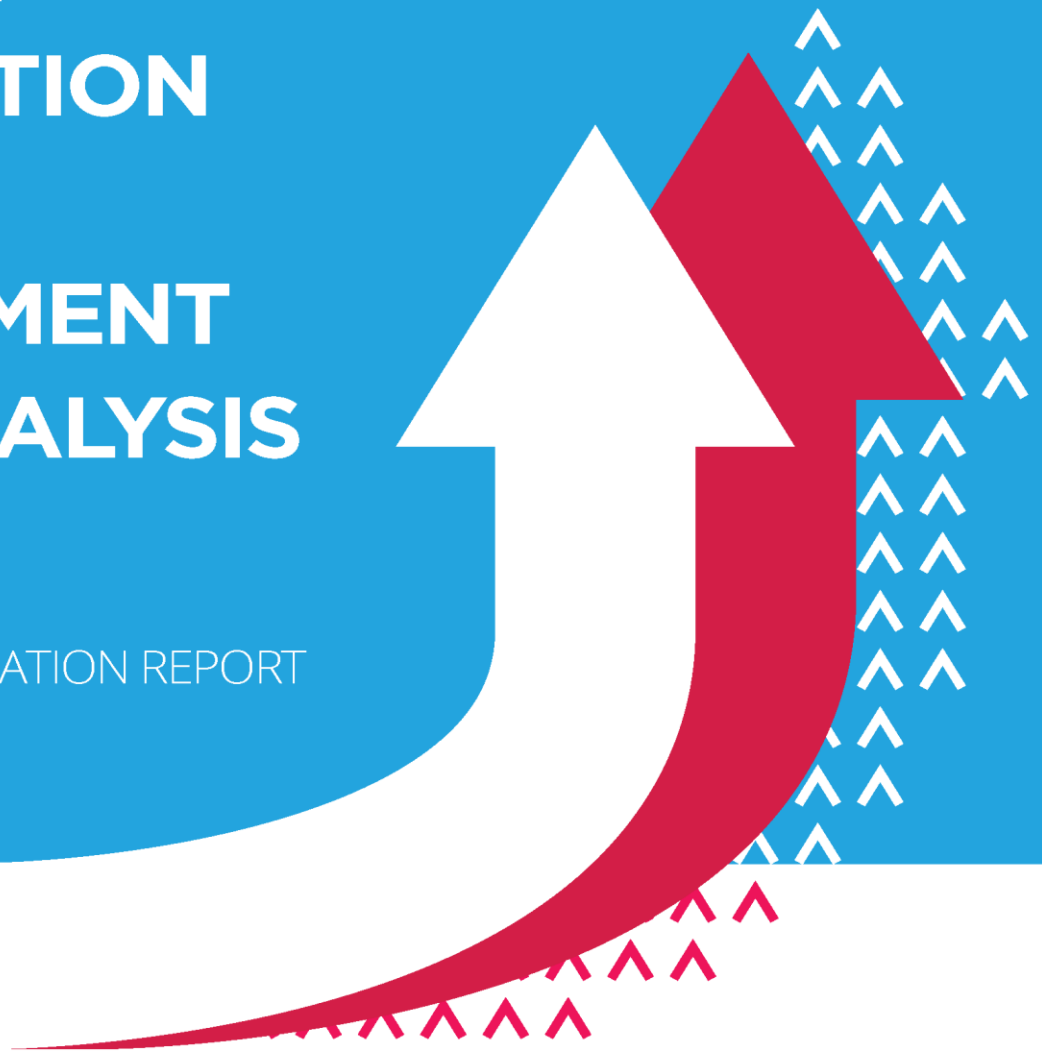


GLOBAL EVALUATION REPORT ASSESSMENT AND ANALYSIS SYSTEM

2018 META EVALUATION REPORT

March 2019



Independent Evaluation Service (IES)
Independent Evaluation and Audit Services (IEAS)



PREPARED FOR UN-WOMEN INDEPENDENT EVALUATION SERVICE (IES)

META EVALUATION MANAGER – Messay Tassew, Evaluation Specialist and Soo Yeon KIM,
Evaluation Analyst

PREPARED BY: Alexandra V. Pittman

Table of contents

Acronyms	4
I. Executive Summary	5
II. Background	7
III. Overview: Geographic Distribution and Characteristics of the Evaluations	7
IV. Findings	8
Overall Rating by Region	9
Overall Ratings per Criteria	10
SECTION 1: OBJECT AND CONTEXT OF THE EVALUATION	12
SECTION 2: PURPOSE, OBJECTIVES, AND SCOPE	14
SECTION 3: METHODOLOGY	17
SECTION 4: FINDINGS	19
SECTION 5: CONCLUSIONS AND LESSONS LEARNED	23
SECTION 6: RECOMMENDATIONS	26
SECTION 7: GENDER AND HUMAN RIGHTS	28
SECTION 8: REPORT PRESENTATION	32
V. Good Practices	34
VI. Conclusions	37
VII. Recommendations to enhance UN Women evaluations in the future	38
ANNEX 1: Contextual Presentation of the Evaluations	41
Annex 2 - List of evaluations for the 2018 GERAAS Review	47

Acronyms

AP	Asia and the Pacific
AS	Arab States
ECA	Europe and Central Asia
EQA	Evaluation Quality Assessment
ESA	Eastern and Southern Africa
GATE	Global Accountability and Tracking of Evaluation Use
GEEW	Gender Equality and the Empowerment of Women
GERAAS	Global Evaluation Reports Assessment and Analysis System
GEWE	Gender Equality and Women's Empowerment (GEWE)
HQ	Corporate
IES	UN Women's Independent Evaluation Service
AC	Americas and the Caribbean
UNEG	UN Evaluation Group
UN-SWAP	UN System Wide Action Plan
VAWG	Prevent violence against women and girls
WCA	Western and Central Africa

I. Executive Summary

In 2013, UN Women's Independent Evaluation Service (IES) developed the Global Evaluation Reports Assessment and Analysis System (GERAAS) as a quality control and learning mechanism to ensure that corporate and decentralized evaluations are, at minimum, compliant with the UN Evaluation Group (UNEG) evaluation quality standards and the UN System Wide Action Plan (UN-SWAP) Evaluation Performance Indicator. In 2018, the GERAAS guidance and the Evaluation Quality Assessment (EQA) matrix have been revised to further enhance the quality and credibility of evaluations. The GERAAS quality review is comprised of eight criteria used to assess each evaluation report: *Section 1: Object and Context of Evaluation; Section 2: Purpose, Objective and Scope; Section 3: Methodology; Section 4: Findings; Section 5: Conclusions and Lessons Learned; Section 6: Recommendations; Section 7: Gender and Human Rights; and, Section 8: Report Presentation.* Reports are independently rated based on the 8 criteria and sub-criteria.

Findings

Nearly 40 percent of the evaluation reports received an *overall* rating of "Very Good" (n=12). Nearly half (n=17, 51%) were rated "Good." There were only four evaluations (12%) which were rated at the lower end of the spectrum - 2 reports were rated "Fair" and two more were rated "Unsatisfactory." The data therefore shows that a majority of the evaluations performed well based on the GERAAS criteria.

Reports that had an overall rating of "Very Good" had the following characteristics:

- Often "Fully" met all sub-criteria under the *Methodology* section, noting gender-responsive methods that were rooted in feminist approaches and provided extensive detail of the human subject research ethics in place, e.g., detailing approaches used to identify participants, and to ensure confidentiality and anonymity.
- Integration of GEWE into the design and presentation of data, resulting in findings, conclusions, and recommendations that were grounded in gender-enriched analysis.
- Conclusions and recommendations that were/are logically linked back to the findings from which they were/are sourced.
- Causal/contributive factors were clearly noted, identified and acknowledged in those frameworks guiding the evaluations.
- Presented content that was clearly written and understandable, and in complete form; no sections of the report were missing and all auxiliary content mentioned under the *Report Presentation* criterion was available for review.

Reports that received an overall rating of "Good" had similar characteristics as those rated "Very Good" but often presented findings without a gender analysis or with one that was limited in scope, or that selectively or haphazardly linked findings to conclusions. Reports with an overall rating of "Good," were also found more often than not to have developed recommendations rooted in a select number of data sources, or provided little to no information outlining the stakeholder engagement process.

The following conclusions were highlighted in the report.

1. More highly rated reports typically articulated a strong gender analysis framework, and carried this lens throughout the evaluation design, implementation, analysis of results, findings and recommendations.
2. The evaluators were consistent in reporting the value of CSO and stakeholder consultations and participation in shaping and achieving key outcomes. However, despite the stated efforts to reach more CSOs and vulnerable communities, many evaluation reports tended to engage with mostly the same type of stakeholder groups, often at the UN or partner level, and had limited engagement with more diverse groups, beneficiaries or other vulnerable populations, which would have deepened insights and strengthened analyses.
3. Most evaluation reports were consistent in highlighting mixed methods to improve the capturing of diverse program and project outcomes.
4. The results of the meta-evaluation revealed declining GERAAS scores of the evaluations in the findings, conclusions and lessons learned, and recommendations sections.
5. Most reports provided limited information on how evaluation implementers engaged with UNEG ethical principles.
6. Some reports noted that limited access to baseline information contributed to low quality reporting.

The following recommendations were made to enhance UN Women evaluations in the future.

1. Underscore the importance of including a detailed summary of stakeholder engagement as part of each evaluation.
2. Incorporate vulnerability assessments into evaluations, or analytic approaches that allow for deeper examination of intersectionality.
3. Use diverse methods to enhance data collection and evaluation practices that allow for more intersectional analyses.
4. Provide more guidance to future evaluators on the presentation of findings, conclusions, lessons learned, and recommendations.
5. Integrate gender considerations and apply a gender lens systematically to improve evaluations.
6. Outline for future evaluators the human research ethics compliance information necessary for sufficient, substantive presentation.
7. Ensure greater access to information to support and strengthen the M&E of development interventions.

II. Background

In 2013, UN Women's Independent Evaluation Service (IES) developed the Global Evaluation Reports Assessment and Analysis System (GERAAS) as a quality control and learning mechanism to ensure that corporate and decentralized evaluations are, at minimum, compliant with the UN Evaluation Group (UNEG) and the UN System Wide Action Plan (UN-SWAP). In 2018, the GERAAS guidance and the Evaluation Quality Assessment (EQA) matrix have been revised to further enhance the quality and credibility of evaluations cognizant of the institutional maturity of UN Women.

To this end, UN Women conducts a quality review by hiring external assessors annually to identify the quality of evaluations on UN Women projects and initiatives. The rating and review feedback are posted in the UN Women publicly accessible database called the Global Accountability and Tracking of Evaluation Use (GATE). This review process is part of a continuous effort to share lessons learned and opportunities for evaluation improvement within UN Women.

The GERAAS quality review is comprised of eight criteria used to assess each evaluation report: *Section 1: Object and Context of Evaluation; Section 2: Purpose, Objective and Scope; Section 3: Methodology; Section 4: Findings; Section 5: Conclusions and Lessons Learned; Section 6: Recommendations; Section 7: Gender and Human Rights; and, Section 8: Report Presentation.* The revised GERAAS rating matrix gave more weight to criteria focused on evaluation methodology, findings, conclusions, and recommendations. This report presents the key findings -- summarizing key trends, assessing report quality, and documenting good practices for evaluations in the future. Additionally, the report also includes recommendations for systems improvement.

III. Overview: Geographic Distribution and Characteristics of the Evaluations

Geographic Distribution

The 33 evaluations were from programmes and interventions that were implemented across the six UN Women operational regions: Arab States (AS), Asia and the Pacific (AP), Eastern and Southern Africa (ESA), Europe and Central Asia (ECA), Americas and the Caribbean (AC), Western and Central Africa (WCA), and Corporate (Independent Evaluation Service and HQ Division). Over half of the evaluations were from the ESA region (n=8) and the AC region (n=7); followed by AS (n=5), AP and WCA (n=4), ECA (n=3), and the fewest number of evaluations were from the Corporate and HQ divisions (n=2).

Evaluation Types (Location, Types, and Themes): Most reports analyzed were country-level evaluations (n=23). There were five regional evaluations, two global evaluations and three multi-country evaluations assessed.

The most frequent type of evaluations analyzed were project evaluations (n=14), followed by programme evaluations (n=7). Five country portfolio evaluations were assessed followed by five

regional, and corporate and global evaluations each. Evaluations implemented at HQ and by the Independent Evaluation Service (IES) covers multiple levels and countries.

In assessing all the evaluations, the top three most frequently noted themes in the evaluations were women’s economic empowerment (n=16); women’s leadership and participation (n=15); and preventing violence against women (VAWG) and expanding access in services (n=12).

The evaluations also focused on women’s leadership in peace, security and humanitarian response (n=11); global norms, policies and standards on gender equality and women’s empowerment (GEWE) (n=9); and gender responsive plans and budgets (n=6) to a lesser extent.

Every evaluation included in the meta-evaluation addressed at least 1 strategic plan/thematic area out of the five outcomes reflected in the UN Women Strategic Plan.¹ Nearly 40% of evaluations (n=13) addressed two or more strategic plan/thematic areas whereas only 15% of evaluations (n=5) addressed all six thematic areas.

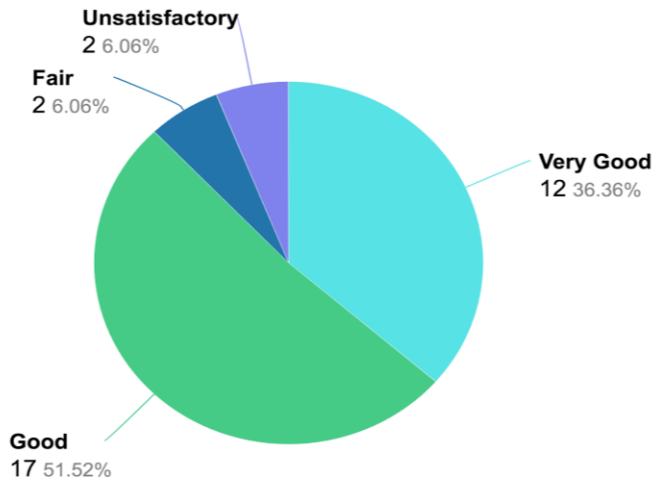
There were 83 evaluators - 61 women and 22 men - who carried out the evaluation work in 2018 in this sample. A more detailed examination of the evaluation types, geographic distribution, and strategic plan/thematic area of focus among those evaluations reviewed, with associated charts, is presented in the Annex below.

IV. Findings

Nearly 40 percent of the evaluation reports received an *overall* rating of “Very Good” (n=12). Nearly half (n=17, 51%) were rated “Good.” There were only four evaluations (12%) which were rated at the lower end of the spectrum - 2 reports were rated “Fair” and two more were rated “Unsatisfactory.” The data therefore shows that a majority of the evaluations performed well based on the GERAAS criteria.

¹ The six (6) strategic plan/thematic areas that are reflected in the GERAAS: Women’s leadership and participation; Women’s access to economic empowerment and opportunities; Prevent violence against women and girls (VAWG) and expand access to services; Women’s leadership in peace, security and humanitarian response; Gender response plans and budgets; Global norms, policies and standards on gender equality and women’s empowerment (GEWE).

Overall GERAAS Rating



Reports that had an overall rating of “Very Good” had the following characteristics:

- Often “Fully” met all sub-criteria under the *Methodology* section, noting gender-responsive methods that were rooted in feminist approaches and provided extensive detail of the human subject research ethics in place, e.g., detailing approaches used to identify participants, and to ensure confidentiality and anonymity.
- Integration of GEWE into the design and presentation of data, resulting in findings, conclusions, and recommendations that were grounded in gender-enriched analysis.
- Conclusions and recommendations that were logically linked back to the findings from which they were sourced.
- Causal/contributive factors were clearly noted, identified and acknowledged in the frameworks guiding the evaluations.
- Presented content that was clearly written and understandable, and in complete form; no sections of the report were missing and all auxiliary content mentioned under the *Report Presentation* criterion was available for review.

