaluation for the Hemayati: Promoting Women and Girls’ health and wellbeing joint programme (#5); Project evaluation for Eid bi Eid (#3); the Programme Presence Portfolio Evaluation (#6).

5. ASSESSMENT STRENGTHS AND WEAKNESSES

The following are key strengths and weaknesses highlighted in the assessment:

Key Strengths:

Across all regions, a large number of evaluations had excellent score. This is likely due to the existence within UN Women Country Offices of capacities (within staff and consultants) to prepare comprehensive Terms of Reference, providing sound direction and guidance regarding evaluation approach, methodology and content of the evaluation reports.

Some evaluations provide good descriptions of the gender-responsive methodological approaches undertaken: there is a wealth of information across the evaluations on methods and principles used, some of which go beyond simply integrating gender in the evaluation scope and methodology but are proactively used to engage and foster gender equality.

UN Women evaluations consistently scores well in UN SWAP assessments: the mandate of the entity largely contributes to good scores. Even the three evaluations that “missed” the GEEW requirements and scored poorly regarding GEEW integration in scope and methodology, integrated GEEW in findings, conclusions and/or recommendations. This is explained by the fact that even in cases where evaluation scope and methodology do not integrate GEEW specifically, these evaluations are assessing frameworks/programmes or projects supporting UN Women initiatives and the implementation of gender initiatives, which impacted on women and girls. As such, there is clear integration of GEEW in evaluation findings, conclusions and recommendations.

A wealth of knowledge about factors and constrains to empowering women: gathered through the excellent analysis presented in many evaluations. These evaluations contain rich, well researched documentation about women reality and obstacles to gender equality worldwide, which constitute a true asset.

Key weaknesses:

Some regions are still lagging behind with the concentration of a few poorer quality evaluation reports in Africa and Arab States primarily, including West and Central Africa (Cote d’Ivoire); and East and Southern Africa (Kenya and Mozambique) and Arab States (Morocco & Jordan). The fact that the Cote d’Ivoire used the same consultant for two equally poor evaluations may be an indication of limited capacities at the level of the CO or limited availability of consultants properly trained in evaluation to undertake the work.

Many evaluations did not provide detailed explanations of the appropriateness of methodology to incorporate GEEW: while most all of the evaluations referred to the adoption of UNEG and/or UN-Women Evaluation policy and guidance for gender-responsive evaluations, several evaluations provided limited information about how such

integration was to be done, with no or only general references to GEEW in the methods and tools used.

Several reports are long, and some are extremely long which could may discourage their use as relevant information to guide decision-making and further programming: out of the 38 evaluations assessed, 15 evaluations (about 40%) contained more than 60 pages (excluding Annexes). About 8 evaluations (20%) were longer than 80 pages with several (5 evaluations) contained more than 100 pages. One evaluation was more than 134 pages long and the findings section started in page # 95. Useful evaluations are typically easy to read and “to the point” regarding their content, clearly addressing the issues and containing high level and yet well substantiated information. Most of these evaluations scored “Good” and some “Fair” and/or “Unsatisfactory” (with only 4 of these evaluations considered Very Good) suggesting that length is not necessarily indicative of quality.

6. RECOMMENDATIONS

Based on these, the following are some high-level recommendations for improvements in the GERAAS for UN Women consideration:

- a) Include assessments of the ToRs as criteria for GERAAS assessments, in addition to the 8 criteria established by UNEG. The ToRs are key elements of the evaluation process. In many cases, the ToRs establish the evaluation criteria and questions, indicate the approach and methodology and set the expectations regarding the content of the evaluation report. Adding the ToRs as an additional criterion will likely draw the attention of the country offices to the need to follow certain parameters for its formulation, ensuring some minimum coverage standards. **It is recommended that UN Women (IES) design a set of criteria and add the ToRs as a stand-alone criterion against which the evaluation reports are assessed in the GERAAS.**
- b) Set up some standards to “qualify” an evaluation for GERAAS assessments: certain evaluations were included in the universe of these assessments but were not “typical” evaluation studies (e.g. the Knowledge Management evaluation was a “study” of practices and organizational culture). Another good example of this is the Evaluation of the Campaign #hearmetoo (#39) which in the end was excluded from the scope of this exercise as it was a “marketing” evaluation geared towards assessing if the “communication” objectives of a marketing campaign/strategy were met. It was not an “UNEG evaluation” which typically assesses the achievement and or contributions by examining results chains, processes and contextual factors and causality.⁸ As a result of the inclusion of these evaluations, which typically would score poorly due to their non-

⁸ **UNEG definition of evaluation:** An evaluation is an assessment, conducted as systematically and impartially as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area^{SEP} or institutional performance. It analyses the level of achievement of both expected and unexpected results by examining the results chain, processes, contextual factors and causality using appropriate criteria such as relevance, effectiveness, efficiency, impact and sustainability.