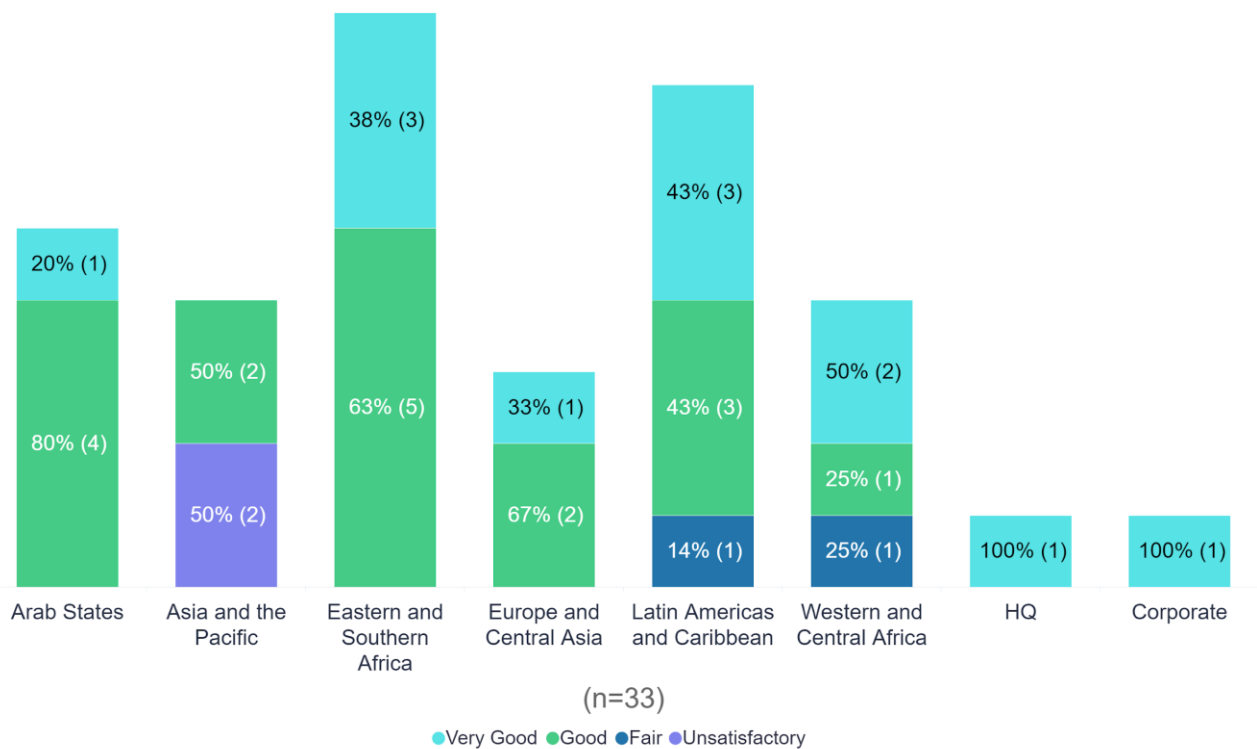
Reports that received an overall rating of “Good” had similar characteristics as those rated “Very Good” but often presented findings without a gender analysis or with one that was limited in scope, or that selectively or haphazardly linked findings to conclusions. Reports with an overall rating of “Good,” were also found more often than not to have developed recommendations rooted in a select number of data sources, or provided little to no information outlining the stakeholder engagement process.

Overall Rating by Region

- All evaluations from IES and HQ (one each) had an overall rating of “Very Good”

- Evaluations from the *Arab States* and *Europe and Central Asia* as well as *Eastern and Southern Africa*, received high overall ratings on the GERAAS as these regions only had “Very Good” and “Good” evaluations. However, a higher proportion of their evaluations are rated as “Good.”
- The regions of *Americas and the Caribbean* and *Western and Central Africa* had more diverse ratings, as both regions garnered “Very Good,” “Good” and “Fair” ratings. But the Americas and the Caribbean is among the regions with higher proportion of evaluations with very good rating.
- The *Asia and the Pacific* region appeared to be the region without any evaluations rated as “Very Good.” Half of the 4 reports from the regions were rated “Good,” and the other half are rated “Unsatisfactory.”

Overall Rating by Region



Overall Ratings per Criteria

The chart below shows a comparison of ratings across the eight GERAAS criteria. The data show that the evaluations had strong ratings in 5 out of 8 criteria, namely: *Object and Context; Purpose, Objectives, and Scope; Methodology; Gender & Human Rights – based on the UN SWAP Evaluation*

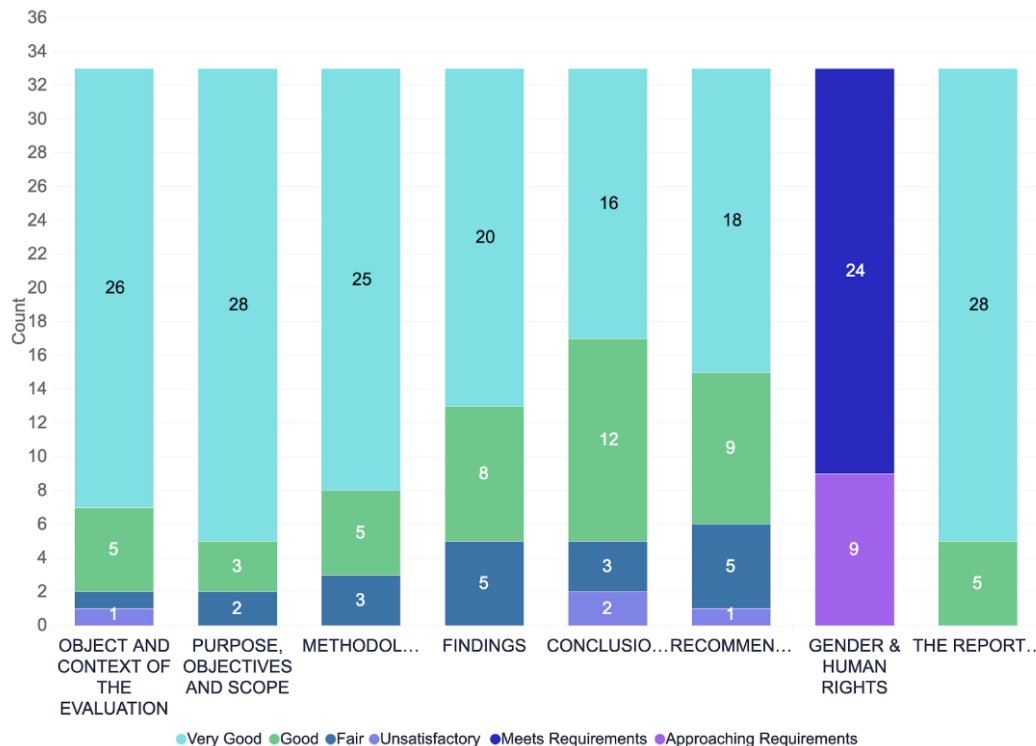
Performance Indicator; and Report Presentation. Under these criteria, 73%-85% of the evaluations garnered a rating of “Very Good.”²

The strongest areas of the evaluations overall included:

- *The Purpose, Objectives, and Scope*, where 85% (n=28) of the evaluations were rated as “Very Good” - the highest proportion of “Very Good” amongst all the 8 criteria.
- The *Report Presentation* section was also strong with 85% (n=28) of the evaluations scored as “Very Good” and with no “Fair” or “Unsatisfactory” ratings.

The criteria with the relatively low ratings were the *Findings, Recommendations, and Conclusions & Lessons Learned*, respectively, as noted in the chart below. This is because there was a decrease in the proportion of evaluations rated as “Very Good” while there were more evaluations rated as “Good.” There was also an increase in the proportion of evaluations that were “Fair” and “Unsatisfactory.” As mentioned above, 73%-85% of the evaluations under the first three criteria were rated “Very Good” and it dipped to around 48-57% for the *Findings, Conclusions, Recommendations* criteria.

Overall Ratings per Criteria



² 73% (n=23) were rated as “Meets Requirements” under the Gender & Human Rights parameters.

The subsequent sections detail trends in each of the eight GERAAS criteria.

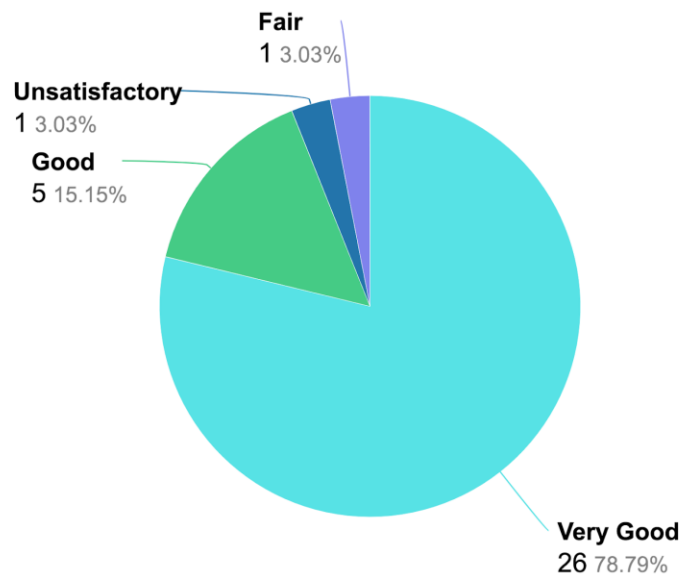
SECTION 1: OBJECT AND CONTEXT OF THE EVALUATION

Does the report present a clear and full description of the 'object' of the evaluation?

The first criterion of the GERAAS focused on the following questions: *Does the evaluation clearly articulate what it aims to assess in the introduction (section)?; Are social, economic, cultural considerations noted in the evaluation?; Are the key stakeholders and users clearly identified?; and Is the implementation status of the object known?*

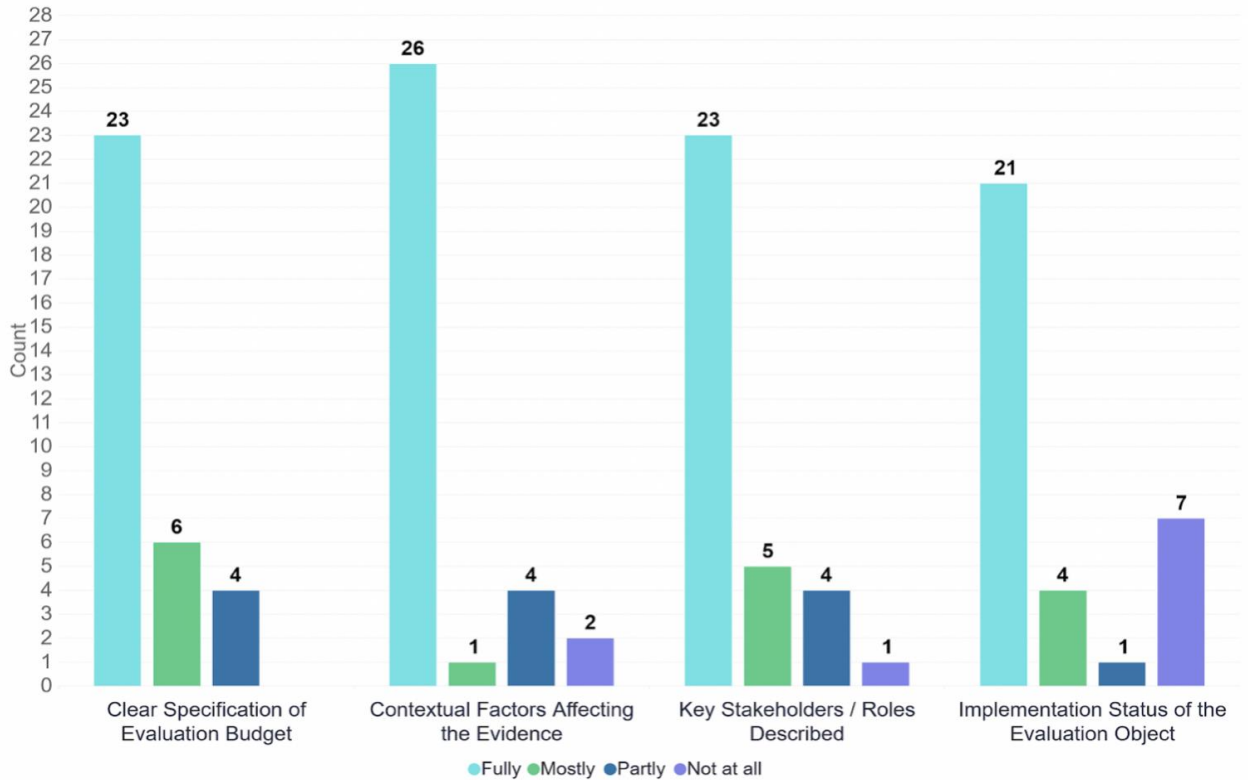
The chart below illustrates that 79% of the evaluations received an overall rating of “Very Good” (n=26).

Overall Rating: Object and Context of the Evaluation



For each of the four sub-criteria for Section 1, between 70%-80% of the reports (n ranges from 21 to 26) received ratings of “Fully” addressing the issues of focus. The sub-criterion of *Contextual Factors Affecting the Evidence* had the highest proportion of reports receiving a “Fully” rating (n=26) when compared to the other 3 sub-criteria. See chart below to see ratings across all four sub-criteria.

Section 1: Object and Context of the Evaluation



In assessing the strengths and weaknesses of evaluation reports, the following key factors were related to stronger ratings. Reports rated “Very Good” or “Good” provided a deeper overview of the context with a clear theory of change and discussed how a program or project would improve gender dynamics in a given cultural context in a country or region. Stronger evaluations outlined the social, economic, and political factors that contributed to gender inequality and noted the role of stakeholders in addressing challenges. Strong reports detailed how the program or project design integrated participation and information from beneficiaries and forecasted how advancing the program or project would improve outcomes in alignment with UN Women priorities.

The “*Final Project Evaluation: Women’s Citizenship for Peace, Justice and Development by Colombia CO*” received the highest ratings “Very Good” on all sub-criteria. The report provided an in-depth overview of the context, including information about the legal framework around peacebuilding in Colombia, gender related data about the war and women’s participation in politics and how the project was developed to address these concerns. A theory of change was also detailed and included information on diverse pathways for change. The report also included information related to the implementation status and shifts that occurred due to budget constraints which was helpful in positioning the effectiveness and efficiency of the initiative. The “*Regional Evaluation on Normative*

Frameworks by the Americas and Caribbean Regional Office” also had high scores on all sub-criteria. A particularly strong part of this evaluation was the context analysis and theoretical framework section, which focused on the analysis of the LAC context and assessed the context as ranging from the transition from favorable governments to a regressive environment characterized by conservative tendencies and then analyzed normative policies, plans and programs regarding gender equality and empowerment of women within that frame.

In contrast, reports with lower scores or weaker evaluations did not provide a clear theory of change and lacked clarity on project or program implementation. These reports also did not clearly highlight the role of stakeholders and key partners or how resources (budget) were used throughout the lifecycle of the project or program. Weaker reports also did not clearly delineate and explore the cultural conditions at play or apply a gender lens throughout the report to assess the project or program.

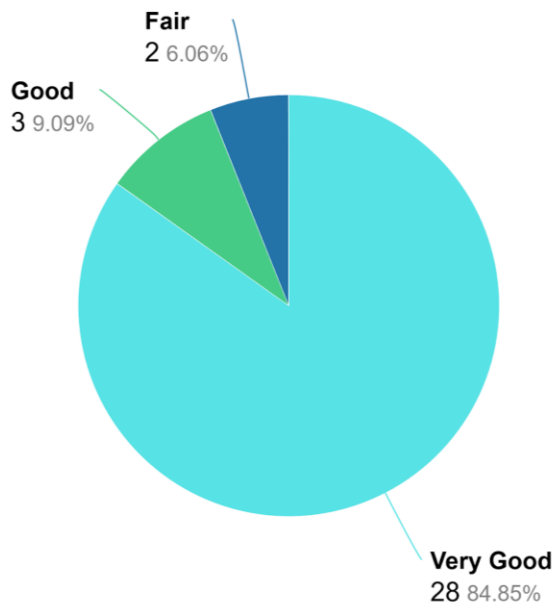
SECTION 2: PURPOSE, OBJECTIVES, AND SCOPE

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

The second criterion of the GERAAS directly refers to the purpose, objectives, and scope of the evaluation. The following questions were posed to rate this criterion: *Is the goal of the evaluation clearly stated?*; and *Is the intended rationale clearly articulated (e.g., making forward-looking recommendations)?* An explanation of the evaluation focus was also needed to provide additional context as well as reference to the geographic scope and justification for the evaluation.

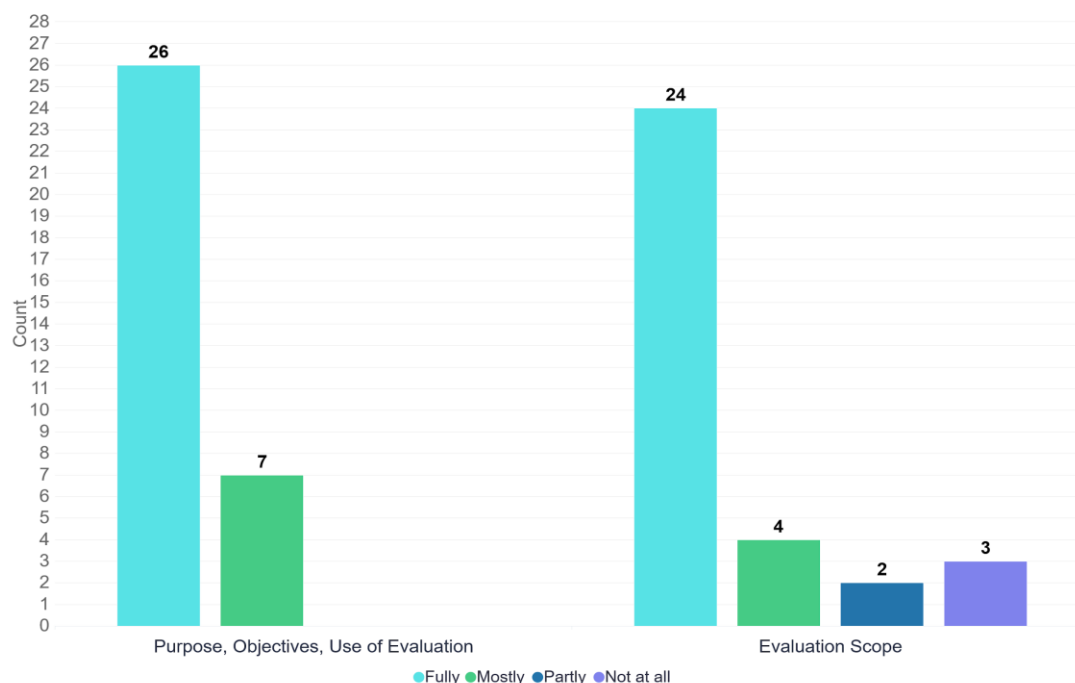
Illustrated in the visual below, 85% of the evaluations (n=28) received an overall rating of “Very Good”. Twelve percent of the evaluations (n=4) were rated either “Good” (n=2) or “Fair” (n=2) for this criterion.

Overall Rating: Purpose, Objectives, and Scope of the Evaluation



The chart presents the sub-criteria, which focused on the evaluation scope and purpose of the evaluation. Overall, both of the criteria around *Evaluation Scope* were “Fully” or “Mostly” met in the evaluations with very few receiving “Partly” or “Not at all” ratings (only 5 reports combined).

Section 2: Purpose, Objectives and Scope of the Evaluation



The purpose, objectives, and use section of reports varied by the degree of detail provided, specifically how projects or programs would contribute to improving outcomes. Stronger reports rated “Very Good” or “Good” clearly articulated the key research objectives and questions, linked the contextual realities of a given intervention and had a clear statement of the key intentions that aligned with UN Women priorities. Strong reports also explained how stakeholder engagement aimed to strengthen implementation and develop lesson learned to advance outcomes in the future. Reports rated “Very Good” or “Good” provided meaningful discussion on the importance of improving accountability mechanisms and governance in collaboration with key stakeholder groups. Most reports that were rated “Very Good” or “Good” in sections 1 and 2 also scored “Very Good” or “Good” in section 3. The “Country Portfolio Evaluations” and the “Building capacity to Prevent Violence Against Women (BCPVAW), Bangladesh ” report are examples of consistent evaluations that maintained strong marks in sections 1 - 3 and scored overall marks of “Very Good” or “Good”, which had strong articulation of objectives, purpose and rationale of the evaluation.

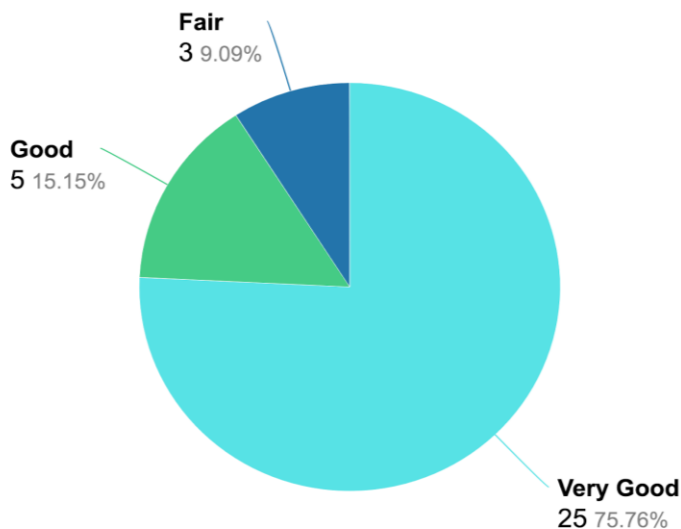
Conversely, weaker reports did not provide context on the project or program scope and lacked the same degree of detail as in other related sections of the evaluation. Weaker reports were too often broad and did not clearly explain the intentions of the project with clear objectives. Reports that received lower scores provide more explanation in other sections of the evaluation. For example, one evaluation rated “Fair” in this section had a “Very Good” in the conclusion and lessons learned section of the report. However, this was more of an exception, most weaker reports in this category, were not consistent in the depth of information provided throughout, which contributed to the lower score.

SECTION 3: METHODOLOGY

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

The third criterion examines the methodology and focused on questions related to *whether the evaluation criteria, the data collection and analysis approaches were effective in providing insight*. Additional questions included: *Were stakeholders consulted during the evaluation exercise?* and *To what extent was the interaction/engagement with the evaluation team, and ethical standards used, presented in the report?* Overall, 76% of the evaluations (n=25) received an overall rating of “Very Good” for the methodology criterion as noted in the chart below.

Overall Rating: Methodology

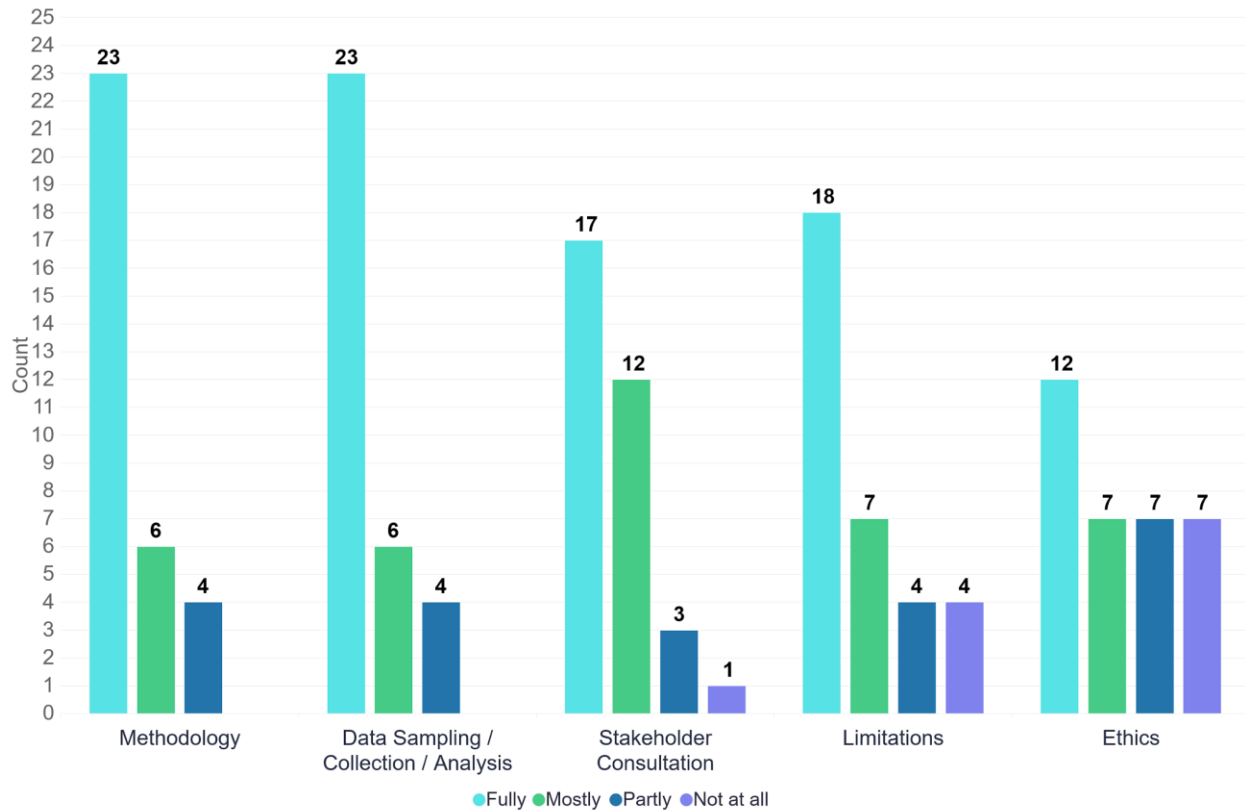


As shown in the chart below, at least two-thirds of the Section 3 sub-criteria were rated “Fully” for the *Data Sampling/Collection/Analysis* (n=23) and *Methodology* (n=23) sub-criteria.

The *Ethics* sub-criterion had the fewest number of evaluation reports which were rated “Fully” (n=12) and “Mostly” (n=6).³

³ Limited information or no mention around maintaining confidentiality, doing no harm or securing informed consent were among the reasons why a report received lower scores of “Partly” and “Not at all” on the Ethics measure. In some evaluations, the mere mention of adhering to UN Evaluation Group guidelines or vague reference to ethical practices (that no one will be harmed, data will be anonymized) might have been considered sufficient to receive a rating. However, those reports that received either a “Fully” or “Mostly” rating for the *Ethics* sub-criterion provided more detailed information about the informed consent process in the evaluation report. For example, one case made reference to providing the sample information sheet to educate prospective participants about the evaluation purpose and the right to withdraw. Reports that received fully/mostly for the *Ethics* sub-criterion also appeared to have offered redress -or stated that they did so- as part of the data collection process.

Section 3: Methodology



Description of *Limitations* was another sub-criterion that was examined under the methodology section of this review. A number of reports that received a rating of “Partly” or “Not at all” for this sub-criterion did not mention limitations at all in the report or provided limited information on how these were addressed during the course of the evaluation.

The *Stakeholder Consultation* process through which individuals were engaged or consulted in the evaluation was rated as either *Fully* or *Mostly* was discussed in almost 88 percent of the evaluation reports (n=29).

Reports classified as strong received “Very Good” or “Good” ratings and included a comprehensive discussion of the methodology, the rationale, and process. Stronger reports detailed the limitations of the project or program and explained at length the mixed methodology, including Key Informant Interview, surveys, multi-sectoral consultations, and quantitative data collection. In addition, these reports included a detailed explanation of how evaluators maintained quality control and articulated the ethical principles and considerations that guided project and program implementation. Furthermore, strong reports outlined engagement with experts to ensure quality

and that significant data and information were collected in alignment with the key objectives. Strong reports also clearly explained key barriers or challenges in collecting some information, which provided greater transparency.

Examples of strong reports with excellent methodology was the evaluation from Kyrgyzstan on the project *“Joint Programme on Accelerating Progress towards the Economic Empowerment of Rural Women (JP RWEE), Kyrgyzstan”* with a perfect score. Diverse mixed methodologies, including the “Change Map” and Women Empowerment in Agriculture Index (WEAI) were used in this project helped to lift up diverse and marginalized perspectives and had a dedicated focus on transformative changes--those changes that shifted the underlying institutional, community, and relational structures of inequalities and discriminations. A particularly strong focus of the corporate evaluation on *“UN Women’s Contribution To Women’s Political Participation And Leadership”* evaluation was the dedicated use of gender analysis and use of the Inclusive Systemic Evaluation for Gender equality, Environments and Marginalized voices (ISE4GEMs) methodology. This methodology focused on rights based and gender-responsive approaches to data collection and analysis.

In contrast, weaker reports lacked clear explanations of the process employed and lacked insight on the limitation of the reports. Reports rated “Fair” did not clearly explain the rationale for the methodology and noted limitations without providing a proxy method or other considerations explored. The level of detail was partly reported but without robust explanations.

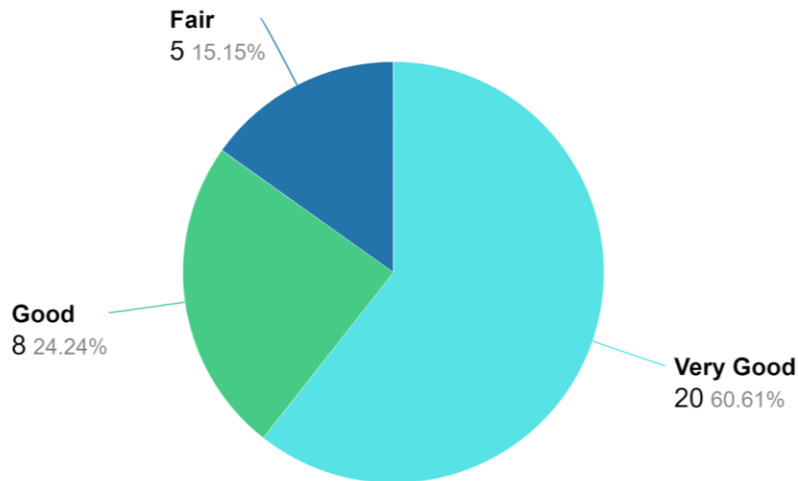
SECTION 4: FINDINGS

Are the findings clearly presented, relevant, and based on evidence?

Section 4 of the GERAAS criteria focused on the presentation of findings. Assessment was focused on the following issues: if the findings were presented with *high evidence data*; if the *findings were supported with evidence*; if the findings *informed causal factors*; and if the *findings were clear, logical, and coherent*.

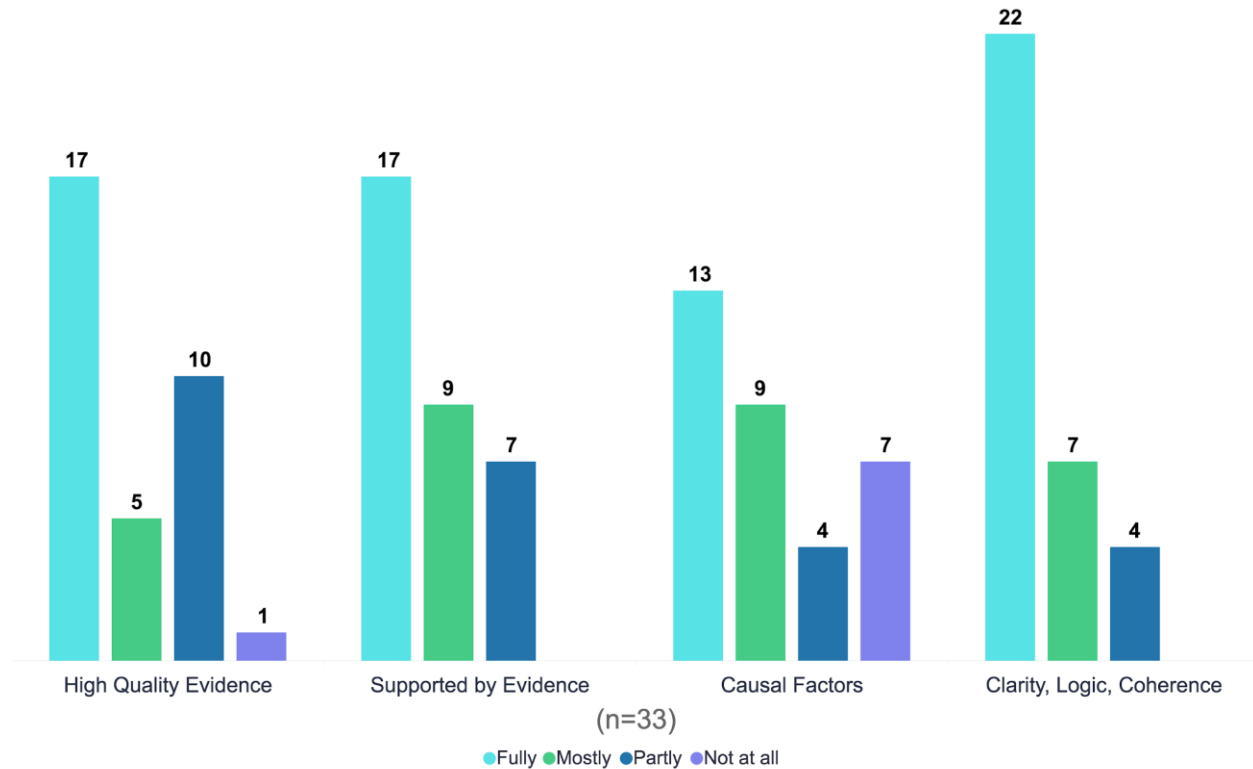
The chart below illustrates that more than half (61% or n=20) of all the evaluations were rated as “Very Good.” There were 8 evaluations (24%) rated as “Good” and 5 evaluations (15%) as “Fair.” No report was rated as “Unsatisfactory.”