- conformance with the criteria, they push the overall scores downward. **It is recommended that UN Women (IES) establish certain standards to ensure the studies can be considered UNEG “evaluations” which can be assessed in the GERAAS.**
- c) Establish parameters to limit the length of the evaluation reports as a good practice to ensure better use of evaluation products and harmonize reports to ensure QA assessments level of comparability is fair. Not only long reports are discouraging for the evaluation user, but in lieu of a parameter to ensure reports contain the same level of information, there are lots of variability in terms of levels of information provided and assessed for quality assurance. For example, certain reports included the entire Inception Report in the Annex, providing a high-level of in-depth information about their methodological approach, maximizing their chances of obtaining high scores in “Methodology”. Other reports which limited their methodology descriptions to synthesis of their Inception Reports, could be penalize for not providing full information on all criteria. One example of this was the 200+ page Annex provided by one of the evaluations. To ensure QA assessments provide fair ratings and evaluation products are more useful, **it is recommended that UN Women (IES) establish a maximum length of pages for the evaluation reports and their annexes, and also establish a limit to the size of document (in MB), which can be uploaded to the platform.**
- d) Similarly, it is important to ensure that evaluations submitted for Quality Assurance Review follow certain parameters as to enhance fairness across participants submitting reports. Deadlines should be firmly established for the submission of the reports. All Evaluators and Country Office staff should be given the same length of time for their submission as extensions should not be granted. On-going granting of extensions penalizes those who rush to provide a product (sometimes without fully completion), while providing more opportunities to those who were late (and have more time to submit a complete product). Similarly, once reports are submitted, their submission should be final, and subsequent revisions to reports should not be accommodated. This not only disrupts the quality review process but creates unfair opportunities for those submitting reports (who can not submit newer and improved versions). **It is recommended that UN Women (IES) establish and honour deadlines established for submission of evaluation reports for quality assurance and reviews its practices of enabling the submission of multiple versions of reports for quality assurance.**
- e) There is a wealth of information across the evaluations on evaluation methods and principles used, some of which go beyond simply integrating gender in the evaluation scope and methodology but are proactively using evaluations to engage and foster gender equality. UN Women should capitalize on this experience and use these evaluations descriptions of methods and tools to provide more guidance to Consultants. **It is recommended that UN Women (IES) consider preparing**

guidance materials on how to integrate GEEW in evaluation methodology, using the evaluations examples and make them available to those consultants contracted to undertake UN Women evaluations, at the minimum.

- f) To address the issue of some regions lagging behind regarding the quality of their evaluation reports, UN Women should invest in targeted training to M&E personnel in selected Country Offices, to build evaluation capacities. Some good training is available in Canada and other countries, which could provide the staff with good solid basis to manage evaluations and provide consultants with some guidance and direction for improvements of evaluation methodologies and overall reporting. **It is recommended that UN Women (IES) offers targeted training to M&E staff in selected COs, including West and Central Africa (Cote d'Ivoire); and East and Southern Africa (Kenya and Mozambique) and Arab States (Morocco & Jordan).**
- g) **There is also an opportunity for UN-Women to systematize and disseminate a wealth of “theoretical” knowledge on gender, which can be applied to programming and capacity building.** Similarly to the example provided above regarding knowledge obtained about factors and constrains to empowering women provided in one evaluation, there is a wealth of information and documentation in all thematic areas of UN-Women interventions gathered through well research evaluations. **It is recommended that UN-Women (IES) undertake a thorough assessments of evaluations rated Very Good (as overall ratings), with a view to extract materials, analytical frameworks, and other “lessons” documented by these evaluations, and to incorporate them in its knowledge management practices.** This includes systematizing these useful information and disseminating them either through publications and/or its use in training, capacity building and/or programming of interventions, consistently. This practice should be adopted by UN-Women on on-going basis, perhaps on an annual basis possibly after each QA cycle rates the evaluations.