Overall Rating: Findings



While the majority of the evaluations are still rated “Very Good”, as noted earlier, the proportion of evaluations rated as “Very Good” declined substantially in this parameter. This is due to the rise of evaluations graded “Partly” and “Not at all” especially on the sub-parameters on *evidence* and *causal factor* as seen in the chart below. As the chart illustrates below, the *High-Quality Evidence* sub-criterion, had more “Partly” (n=10) scores than “Mostly” (n=5) while the sub-criterion *Supported by Evidence* had 7 evaluations with “Partly” scores. Evaluations that had “Mostly” grades are usually also weak in the same sub-parameters. Overall, the low grades in these aspects pulled the scores down.

Section 4: Findings



Conversely, evaluations that garnered a rating of “Very Good” in the Findings parameter would have some or most of the following attributes:

- Adequately able to address in their findings the evaluation criteria and questions. This also contributed to a clear, logical, and coherent evaluation.
- Objective, clear, and comprehensive presentation of findings through the use of tables, graphs, quotes, and testimonials.
- Clear link of data collection strategies employed and data sources (i.e., stakeholders interviewed) when presenting data. This also showed that triangulation was indeed done.
- Integration of human rights and gender lens in the analysis of findings.
- Presentation of impact-level, gender-transformative data.
- Deliberate discussion of factors that led to achievement or non-achievement of project goals. Some evaluations even had a separate section for this.
- Accounting both the positive gains and weaknesses of a development intervention, which made the findings robust and more comprehensive.

An example of a strong report is the *Independent Evaluation of UN Women's Fund for Gender Equality (2009-2017)* which garnered a perfect score across all sub-criteria for the Findings parameter. This particular evaluation assessed data from 80 countries yet it was able to present

information clearly and with ample evidence through the consistent use of charts, tables, and quotes from data sources. The findings section was divided into 3 parts: 1. Did the Fund do things right?; 2. Did the Fund do the right things?; and 3. Evaluation Case Studies and in each part, specific evaluation criteria aligned with the part was discussed (i.e., effectiveness and efficiency criteria was discussed under the *Did the Fund do the right things?* section). This allowed for a very clear, systematic, and contextualized discussion of the findings. This evaluation also provided a comprehensive picture as it noted both the strengths and weaknesses of the intervention being assessed. Finally, the evaluation did not fail to explain the factors behind the achievement or non-achievement of the project goals.

Another good example is the evaluation *Joint Programme On: Accelerating Progress Towards The Economic Empowerment Of Rural Women (Jp-Rwee) in Kyrgyzstan*. This was especially strong as it was able to present gender-transformative, impact-level data of the project -- illustrating clearly how the project made progress at the grassroots level. It was also consistent in explaining the reasons which contributed to the realization of the project goals and also captured the negative effects of the project and illustrating the reasons why.

The *evaluation of the Safe Cities Campaign #Noesdehombres, Mexico* and *Mid Term Evaluation of the WEE BE Funded Project in Gaza Province, Mozambique* did a good job in clearly linking data being presented and the data collection strategies undertaken. Data sources (i.e., stakeholder or data gathering conducted such as interview etc.) are cited as necessary in the presentation of data. This in turn provided clarity on how the data collection strategies were reflected in the findings and at the same time, provided better contextualization of the data being discussed.

Evaluations which showcased triangulation were *Evaluación Del Proyecto "Mujeres Liderando el Desarrollo Inclusivo Sostenible de la Provincia de Loja -Ecuador*, *Regional Evaluation on Normative Framework (Americas and the Caribbean Regional Office)*, and *Evaluation of Women, Peace and Security in the Arab States*. These evaluations consistently cited diverse data sources and research connected to different findings. As such, it was clear how findings were validated.

Meanwhile, evaluations that did not achieve higher ratings in this section primarily did not include sufficient and detailed tables, charts, quotes, and testimonials to support their findings. These evaluations in turn appeared to be subjective and ambiguous without data to back up findings.

Meanwhile, a significant number of evaluations were also graded low in highlighting *causal factors*. These evaluations were unclear or had minimal discussion of factors of success/failure or did not tackle causal or contributory factors in their findings. In some assessments, there was difficulty in identifying causality as these evaluations did not have explicit discussion of theory of change or results framework. On the other hand, assessments often struggled to establish causality due to the lack of baseline data.

Another trend found in lower scores in this section was the weak logical presentation of data. There were a few evaluations that did not present findings based on the evaluation criteria or questions.

The findings of one evaluation appeared to be more like recommendations or conclusions rather than presenting the analysis of data gathered. There was an evaluation where findings were presented per project component; while there was another where findings were presented by results/success. Thus, it was difficult to determine if the evaluation criteria and questions were addressed.

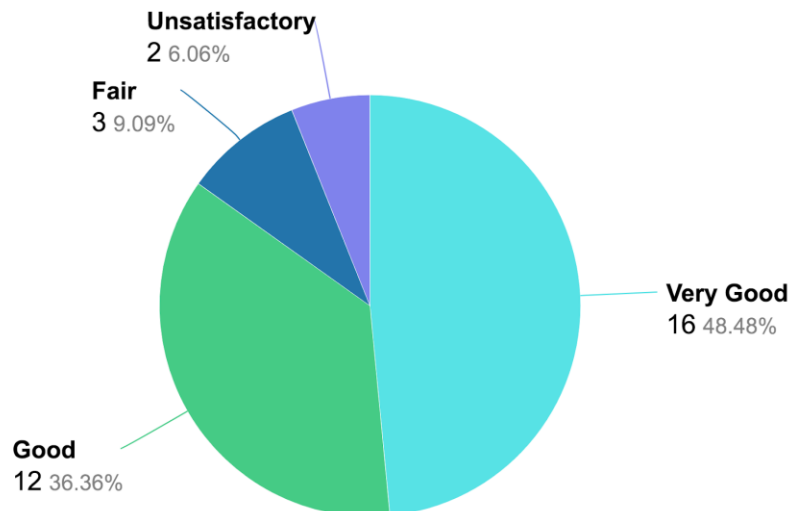
SECTION 5: CONCLUSIONS AND LESSONS LEARNED

Are the conclusions clearly presented based on findings and substantiated by evidence?

Section 5 assessed how conclusions and lessons learned were discussed in the evaluation. Assessment of conclusions and lessons learned were rated based on the following sub-criteria: *substantiated by evidence and connected to the findings; provided further insights and went beyond findings; presented strengths and weaknesses; and lessons learned are based on findings and accounts for applicability to sectors/contexts.*

In the presentation of conclusions and lessons learned, almost half (49%; n=16) received a rating of “Very Good.” There were 12 reports that were rated “Good.” On the other hand, 3 reports were rated “Fair” and 2 were “Unsatisfactory.”

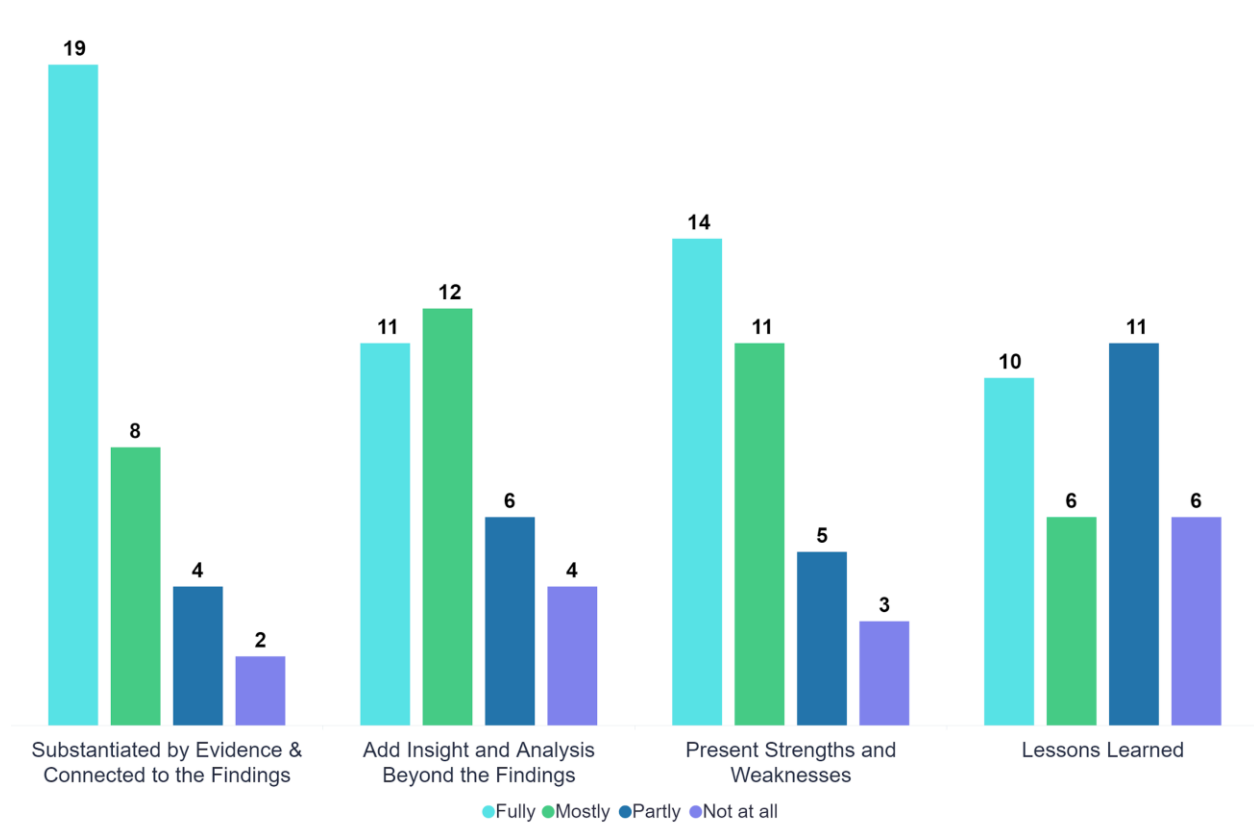
Overall Rating: Conclusions & Lessons Learned



Like the analysis on the *Findings* criteria, the proportion of evaluations rated “Very Good” declined in the *Conclusion and Lessons Learned* criteria. This is driven by the increase in the proportion of “Mostly,” “Partly,” and “Not at All” scores.

As seen in the chart below, evaluations with scores “Mostly” and “Partly” were distributed across all sub-criteria. Evaluations with conclusions that *add insights and went beyond findings* had a higher proportion of “Mostly” (n=12) scores than “Fully” (n=11). Notably, there were more evaluations rated as “Partly” (n=11) than “Fully” (n=10) and “Mostly” (n=6) for the parameter of *Lessons Learned*. Likewise, *Lessons Learned* had the highest proportion of evaluations scored as “Not at all” (n=6).

Section 5: Conclusions & Lessons Learned



The strongest evaluations in the presentation of *Conclusions and Lessons Learned* area had some or most of the following characteristics present:

- Conclusions are aligned to the findings and described a clear link of conclusions to the findings.
- Conclusions added further insights and analysis to the findings.
- Strengths and weaknesses of the development intervention were analyzed.
- Lessons learned were presented, practical and action-oriented.

There were various evaluation reports that clearly cited the specific findings that led to the creation of a conclusion, including: *Evaluation of the Safe Cities Campaign #Noesdehombres (Mexico)*, *Final Evaluation Project: Women's Citizenship for Peace, Justice and Development (Colombia)*, *Mid Term Evaluation of the Programme: Overcoming Gender-Based Violence to Ensure Women's Full Enjoyment of Rights (Colombia)*, *Kenya Country Portfolio Evaluation and Independent Evaluation of UN Women's Fund for Gender Equality (2009-2017)*.

Stronger evaluations synthesized high level findings, added an additional level of analysis, which facilitated deeper insights. Among the evaluations that carried out this good practice were the *Independent Evaluation of UN Women's Fund for Gender Equality (2009-2017)*, *Joint Programme On: Accelerating Progress Towards The Economic Empowerment Of Rural Women (Ethiopia)*, and *Mid Term Evaluation of the WEE BE Funded Project in Gaza Province, Mozambique*.

A handful of evaluations also scored high as they were able to elaborate on the strengths and weaknesses of development interventions in the conclusions. Examples of evaluations that did well in incorporating strengths and weaknesses in the conclusions section, included: *Regional Evaluation on Normative Framework (Americas and the Caribbean)*, *Final Evaluation of the Joint Programme "Advancing and Sustaining Gender Equality Gains in Rwanda, UN Women's Contribution to Women's Political Participation and Leadership (Corporate Evaluation)*, and *UN Women Country Portfolio Evaluation South Sudan*. Specifically, these evaluations elucidated strengths and weaknesses by highlighting the diverse point of view of stakeholders relevant to the intervention being evaluated.

Meanwhile, evaluations that garnered higher scores discussed lessons learned, and ensured they were actionable and practical. This meant that lessons learned were fully contextualized, taking into account the limitations and possibilities of the environment and the stakeholders involved. Examples of evaluations that reflected this good practice were *Independent Evaluation of UN Women's Fund for Gender Equality (2009-2017)*, *Evaluation of the Regional Project on Preventing Exploitation Of Women Migrant Workers (Regional Office for Asia and the Pacific)*, and *UN Women's Contribution to Women's Political Participation and Leadership* to name some. Meanwhile, some evaluations went an extra mile by including an additional section to highlight lessons learned - such as "Good Practices" (Thematic Evaluation Of Women's Political Empowerment) and "Best Practices" (Final Evaluation Sweden Project: Women's Citizenship for Peace, Justice and Development).

One of the main reasons evaluations garnered lower scores was because the *Conclusions* were just a re-articulation of the findings. It offered no additional insights. A few evaluations also had weak conclusions because the findings had insufficient data. Some evaluations had very few conclusions or had concluding statements that were too brief, which made the conclusions unclear. Meanwhile, there were evaluations that had no conclusion section at all. In these cases, the conclusions were incorporated in other parts such as the "Recommendation" section.

Evaluations that were scored weaker in the *Lessons Learned* criteria often did not have a dedicated section for this particular criterion. Some of the evaluations that had missing lessons learned section incorporated learnings into other parts such as the Conclusion or Recommendation section. Other weaker evaluations in this aspect did have the lessons learned section but they were not easy to implement as they lacked grounding in contexts and practical limitations.

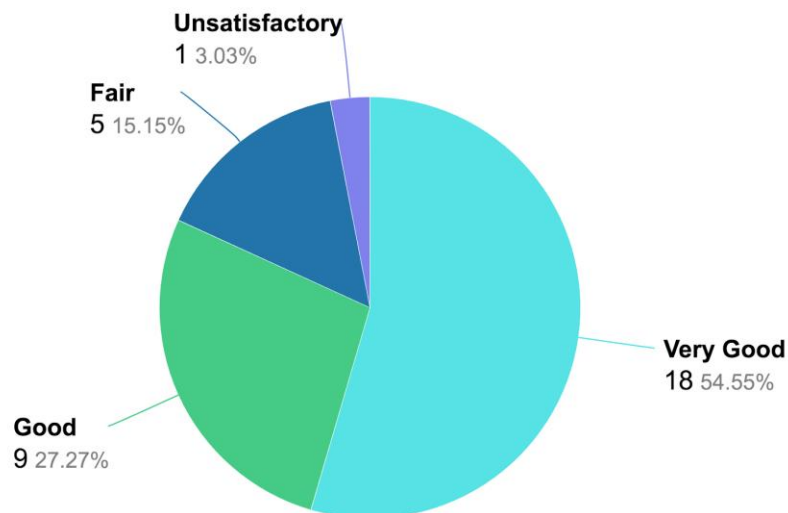
SECTION 6: RECOMMENDATIONS

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

Section 6 assessed how the recommendations were presented in the evaluation report. The ratings were determined by the following sub-criteria: *recommendations are based on conclusions/findings; presence of discussion of the process in arriving at the recommendation (which includes consultation with the stakeholders); clear and realistic recommendations; and prioritization/classification of the recommendations.*

The assessment found that over half (55%) of the reports (n=18) had ratings of “Very Good” in the Recommendation sections. There were 9 reports rated as “Good.” Five evaluations had a weaker Recommendation section, i.e., 5 reports were rated as “Fair”, and one report as “Unsatisfactory.”

Overall Rating: Recommendations

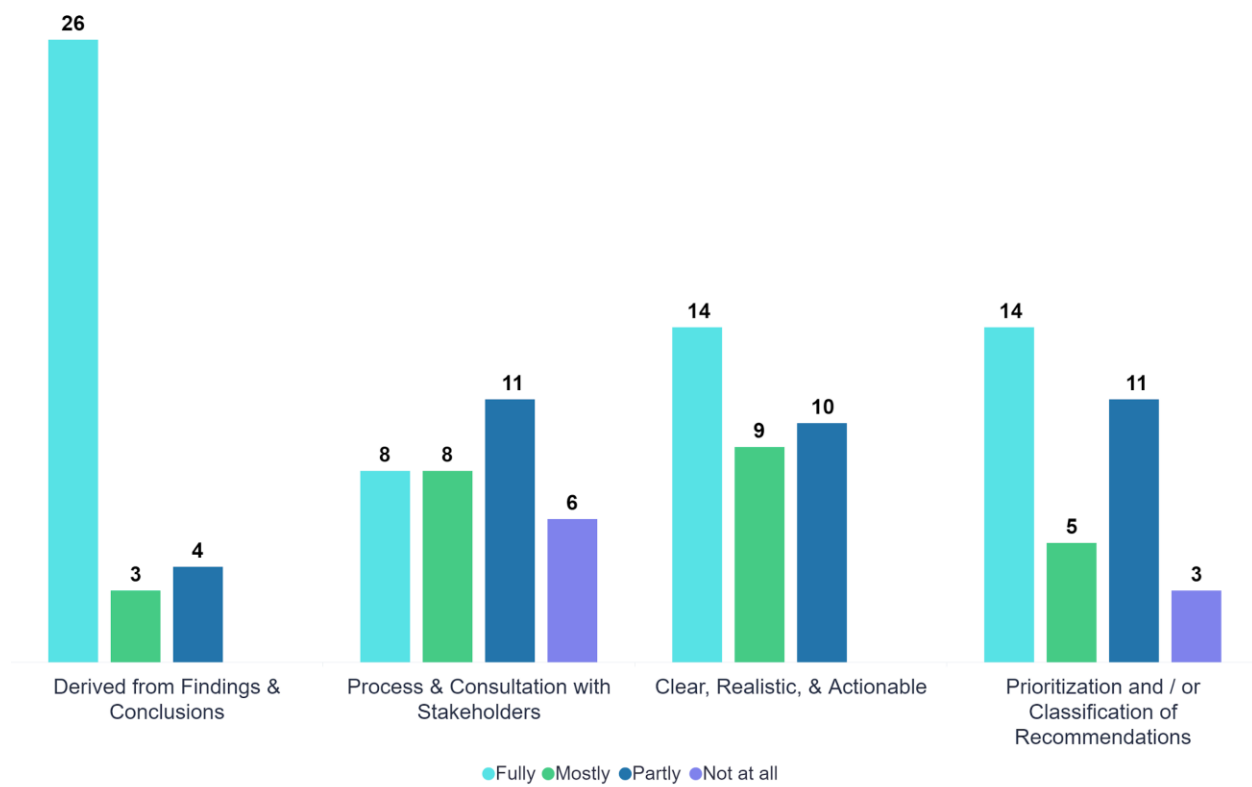


The criteria on *Recommendations* showed that majority of evaluations is still rated as “Very Good.” But similar to *Findings* and *Conclusions and Lessons Learned*, the proportion of evaluations rated as

“Very Good” declined following the increase in the proportion of evaluations scored “Mostly,” “Partly,” and “Not at All,” for the sub-criteria.

As seen in the chart below, “Partly” scores dominated the sub-criteria *Process & Consultation with Stakeholders*. “Partly” scores also comes next to “Fully” in terms of the recommendation being realistic and have proper prioritization and classification.

Section 6: Recommendations



The strongest evaluations under this parameter would feature some or most of the description below:

- Provided a strong link between recommendations and the findings and conclusions.
- Able to describe the process in developing the recommendations.
- Had clear, realistic, and actionable recommendations.
- Recommendations had clear prioritization and classifications.

For the aspect on clear linkage of recommendations to the findings and conclusions, 78% (n=26) of the evaluations demonstrated this. This was the foremost strength of evaluations under this criteria.

The *CVE Project Evaluation, End Evaluation Of “Preventing And Responding To Violence Against Women And Girls In Ethiopia”* and *UN Women's Contribution to Women's Political Participation and Leadership* were some evaluations that were able to provide good details on how they arrived to their recommendations. It was able to explain the processes and the people involved in validating and finalizing the recommendations for the evaluation.

There were a handful of evaluations that rated high in terms of providing *clear, actionable, and realistic recommendations*. The strongest ones were realistic enough to identify possible limitations and constraints in undertaking the recommendations, clear enough to identify the diverse stakeholders delegated for the recommendations, and detailed enough on the action steps to be taken. Meanwhile, there were also evaluations that demonstrated good practices in terms of ensuring recommendations have prioritization levels and classification. These evaluations were able to identify the specific stakeholders to undertake the next steps, and the level of urgency, feasibility, difficulty, and impact of each recommendation.

The strongest evaluations demonstrating such good practices included *Evaluation of the Safe Cities Campaign #Noesdehombres (Mexico)*, *Promote and strengthen the institutionalization of gender equality in the public policies (Morocco)*, *Burundi Country Portfolio Evaluation*, and *UN Women Rwanda Country Portfolio Evaluation 2014-2018*.

Meanwhile, evaluations that garnered low scores in this criteria essentially fell short in meeting the basic requirements. These evaluations had no discussion or were unclear on the process carried out in generating the recommendations. The recommendations were not practical and actionable. Some were assessed to have recommendations that seemed to be ambitious or easy to be achieved when there could be actual barriers, or not SMART (specific, measurable, attainable, realistic, timely). Some recommendations were too brief that it did not convey understandable information. Finally, some evaluations only noted the stakeholders in-charge of an action step, yet, there was no prioritization or categorization of these recommendations.

SECTION 7: GENDER AND HUMAN RIGHTS

Does the evaluation meet UN-SWAP evaluation performance indicators?

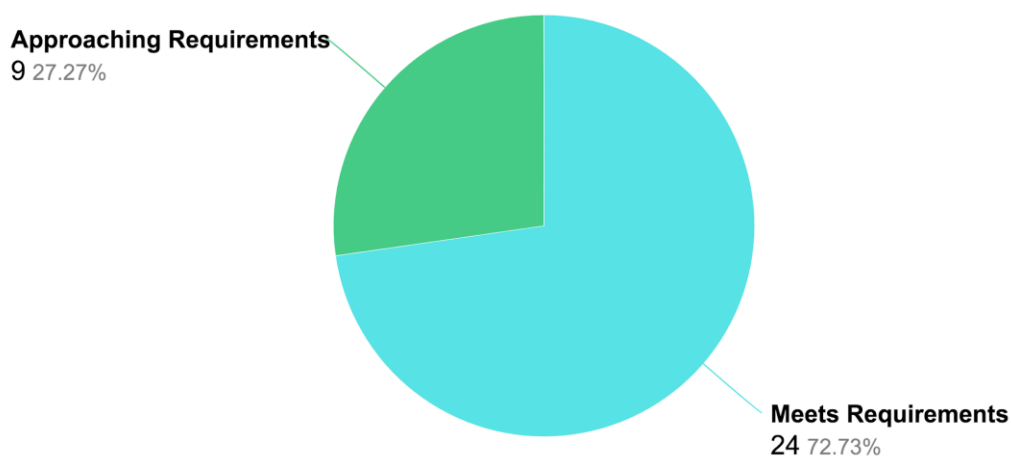
In January 2019, an in-depth assessment of these 33 evaluation reports using the UN-SWAP⁴ criteria was conducted. The UN-SWAP assesses the integration of gender equality and women's empowerment into evaluations. It assessed the use of gender-responsive methodology and the integration of gender analysis of findings, recommendations, and lessons learned, among other issues.

⁴ UN System-wide Action Plan (UN-SWAP) on Gender Equality and the Empowerment of Women (GEEW)

There were three main criteria used in the SWAP analysis: the integration of and attention to GEEW and gender mainstreaming in the evaluation framework, methodology, and findings. Integrated into each criterion were key questions that served as a guide on what to specifically assess in each evaluation report. A separate fourth criterion focused on identifying if at least one evaluation was conducted to assess corporate performance on gender mainstreaming.

Overall, UN Women evaluation reports were strong with the majority meeting SWAP requirements. Specifically, 73% of the evaluations (n=24) received the “Meets Requirements” rating, and 27% (n=9) of the evaluations received the “Approaches Requirements” rating. No reports “Missed requirements.” See chart below.

Overall Rating: Gender and Human Rights



All reports incorporated the Gender and Human Rights Criteria.⁵ As seen in the chart below, overwhelmingly reports integrated GEWE in their scope (n=24). There was a lot of variation in the sub-criterion of *gender methodology and tools* with 14 reports fully integrating these, 15 partly integrating these and 4 only partially integrating these tools. For the sub-criterion of *findings/conclusions/recommendations that reflect a gender analysis* had the lowest score with most of the reports (n=16) satisfactorily integrated this. Therefore, this warrants the acknowledgement and incorporation of systemwide influences and impacts of gender norms into the evaluation.

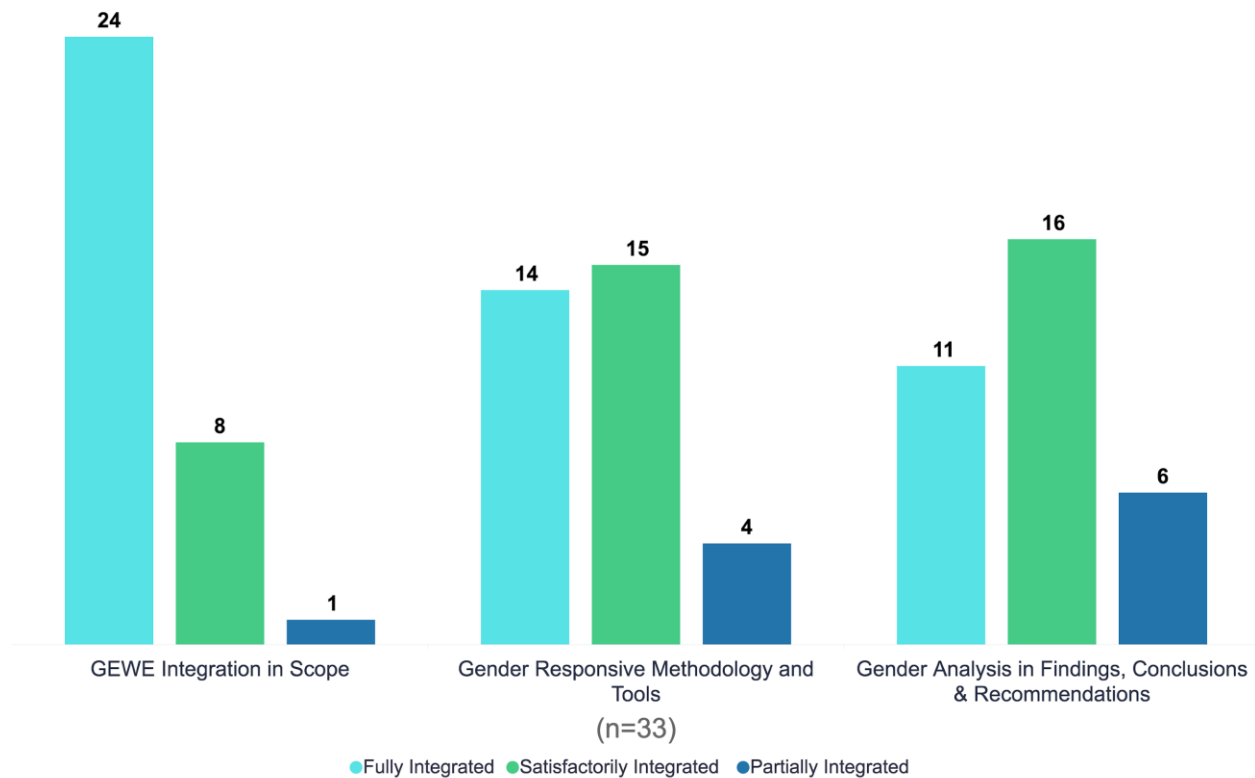
⁵ Full report on Section 7 is accessible via this link:
<https://docs.google.com/document/d/1YXtPso0EBfo9cuvzkQiwRWTQx2Sf77LkeMtIjp63u4/edit>

Evaluation reports that were rated as the strongest clearly illustrated the project gains and impacts with triangulated supportive evidence, statistics, and quotes from the project stakeholders as necessary, while also exploring the negative or unintended effects of the intervention. Strong reports also fully engaged in gender analysis throughout the results, implications and lessons learned and findings sections. A few examples appear below.

- From the Latin America and Caribbean region, the *“Evaluation Of The Safe Cities Campaign #Noesdehombres” (Mexico)* applied a robust methodology using mixed method approach with five central components: reach assessment, pre and post-campaigns surveys, focus groups (men and women), intuitive associations for the campaign images, and social media coverage analysis. Further, the evaluation documented changes in men’s and women’s perceptions of sexual harassment. The evaluation also tracked and explored underlying key gender norms that maintain sexist behavior (such as awareness from men and women that sexual harassment is a problem and should not be treated as a norm and that men understood that they play a major role in addressing the issue of sexual harassment). It also utilized gender analysis and gender lens throughout the methodology, results and recommendations.
- Another example of a strong report was the evaluation from Kyrgyzstan on the project *“Joint Programme on Accelerating Progress towards the Economic Empowerment of Rural Women (Ethiopia).”* The methodologies used in this project helped to lift up diverse and marginalized perspectives and had a dedicated focus on transformative changes--those changes that shifted the underlying institutional, community, and relational structures of inequalities and discriminations. Specifically, the evaluation used the “Change Map” methodology during focus group discussions as a way for rural women to illustrate the changes they encountered through the program. Further, this particular evaluation employed the Women Empowerment in Agriculture Index (WEAI) as a framework to structure their data gathering (such as the Change Map). The WEIA was also used as a framework to analyze the impact of the project. The evaluation also provided specific details on the transformative changes that took place.
- The Corporate Office evaluation, entitled, *“UN Women's Contribution To Women's Political Participation And Leadership”* rated high in the assessment with 8/9 points. A particularly strong focus of this evaluation was the dedicated use of gender analysis and use of the Inclusive Systemic Evaluation for Gender equality, Environments and Marginalized voices (ISE4GEMs) methodology. This methodology focused on rights based and gender-responsive approaches to data collection and analysis, such as a stakeholder analysis (to identify vulnerable groups likely to be affected by human rights violations and discriminations that should be included in the evaluation), boundary story analysis, and case studies. Through these different analytic approaches, the evaluation examined how human rights were accessed, exercised, and strengthened in addition to analysis of the barriers and facilitators to women’s political participation and leadership.