ANNEX 1: List of evaluations covered in the 2019 GERAAS and UN-SWAP Assessments

#	Region	Country	TITLE	TYPE
1	Americas and the Caribbean	Bolivia	Evaluation of UNW's programme support (partnerships)	Thematic Evaluation
2	Arab States	Regional	Formative Evaluation of Community and National Level Actions for Promoting Gender Equality and Engaging Men and Boys	Regional Evaluation
3	Arab States	Jordan	UN Women Programme Eid Bi Eid I & II Final Evaluation	Project/Programme Evaluation
4	Arab States	Jordan	Final evaluation for the National Strategy for Women	Project/Programme Evaluation
5	Arab States	Jordan	Final evaluation for the Hemayati: Promoting Women and Girls' health and wellbeing joint programme	Joint Evaluation
6	Asia and the Pacific	Regional	Programme Presence Portfolio Evaluation in Asia and the Pacific	Country Portfolio Evaluation
7	Corporate	IES	Corporate Evaluation of UN-Women's Contribution to Governance and National Planning	Corporate Evaluation
8	Corporate	IES	Corporate Evaluation of UN-Women's Contribution to Humanitarian Action	Corporate Evaluation
9	East and Southern Africa	Zimbabwe	Final Evaluation of the Gender, Peace and Security Programme in Zimbabwe	Project/Programme Evaluation
10	East and Southern Africa	Ethiopia	Evaluation of JP Rural Women Economic empowerment	Project/Programme Evaluation
11	Europe and Central Asia	Regional	Evaluation of the third phase of the Promoting Gender Responsive Policies in South East Europe Programme	Regional Evaluation
12	Europe and Central Asia	Georgia	Final evaluation of the EU supported EVAWG project "UNiTe to Fight Violence against Women in Georgia"	Project/Programme Evaluation
13	Europe and Central Asia	Georgia	Mid-term evaluation of UN Joint Programme For Gender Equality	Joint Evaluation
14	Europe and Central Asia	Bosnia & Herzegovina	Final Evaluation of the project "Standards and Engagement for Ending Violence against Women and Domestic Violence in Bosnia and Herzegovina"	Project/Programme Evaluation
15	Europe and Central Asia	Serbia	Evaluation of the Serbia National Action Plan for GE	Project/Programme Evaluation
16	Europe and Central Asia	Ukraine	Mid-Term Evaluation: Enhancing Accountability for Gender Equality and Women's Empowerment in National Reforms, Peace and Security	Europe and Central Asia
17	West and Central Africa	Mali	Final Evaluation Report Mali Country Programme 2014 – 2019	Country Portfolio Evaluation
18	West and Central Africa	Liberia	Final Project Evaluation: "Inclusive Security Nothing for Us Without Us"	Joint Evaluation
19	Americas and the Caribbean	Mexico	Final Evaluation of the Strategic Note of UN Women Mexico, 2014-2019	Country Portfolio Evaluation
20	Arab States	Palestine	Final Evaluation of UN Women Project "Advancing the Implementation of UNSCR 1325 in the oPt"	Project/Programme Evaluation
21	Arab States	Regional	UN Women's Supporting Syrian Women's Engagement in the Syrian Political Process	Regional Evaluation
22	Asia and the Pacific	Papua New Guinea	Evaluation of the 'Port Moresby: A Safe City for Women and Girls Programme'	Project/Programme Evaluation
23	East and Southern Africa	Mozambique	Final Evaluation of the WEE Project in Gaza, Mozambique	Project/Programme Evaluation
24	East and Southern Africa	Kenya	Final Evaluation Integrating Gender into Peace Support Operations in Eastern Africa	Project/Programme Evaluation

25	West and Central Africa	Sierra Leone	Evaluation of the Improving Women's Participation in Political Processes as Peace building Ambassadors Project - Sierra Leone	Joint Evaluation
26	West and Central Africa	Cote d'Ivoire	Evaluation "Women and Young Girls, Actors in Conflict Prevention through Early Warning and Information Networks"	Joint Evaluation
27	West and Central Africa	Regional	Knowledge Management Evaluation	Regional Evaluation
28	Asia and the Pacific	Bangladesh	Country Portfolio Evaluation	Country Portfolio Evaluation
29	Arab States	Algeria	Strengthening the effectiveness of equal rights between men and women in Algeria	Project/Programme Evaluation
30	Arab States	Morocco	Evaluation of the Project "Support for the economic empowerment of women through the promotion of the agro-ecological value chain for better resilience to climate change"	Project/Programme Evaluation
31	Arab States	Morocco	Evaluation of the project "Access to judicial services for women and children victims of human trafficking in Morocco"	Project/Programme Evaluation
32	West and Central Africa	Cote d'Ivoire	Evaluation of the project Women Empowerment in the Shea Sector	Project/Programme Evaluation
33	East and Southern Africa	Malawi	End of Project evaluation for the Advancing and Sustaining Gender Based Governance Project in Malawi	Project/Programme Evaluation
34	East and Southern Africa	Malawi	Mid term evaluation for the Women Empowerment Programme	Project/Programme Evaluation
35	Americas and the Caribbean	Guatemala	Evaluation of UN-Women Guatemala Strategic Note 2015-2019	Country Portfolio Evaluation
36	Americas and the Caribbean	Ecuador	Impact Evaluation of the "Quito: Safe City for Women and Girls Program"	Project/Programme Evaluation
37	Asia and the Pacific	Nepal	UN Women's Contribution to Women's Economic Empowerment	Thematic Evaluation
38	Asia and the Pacific	Papua New Guinea	UN Women Country Portfolio Evaluation Papua New Guinea	Country Portfolio Evaluation