- Reports rated as “Approaches Requirements” most often had methodological issues. For example, often results were not backed up by evidence and appeared to be high-level summaries, without substantiating or exploring the effects of the intervention, findings appeared often at the output, instead of outcome level, and the evaluations did not explore unanticipated results and effects.
- In the future, UN Women can continue to emphasize strong characteristics with future evaluators and ensure the above weaknesses are fully addressed in the evaluations.
- A few recommendations for improvement to the SWAP tool and process follows.
 1. Incorporate some standardized gender analysis and vulnerability assessment into the SWAP criteria to help align agencies around a similar approach and improve attention to GEEW and gender mainstreaming.
 2. Offer further guidance and questions to examine intersectionality and vulnerabilities as they pertain to gender equality and human rights. Identifying and understanding the root causes of vulnerabilities is critical when evaluating GEEW programming because it provides an added lens of understanding how gender inequalities and human rights violations are faced by these vulnerable populations affect GEEW program design/implementation.
 3. Add a criteria to identify if transformation of power, norms, structures, relations or behaviours regarding GEEW is analyzed in the evaluation results. Currently, the criteria is limited to looking if there is a triangulation of different voices in the findings.

Section 7: Gender and Human Rights



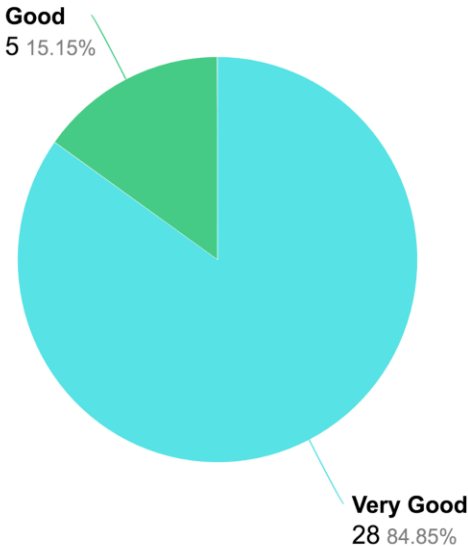
SECTION 8: REPORT PRESENTATION

Is the report well structured, written in accessible language and well presented?

The Report Presentation section focused on the organization and delivery of the evaluation content. Logical structure and readability were of importance in rating this criterion.

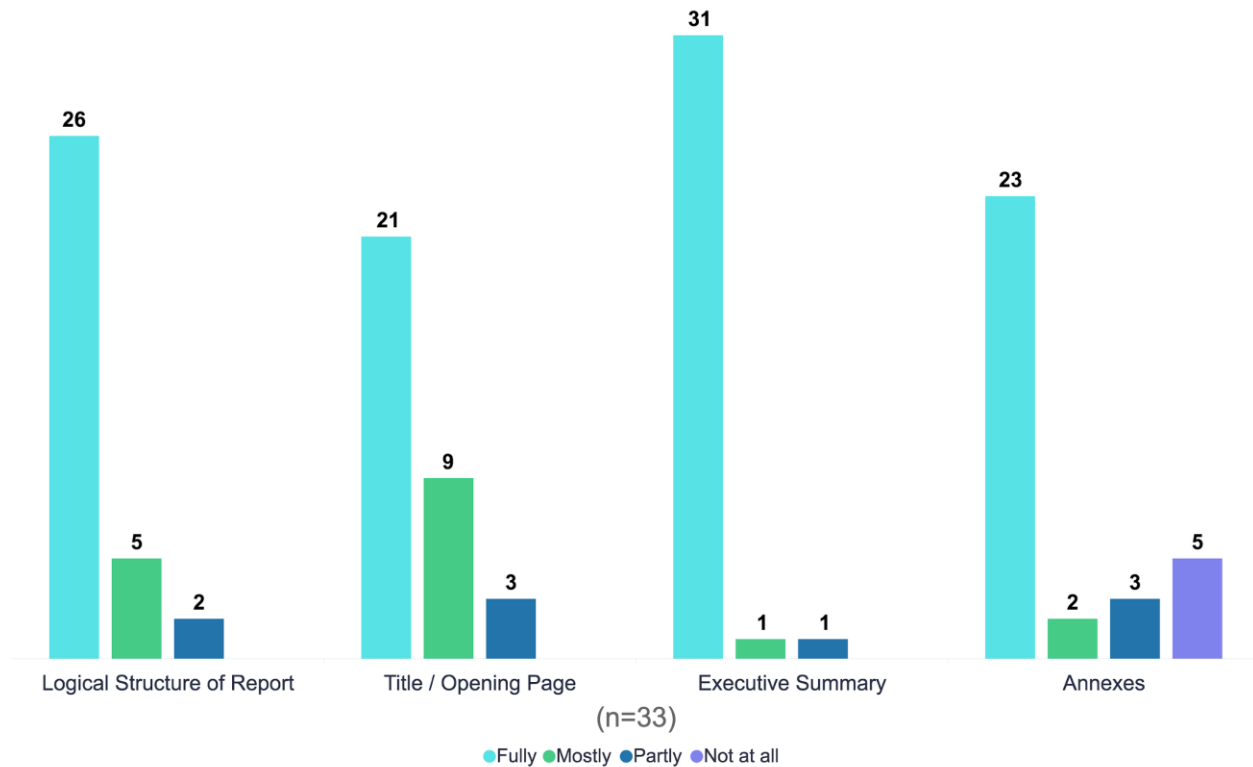
Eighty-five percent of reports scored “Very Good” in the Report Presentation criterion, and the remaining 15% received a score of “Good”. This was the only GERAAS criterion where reports received a minimum overall rating of “Good.”

Overall Rating: Report Presentation



Over 90% of the reports presented an executive summary (n=31), which summarized the key methodology, findings and recommendations, providing an overview of the whole evaluation report. The weakest sub-criterion in this section was the *Annex* sub-criterion where 15% (n=5) of reports received a rating of “Not at all”, and mostly this was due to some evaluators’ failure to include annexes in the printed report.

Section 8: Report Presentation



V. Good Practices

There were a variety of good practices presented in the evaluation reports. The examples cited below had strong ratings and exceeded standards on specific aspects of the GERAAS criteria.

Context and object of evaluation was extensively discussed. While it was required for each evaluation to explain the context (e.g., social, political, economic, demographic, and institutional) in which the intervention was implemented, there were evaluations that fared better on this criterion compared to others.

The “*Building Capacity to Prevent Violence Against Women (BCPVAW) in Bangladesh (Dec 2014 – Sept 2018)*”, aimed to end violence faced by women in universities. The evaluation extensively detailed the current situation of women by citing scientific studies, which revealed the incidence of sexual harassment at select local universities. It also included a presentation of the legal measures in place to prevent such violence and the weaknesses associated with the actual implementation of these policies. Like the previous example, it clearly laid out the specific problem and through this, the rationale of the project was also elucidated.

Presentation of stakeholder analysis and their level of engagement were important components of contribution analysis under the Methodology criterion. While stakeholder mapping was incorporated into most evaluations, a handful clearly articulated the role that stakeholders played in formulating/reviewing evaluation findings and recommendations.

- The stakeholder analysis matrix of the *Evaluation of Women's Employment Promotion Project (WEPP)* in Egypt not only outlined which stakeholders were engaged in the process, but also articulated their roles in the project, how they engaged in the evaluation and when, and expected gains from each stakeholder group's participation. Assigning priority levels of engagement to each stakeholder group also helped the evaluators determine which stakeholder groups were essential to inform/maintain contact throughout the entire evaluation exercise versus for only one segment (e.g., data collection).
- The *Rwanda Country Portfolio Evaluation* and the *South Sudan Country Portfolio Evaluation* both provided detailed stakeholder analysis as part of their evaluations. The Country Portfolio Evaluations (CPE) assessed UN Women contributions to three mandates - normative, coordination, and programming- while informing future programming. In the Rwanda CPE, stakeholder analysis was conducted, using a snapshot of those stakeholders engaged in country portfolio (programming) during the period of interest (2014-2018). Boundary analysis was part of this process, and through examination of each of the three UN Women pillars (normative, coordination, programming) and application of its theory of change, the Rwanda CPE noted partners, UN Women activities, the activities' target groups, target changes, and links to other activities. The *South Sudan Country Evaluation* used a similar analytic approach, using boundary analysis to note each stakeholder group (normative partners [e.g., Government of South Sudan], coordination partners [e.g., UN agencies], implementation partners) operating under each of the 3 UN Women pillars, and then aligned each group to a specific result area under the Development Results framework in place. For example, under the UN Women South Sudan Development Results framework result 3 - *women and girls live lives free from violence* - the South Sudan national gender machinery (normative partner), UN Women South Sudan Country Office staff (coordination partner), and implementing partner South Sudan Ministry of Health were all engaged in reviewing, adopting and implementing laws to protect women and girls from experiencing violence in their country.
- Outlining the vulnerabilities faced and ethical considerations for each stakeholder group contributed to extensive details regarding stakeholder engagement in the *UN Women's Contribution to Women's Political Participation and Leadership* evaluation. In addition to the multilevel stakeholder analysis that was conducted at evaluation inception, a vulnerability assessment accompanied a second round of stakeholder analysis at the start of data collection, which facilitated the application of methods and refinement of ethical safeguards, depending on the vulnerabilities identified.

Evaluations that utilized specific gender and human rights frameworks and tools to analyze findings and conclusions. Some evaluations stood out from the others as they made use of gender analysis and human rights frameworks and tools.

- One good example was the “*Joint Programme on Accelerating Progress towards the Economic Empowerment of Rural Women (JP-RWEE)*” in Kyrgyzstan. This particular evaluation employed the Women Empowerment in Agriculture Index (WEAI) and the Change Map approach to assess programme impact as it related to women’s economic empowerment. The WEAI quantitatively measures women’s empowerment, agency, and inclusion as connected to production, and agricultural growth, across 5 domains.⁶ In addition, “Change map(ping)” accompanied focus group discussions to document change within each of the five WEAI dimensions that rural women encountered as program participants. (Non-WEAI dimensions were included to allow non-index change to be acknowledged). Eighteen change maps were created, and in conjunction with the WEAI, facilitated the evaluation team’s substantiation of their findings, which contributed to good evaluation practice.
- Another example is the *UN Women’s Contribution to Women’s Political Participation and Leadership*, a Corporate (HQ) evaluation. This utilized the rights-based Inclusive Systemic Evaluation for Gender Equality, Environments and Marginalized Voices (ISE4GEMs) approach that brought together systemic evaluation with intersectional analyses. Boundary stories, an integral part of this approach, combine evaluability assessments, scoping studies and interviews-- supporting systemic triangulation of data, which aided the development of narrative summaries of UN portfolio activity and progress at the global level.
- The *Independent Evaluation of UN Women’s Fund for Gender Equality (FGE)* presented a theory of change for results-based programming and management, tying grantmaking with technical assistance, pivoted around three pillars (support, sustain, strengthen). This *Fund for Gender Equality (FGE)* was designed to address the funding shortfalls experienced by women’s rights organizations globally and subsequently, policy frameworks were developed to address these.
- The *Evaluation of Women’s Employment Promotion Project (WEPP in Egypt)* used the Gender Results Effective Scale (GRES) to assess the transformative nature of the outcomes which the program contributed to, tracking the presence of gender negative, blind, targeted, responsive and transformative results.

⁶ The WEIA 5 dimensions of women empowerment are: i) decisions about agricultural production; ii) access to and decision-making power over productive resources, iii) control over use of income, iv) leadership in the community, and v) time use.

Clear articulation/alignment to results frameworks, or the inclusion/depiction of explicit theories of change.

- Evaluations that received high overall ratings on the Methodology sub-criterion had their results framework or theory of change (TOC) available for the reviewer, noting outputs/outcomes (results) or directional change (for impact). The *Regional Evaluation on Normative Frameworks in the Americas and the Caribbean region*, which “Fully” met the Methodology sub-criterion and “Fully” presented its evaluation object including its theory of change, articulated a theoretical framework that guides the creation of UN Women legal and normative standards as applied to the global, regional and national spheres.

Detailed description of the ethical practices used in the evaluation that go beyond basic mentioning of adherence to UN Evaluation Group (UNEG) norms and standards, securing informed consent, maintaining confidentiality and do no harm.

- The “*End Evaluation of the Preventing and Responding to Violence Against Women and Girls in Ethiopia Programme*” team shared how they created an information sheet and translated it into the local language to facilitate the recruitment of participants. The *Evaluation of Women, Peace and Security in the Arab States* highlighted guidelines for participant engagement via interviews and group discussions, and noted the various UN ethics guidelines including the UN Women Evaluation Policy, UN Evaluation Group Ethical Guidelines and UN Code of Conduct for Evaluation in the UN System, that were put in place during the evaluation.
- The team in the “*Building Capacity to Prevent Violence Against Women (BCPVAW) in Bangladesh*” evaluation held a targeted data collection training on how to collect highly sensitive data (women were VAW survivors) and noted approaches for addressing ill treatment as a direct result of participation. Opportunities for redress were not limited to project or programme evaluations; country program evaluations such as the *Rwanda Country Evaluation 2014-2018* also noted opportunities for redress in addition to compliance for specific codes for engaging with vulnerable population groups.

VI. Conclusions

Conclusion 1: More highly rated reports typically articulated a strong gender analysis framework, and carried this lens throughout the evaluation design, implementation, analysis of results, findings and recommendations. Reports that scored higher on Gender and Human Rights criteria typically used a strong gender analysis framework, and weaved a gender lens throughout the evaluation design, implementation, analysis of results, findings and recommendations. A number of these evaluations and good practices are highlighted in the sections below and could be shared with evaluators to support future work.

Conclusion 2: The evaluators were consistent in reporting the value of CSO and stakeholder consultations and participation in shaping and achieving key outcomes. However, despite the stated efforts to reach more CSOs and vulnerable communities, many evaluation reports tended to engage with mostly the same type of stakeholder groups, often at the UN or partner level, and had limited engagement with more diverse groups, beneficiaries or other vulnerable populations, which would have deepened insights and strengthened analyses. Some reports needed to expand on how they would solve for these gaps using analytical approaches to improve visibility of intersectionality among historically marginalized groups in particular.

Conclusion 3: Most evaluation reports were consistent in highlighting mixed methods to improve the capturing of diverse program and project outcomes. Stronger evaluations noted the triangulation of data sources to enhance project or program implementation and identified data and information gaps, which affected the quality of the analyses. Weaker reports often had gaps in baseline information, lacked strong methods for data analysis, and especially a systematic approach to qualitative data analysis. Moreover, weaker evaluation reports tended to have less information on the diverse methodologies used, and a more superficial interpretation of the data and results.

Conclusion 4: The results of the meta-evaluation revealed declining GERAAS scores of the evaluations in the findings, conclusions and lessons learned, and recommendations sections. Findings are one of the most vital evaluation components - documenting the intervention effect and impact that were invested with time, money, effort, and personnel resources. Weaker reports often lacked strong evidence to underpin findings, analyses and support lessons learned. While some evidence in reports were more substantial than others, overall the assessment results revealed the need for the more consistent adoption of diverse data collection practices in alignment with the outcomes that were being measured.

Conclusion 5: Most reports provided limited information on how evaluation implementers engaged with UNEG ethical principles. These principles are important in understanding the practices used to engage and conduct ethical work with stakeholder groups that respect human rights and gender equality concerns. Most reports required examples of how principles were made actionable through the project or program.

Conclusion 6: Some reports noted that limited access to baseline information contributed to low quality reporting. This critical challenge resulted in a lack of deep analysis and lessons learned, thus highlighting the need for additional guidance for evaluators to solve for these challenges.

VII. Recommendations to enhance UN Women evaluations in the future

Recommendation 1: Underscore the importance of including a detailed summary of stakeholder engagement as part of each evaluation. To strengthen the relevance and validity of evaluations, additional information is needed to ensure the right internal and external stakeholders have provided meaningful input. At times, reports seemed to highlight a UN Women and partner

heavy stakeholder consultation process with limited participation of CSOs, communities and people affected by the projects. Outreach to a diverse set of stakeholders could strengthen work and provide new recommendations on future directions given a holistic perspective on UN Women's contributions.

Recommendation 2: Incorporate vulnerability assessments into evaluations, or analytic approaches that allow for deeper examination of intersectionality. A vulnerability assessment helps identify specific populations to target for engagement, moving beyond target groups of *just* women/girls, and include those groups who experience human rights violations and gender inequalities as well. This assessment should then shape and infuse the data collection, data analysis, and recommendations process.

Recommendation 3: Use diverse methods to enhance data collection and evaluation practices that allow for more intersectional analyses. UN Women should advise its evaluators to consider the use of diverse evaluation methods and data collection processes to facilitate analyses of multiple, complex data sources. Less commonly used methods incorporated into a couple evaluations include longitudinal social media analysis of a campaign, or the integration of diverse intersectional methods and analysis. For example, the *“UN Women's Contribution To Women's Political Participation And Leadership”* used ISE4GEMs methodology, which combined multiple methods (boundary analysis, FGDs) to conduct a range of intersectional analyses that allowed for examination of agency contributions across different institutional/activity pillars.

Recommendation 4: Provide more guidance to future evaluators on the presentation of findings, conclusions, lessons learned, and recommendations. Conclusions, lessons learned, and recommendations are important pieces of information that can guide better project design and implementation should there be future interventions. UN Women should emphasize with future evaluators that these sections necessitate substantiation with high-quality evidence.

Recommendation 5: Integrate gender considerations, and apply a gender lens systematically to improve evaluations. Ensure that evaluators have a specific gender lens that will be integrated throughout the evaluation process and the report. Providing guidance and information to support evaluators adoption of a gender lens to assess programs will help strengthen reporting and provide greater nuance and relevance of the analysis to help UN Women's future performance.

Recommendation 6: Outline for future evaluators the human research ethics compliance information necessary for sufficient, substantive presentation. Many of the reports merely presented a sentence that said they followed UNEG ethical principles without elaborating on them. While there were some time and resource limitations noted, UN Women should ask future evaluators to document how they are meeting ethical principles with substantive examples as it is a central part of ensuring a strong research methodology and process that is not biased and respects communities' rights, privacy, and confidentiality.

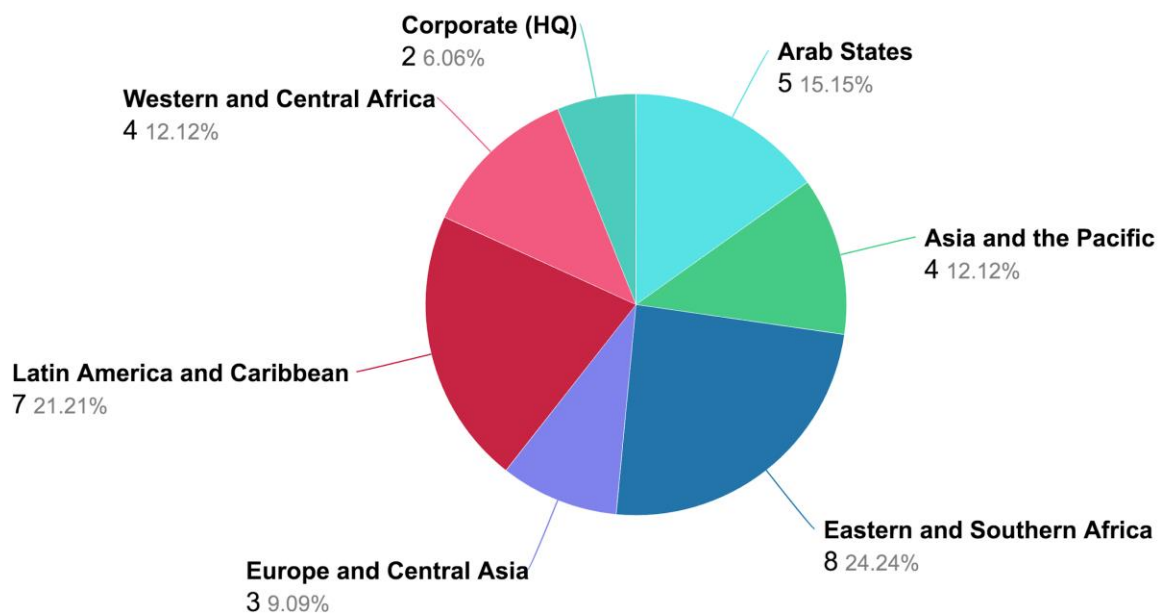
Recommendation 7: Ensure greater access to information to strengthen M&E of development interventions. The assessment results found that some evaluations did not fully analyze project gains or causal or contributory factors, in part because of limited baseline information. Some evaluations also noted the lack of project monitoring documents, which affected the quality of the evaluation. UN Women should consider greater support and reinforcement of M&E systems for development projects.

ANNEX 1: Contextual Presentation of the Evaluations

This annex presents additional insights on the regional distribution geographic scope, evaluation type, and strategic plan thematic area. This information helps lay the overall context of activity.

A. Regional Distribution of Reports

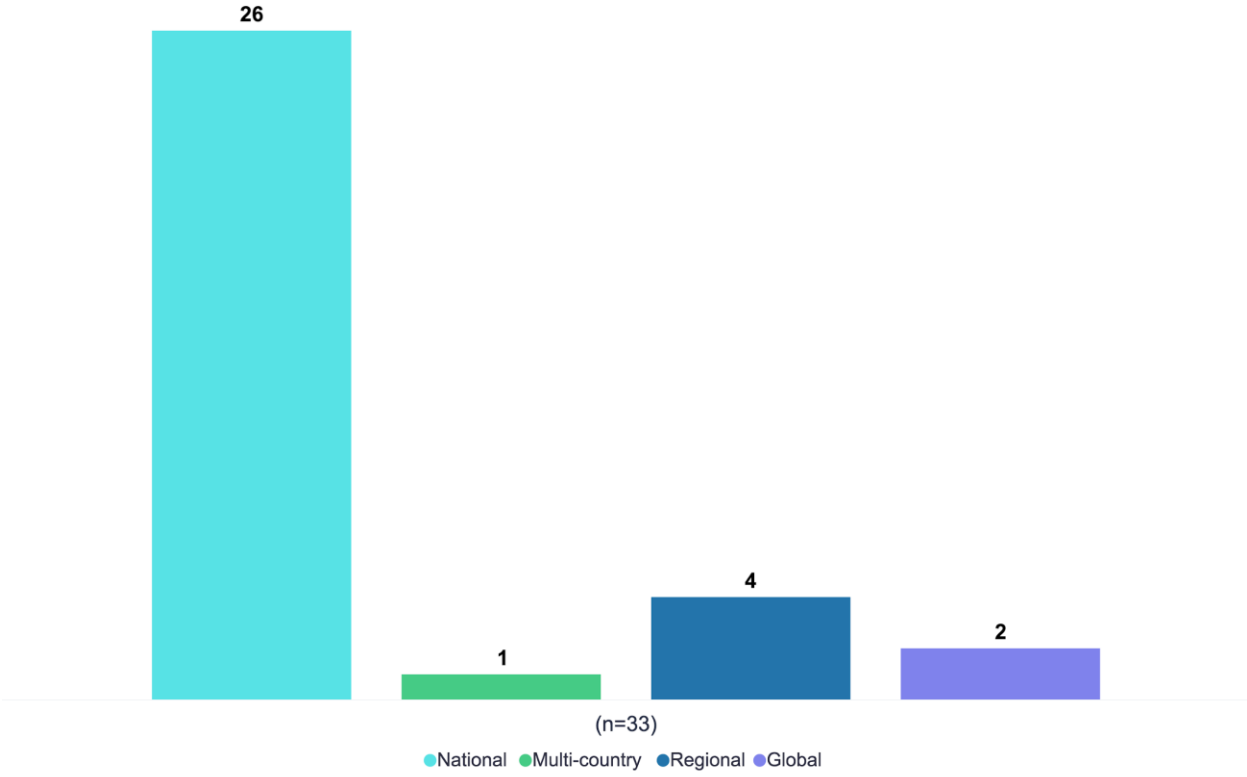
Evaluation Report, by Region



Nearly half of the evaluations were carried out in either the *Eastern and Southern Africa (ESA)* region (n=8), or the *Latin America and Caribbean (LAC)* region (n=7). Four evaluations were implemented in both *Asia and the Pacific (AP)*, and the *Western and Central Africa (WCA)* regions, while the *Arab States (AS)* region contributed 5 evaluations. Few evaluations were implemented in the *Europe and Central Asia* (n=3) and *Corporate* regions (n=2).

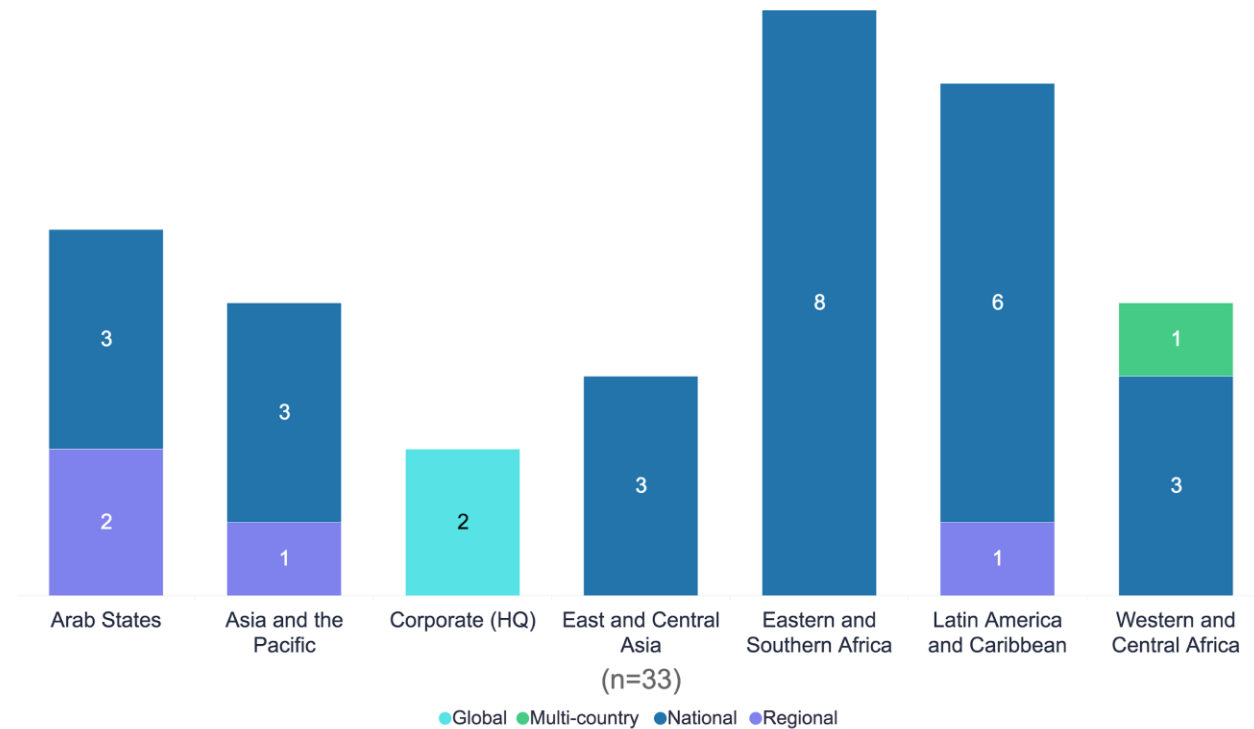
Geographic Coverage

Geographic Coverage of GERAAS Reports



The majority of evaluation reports that were included in the meta-evaluation were carried out at the national level (n=26). Only one evaluation was considered to be multi-country in orientation, and four were carried out at the regional level. Two evaluations were operationalized at the global level.

Geographic Coverage, by Region

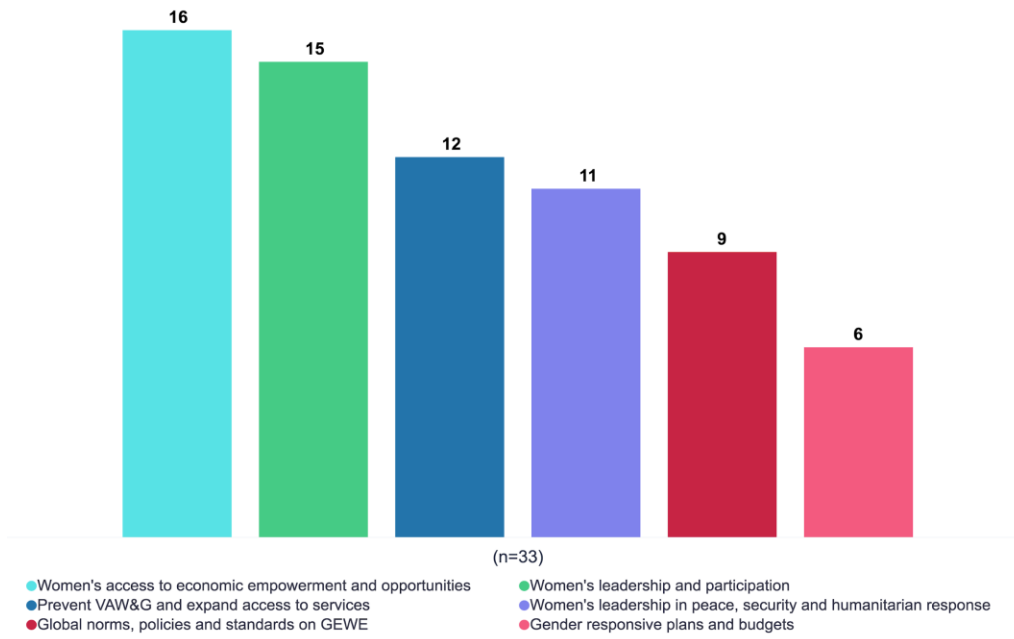


National-level evaluations accounted for the majority of those evaluations conducted in the Arab States, Asia and the Pacific, Latin America and the Caribbean, Asia and Western and Central Africa. National-level evaluations were the only type of evaluations conducted in both *East and Central Asia*, and *Eastern and Southern Africa* regions. Corporate evaluations (n=2) accounted for those at the global level, while the one multi-country evaluation that was included in the meta-evaluation was implemented in the Western and Central Africa region.

B. Strategic Plan Thematic Area

All evaluations addressed at least one strategic plan area. Thirteen evaluations addressed two or more thematic areas whereas five evaluations addressed all six thematic areas within their scope. *Women's access to economic empowerment and opportunities* was the most frequently reported thematic area addressed (n=16), followed by *women's leadership and participation* (n=15), as noted in the graph down below. The two strategic plan thematic areas least reflected in the meta-evaluation were *global norms, policies, and standards on GEWE* (n=9), and *gender responsive plans and budgets* (n=6).

Strategic Plan Thematic Areas in GERAAS Evaluation Reports

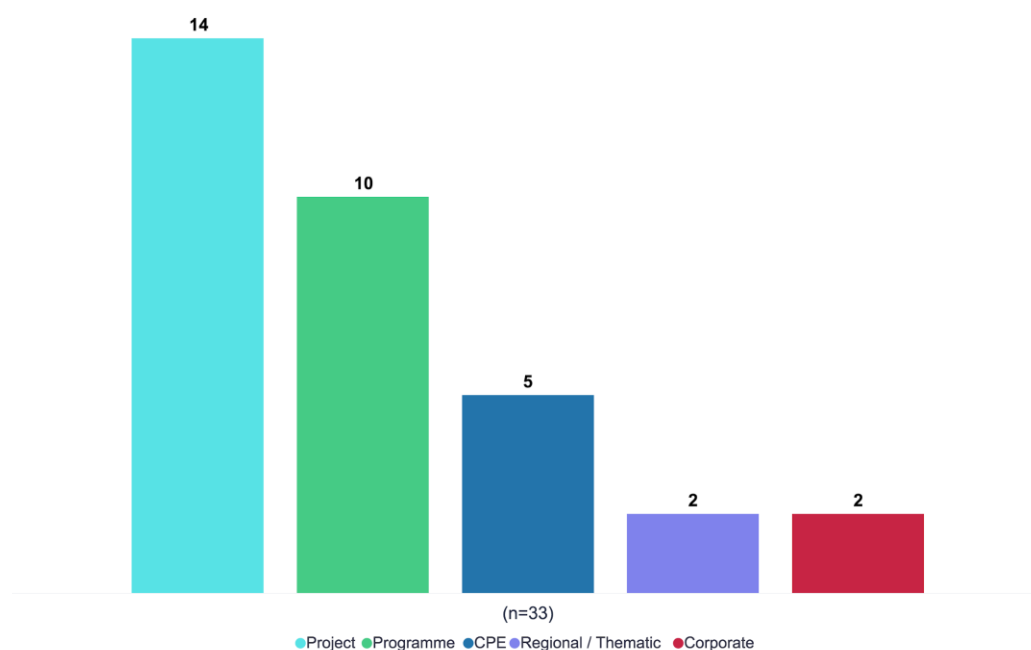


Regional Distribution of Evaluations, by Strategic Plan Thematic Area



Nearly 40 percent of the 16 evaluations that acknowledged *women's access to economic empowerment and opportunities* as one of its strategic areas, were carried out in Eastern and Southern Africa. One-third of the 15 evaluations (n=5) that featured *women's leadership and participation*, were reported in Eastern and Southern Africa. The Corporate (HQ) region, which had the fewest evaluations for this meta-evaluation, worked across three strategic areas: *women's leadership and participation*, *women's access to economic empowerment and opportunities*, and *global norms, policies, and standards on GEWE*. The East and Central Africa region had the fewest number of strategic thematic areas represented among those evaluations implemented in this region, which are 1) *women's access to economic empowerment and opportunities*, and 2) *women's leadership and participation*.

GERAAS Reports, by Evaluation Type



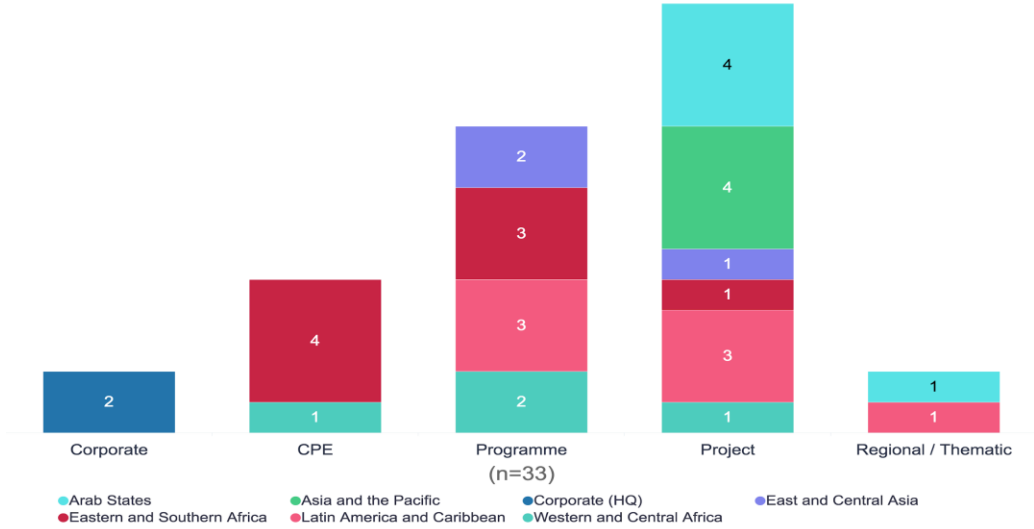
Over 40 percent of the evaluations (n=14) were project evaluations, and programme evaluations accounted for another 30 percent (n=10). Regional/thematic evaluations (n=2), and corporate evaluations (n=2) were the two types of evaluations least carried out by evaluators.

In assessing the geographic distribution of evaluation types, we found the following:

- *Corporate* evaluations exist only at the Global level;
- Evaluations carried out in the Arab States were mostly *project* type evaluations (n=4);
- CPEs- *country portfolio evaluations*- were only carried out on the African continent- either in Eastern and Southern Africa region, or Western and Central Africa region.

- The *project* evaluation type was the most widely documented in all regions except for *Corporate*.

GERAAS Reports, by Evaluation Type



Annex 2 - List of evaluations for the 2018 GERAAS Review

Region	Office/division	Title of evaluation	GERRAS Report quality rating
West and Central Africa	Mali	L'évaluation Finale Du Projet « Projet D'amélioration De L'accès Des Femmes Victimes De Violences Sexuelles Et Basées À La Justice Et À La Sécurité Dans Le Processus De Consolidation De La Paix Au Mali »	Fair
	Regional Office for West and Central Africa (Senegal)	End-Term Evaluation of the Joint Programme on Gender, Menstrual Hygiene And Sanitation	Very Good
	Cameroon	Prise En Charge Adéquate Des Femmes Victimes De Violences Dans La Région De L'extrême Nord	Good
	Liberia	Country Portfolio Evaluation	Very Good
East and Southern Africa	Burundi	Country Portfolio Evaluation	Very Good
	Kenya	Country Portfolio Evaluation	Good
		Evaluation of "Countering Violent Extremism" Project in Kenya	Good
	Rwanda	Final Evaluation of the Joint Programme "Advancing and Sustaining Gender Equality Gains in Rwanda"	Very Good
		Country Portfolio Evaluation	Good
	South Sudan	Country Portfolio Evaluation	Very Good
	Ethiopia	Evaluation of "Preventing and Responding to Violence Against Women and Girls in Ethiopia" Programme	Good
Mozambique	Mid-term Evaluation of WEE project in Gaza, Mozambique	Good	
Asia and the Pacific	Bangladesh	Evaluation of "Building capacity to Prevent Violence Against Women (BCPVAW)"	Good
	Multi-Country Office for the Pacific (Fiji)	End of Project Evaluation of the EC Strongim Mere: "Promoting Women's Political Participation And Representation in the Solomon Islands" (2014 - 2016)	Unsatisfactory
	Regional Office for Asia and the Pacific (Thailand)	Final Evaluation of Regional Project "Preventing Exploitation of Women Migrant Workers"	Good
	India MCO	Final Evaluation of RNE supported project "Promoting women's political participation in Sri Lanka"	Unsatisfactory
	Regional Office for Arab States (Egypt)	Evaluation of "UN Women's Economic Interventions under LEAP/HA programming in the Arab States region"	Good
	Jordan	Evaluation of UN Women's "Peace and Security in the Arab States" Regional Project	Good

	Egypt	Evaluation of “Securing Rights and Improving Livelihoods of Women (SRILW)” Action	Good
		Final Evaluation of UN Women’s “Promoting Women’s Employment by Creating Safe and Women-Friendly Workplaces” Programme (WEPP)”	Good
	Morocco MCO	L’évaluation Finale du Projet de l’opérationnalisation de l’entente de partenariat pour la promotion de la participation des Femmes à la gestion des affaires locales et le renforcement de la gouvernance territoriale sensible au genre (2014–2017)	Very Good
Europe and Central Asia	Albania	UN Women Albania Outcome Evaluation on “Women’s Leadership and Political Participation”	Good
	Kyrgyzstan	Joint Programme on: “Accelerating Progress Towards the Economic Empowerment of Rural Women”	Very Good
		Livelihoods Through Participation and Equal Access to Water	Good
Americas and the Caribbean	El Salvador	Evaluación Final Del Proyecto: “Ciudad Mujer/Onu Mujeres”	Good
	Mexico	Evaluation of “The Safe Cities Campaign #Noesdehombres”	Very Good
	Ecuador	Evaluación Del Proyecto ““Mujeres Liderando el Desarrollo Inclusivo Sostenible de la Provincia de Loja ”	Good
	Colombia	Final Evaluation of the “Women’s Citizenship for Peace, Justice and Development”	Good
	Colombia	Mid-term Evaluation of the Programme: “Overcoming Gender-Based Violence to Ensure Women’s Full Enjoyment of Rights”	Very Good
	Regional Office for Americas and the Caribbean (Panama)	Regional Evaluation on Normative Frameworks	Very Good
	Ecuador	Mid-term Evaluation of “Sustainability of the wasteland from a gender perspective”	Fair
Corporate	Independent Evaluation Service (IES)	UN Women’s Contribution to Women’s Political Participation and Leadership	Very Good
Headquarters	Fund for Gender Equality	Independent Evaluation of “UW Women’s Fund for Gender Equality” (2009–2017)	Very Good

Annex 3. UN Women Global Evaluation Quality Assessment and Rating Matrix (sample)

Independent Evaluation and Audit Services (IEAS)
UN Women Global Evaluation Quality Assessment and Rating

Rating Scale	Very Good	Good	Fair	Unsatisfactory	Reviewer Guidance :	
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.	<ul style="list-style-type: none"> - Overall reports are rated against a 4-point scale (Very Good, Good, Fair and Unsatisfactory), which is an aggregated rating of eight parameters. - Each overarching parameter is rated against a 4-point scale (Fully, Mostly, Partially and Not at all). - Parameters such as evaluation methodology, findings, conclusions and recommendations are given more weight. - Executive feedback - provide summary of the extent to which the report meets or fails to meet the criteria provided under each parameter. Please also include suggestion on how to improve future evaluation practice. The overall review, rating , and the executive feedback will be provided to the evaluation commissioning office. 	
Parameter Weight (%)	1: Object and context	5	5: Conclusions and lessons learned	20	<i>Are weightings equal to 100%?</i> OK	
	2: Purpose and scope	5	6: Recommendations	15		
	3: Methodology	15	7: Gender Equality and Human Rights (UN-SWAP)	10		
	4: Findings	20	8: Presentation	10		
PART I: REPORT DETAILS						
Report title	Corporate Evaluation on Strategic Partnerships for Gender Equality and the Empowerment of Women			Geographical Coverage		Global
Sequence number		Evaluators	[Female]	[Male]	Year	2018
Region	Europe and Central Asia	Country(ies)			Type of intervention evaluated	CPE
Portfolio Budget (USD)		Evaluation Budget (USD)			Reviewer	

Strategic Plan Thematic Area (select all that apply)	Women's leadership and participation	Women's leadership in peace, security and humanitarian response	Women's leadership in peace, security and humanitarian response	Review Date	February 22nd, 2018
	Women's access to economic empowerment and opportunities				

PART II: THE EIGHT KEY PARAMETERS

SECTION 1: OBJECT AND CONTEXT OF THE EVALUATION (weight 5%)	RATING	Good	Criteria Weight	Weighted increments	Raw point score
Does the report present a clear and full description of the 'object' of the evaluation?	50%	Executive Feedback on Section 1			
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.	Fully		25%	0.42	1.25
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.	Mostly		25%	0.42	0.83
1.3 The key stakeholders involved in the implementation, including the implementing agency(s) and partners, other stakeholders and their roles are described.	Partly		25%	0.42	0.42
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.	Not at all		25%	0.42	0.00
SECTION 2: PURPOSE, OBJECTIVES AND SCOPE (weight 5%)	RATING	Very Good			
Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation?	83%	Executive Feedback on Section 2	Criteria Weight	Weighted increments	Raw point score

2.1 Purpose, objectives and use of evaluation: The evaluation report provides 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.	Fully		50%	0.83	2.50
2.2 Evaluation Scope: The evaluation report provides 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 ToRs, 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).	Mostly		50%	0.83	1.67
SECTION 3 : METHODOLOGY (weight 15%)	RATING	Very Good			
Is the methodology used for the evaluation clearly described and appropriate, and the rationale for the methodological choice justified?	83%	Executive Feedback on Section 3	Criteria Weight	Weighted increments	Raw point score
3.1 Methodology: The report specifies and provides complete description of a relevant design and sets of methods including the chosen evaluation criteria, questions, and performance standards. The methods employed are appropriate for analyzing gender and rights issues identified in the evaluation scope.	Fully		35%	1.75	5.25
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.	Mostly		40%	2	4.00
3.3 Stakeholders Consultation: The evaluation report gives a complete description of stakeholder's consultation process in the evaluation, including the rationale for selecting the particular level and activities for consultation.	Fully		10%	0.50	1.50
3.4 Limitations: The report presents clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias.	Partly		5%	0.25	0.25
3.5 Ethics: The evaluation report includes a discussion of the extent to which the evaluation design included ethical safeguards and mechanisms and measures that were implemented to ensure that the evaluation process conformed with relevant ethical standards including but not limited to informed consent of participants, confidentiality and avoidance of harm considerations.	Fully		10%	0.5	1.50
SECTION 4: FINDINGS (weight 20%)	Rating	Good			

Are the findings clearly presented, relevant and based on evidence?	70%	Executive Feedback on Section 4	Criteria Weight	Weighted increments	Raw point score
4.1 The evaluation report findings provide sufficient levels of high quality evidence to systematically address all of the evaluation questions and criteria.	Partly		30%	2.00	2.00
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 made.	Mostly		30%	2.00	4.00
4.3 The causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified.	Fully		20%	1.33	4.00
4.4 Findings are presented with clarity, logic and coherence (e.g., avoid ambiguities).	Fully		20%	1.33	4.00
SECTION 5: CONCLUSIONS AND LESSONS LEARNED (weight 20%)	Rating	Fair			
Are the conclusions clearly presented based on findings and substantiated by evidence?	40%	Executive Feedback on Section 5	Criteria Weight	Weighted increments	Raw point score
5.1 Conclusions are well substantiated by the evidence presented and are logically connected to evaluation findings.	Fully		40%	2.67	8.00
5.2 The conclusions reflect reasonable evaluative judgments that add insight and analysis beyond the findings	Not at all		40%	2.67	0.00
5.3 Conclusions present strengths and weaknesses of the object (policy, programmes, project's or other intervention) being evaluated, based on the evidence presented and taking due account of the views of a diverse cross-section of stakeholders.	Not at all		15%	1.00	0.00
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.	Not at all		5%	0.33	0.00
SECTION 6: RECOMMENDATIONS (weight 15%)	Rating	Very Good			
Are the recommendations relevant, useful, and actionable and clearly presented in a priority order?	100%	Executive Feedback on Section 6	Criteria Weight	Weighted increments	Raw point score
6.1 Recommendations are logically derived from the findings and/or conclusions.	Fully		30%	1.50	4.50

6.2 The report describes the process followed in developing the recommendations including consultation with stakeholders.	Fully		20%	1.00	3.00
6.3 Recommendations are clear, realistic (e.g., reflect an understanding of the subject's potential constraints to follow-up) and actionable.	Fully		30%	1.50	4.50
6.4 Clear prioritization and/or classification of recommendations to support use.	Fully		20%	1.00	3.00
SECTION 7: GENDER AND HUMAN RIGHTS (weight 15%)	Score	Approaching Requirements			
Does the evaluation meet UN SWAP evaluation performance indicators? Note: this section will be rated according to UN SWAP standards.	67%	Executive Feedback on Section 7	Criteria Weight	Weighted increments	Raw point score
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.	Partially integrated (1)		33%	1.11	1.11
7.2 A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated (2)		33%	1.11	2.22
7.3 The evaluation findings, conclusions and recommendation reflect a gender analysis.	Fully integrated (3)		33%	1.11	3.33
SECTION 8: THE REPORT PRESENTATION (weight 10%)	Rating	Fair			
Is the report well structured, written in accessible language and well presented?	40%	Executive Feedback on Section 8	Criteria Weight	Weighted increments	Raw point score
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 (for instance, with numbered sections, clear titles and sub-titles; context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations) and written in an accessible language with minimal grammatical, spelling or punctuation errors.	Mostly		40%	1.33	2.67
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.	Partly		10%	0.33	0.33

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.	Not at all		40%	1.33	0.00
8.4 Annexes should include, when not present in the body of the report: Terms of Reference, 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).	Fully		10%	0.33	1.00
Additional Information					
Identify aspects of <i>good practice</i> of the evaluation					

PART III: THE OVERALL RATING			
Key Guiding Question	Total weighted score %	Overall Rating	Other reviewer's comments
Is this a credible report that addresses the evaluation purpose and objectives based on evidence, and that can therefore be used with confidence?	66.83	Good	