ImpactReady™

UN Women
GERAAS 2016

*Global Evaluation Reports Assessment and Analysis System*

META EVALUATION REPORT

Version 2.0

March 2017

Prepared for UN Women Independent Evaluation Office
Marco Segone, Messay Tassew (Evaluation Manager)

Prepared by ImpactReady LLP
Joseph Barnes

# Contents

# 1. Acronyms

| | |
|---|---|
| ECOSOC | United Nations Economic and Social Council |
| EO | UN Women Evaluation Office |
| EVAW | Ending Violence Against Women |
| GBV | Gender Based Violence |
| GEEW | Gender Equality and Empowerment of Women |
| GERAAS | Global Evaluation Reports Assessment and Analysis System |
| HQ | UN Women Head Quarters |
| M&E | Monitoring and Evaluation |
| OECD DAC | Organisation for Economic Cooperation and Development – Development Assistance Committee |
| SGBV | Sexual and Gender Based Violence |
| ToR | Terms of Reference |
| UN SWAP | United Nations System Wide Action Plan for Gender-Responsive Evaluation |
| UNEG | United Nations Evaluation Group |
| UNSCR | United Nations Security Council Resolution |
| VAW | Violence Against Women |

# 2. Executive Summary

## 2.1 Background

The Purpose of this meta-evaluation is to report on the quality of evaluation reports from 2016, and the trends in evaluation quality since 2013. Evaluation in UN Women contributes to learning on the best ways to promote gender equality and women's empowerment, enhancing UN Women's accountability, and informing decision-making. To address the organizational demands for ensuring good quality and credible evaluations particularly at decentralized level, the Independent Evaluation Office has designed a Global Evaluation Reports Assessment and Analysis System (GERAAS).

GERAAS uses the UNEG evaluation reports standards as a basis for review and assessment of final evaluation reports, while ensuring specific standards relevant to UN Women. Reports are rated, analyzed and synthesized through a 5-step process that has been refined over 4 years.

Reports are independently rated and quality assured based on 8 parameters of quality and 39 indicators. Each indicator, each parameter and the overall report is rated on a four-point scale: Very Good, Good, Satisfactory, Unsatisfactory. Reports rated Good or Very Good meet UNEG standards and can be used with confidence. Reports rated Satisfactory should be used with care. It is recommended that reports rated Unsatisfactory are only used with extreme caution.

The Independent Evaluation Office (IEO) oversaw, coordinated and supported the review process.

## 2.2 Findings

**Evaluation coverage is increasing.** The meta-evaluation assessed 36 reports from 2016, a substantive increase from the 27 reports rated from 2015 and 21 from 2014. Over the course of the 2014-2017 Strategic Plan, UN Women has completed 84 evaluations[1]. 52 countries[2] including programme presence were covered by evaluations in 2016[3]. Since 2013, a total of 82 regional, country and programme presence have been included in evaluations. 12 countries including programme presence were included in evaluations in 2016 that have not previously be covered. Despite the overall improvement in coverage, however, it remains geographically uneven.

**Quality of evaluation reports is stabilising.** In 2016, 33% of reports were Very Good, 50% Good, and 17% Satisfactory, an incremental improvement over the previous year. For the third year in a row, no reports were Unsatisfactory. The majority of evaluations (83%) can be used with full confidence. The remaining evaluation reports can still be used for management, accountability, and learning; but with awareness of the limitations that they face.

**Evaluations are becoming more strategic.** The portfolio of project and programme evaluations is more strategic than in previous years. No evaluation in 2016 was limited to the output level of analysis, with 64% attempting to measure outcome-level changes (the result of interventions

---

[1] 108 evaluations have been rated in the four years since GERAAS began, covering 2013-2016
[2] This figure includes countries covered through country case studies for corporate, regional and HQ evaluations
[3] Countries are only counted where they are explicitly covered as a case study in an evaluation, rather than implicitly as part of a desk review.

by multiple actors) and the remaining 46% attempting to measure higher-level outcomes (the cumulative effects on people's lives and human rights). Evaluations have also started to cover multiple impact areas under the Strategic Plan 2014-2017. The most evaluated impact area was Women's Economic Empowerment.

**Evaluations meet UN SWAP standards for gender responsiveness.** The meta evaluation found a significant increase in performance of evaluation reports with regard to the UN SWAP EPI, with the highest recorded performance for UN Women since tracking of this indicator began. UN Women evaluations were found to meet the UNEG requirements, with an average performance rating of 8.31 (out of a maximum score of 12). This is a substantive improvement over the 2015 and 2014 averages. As a result UN Women has met the agreed deadline for reaching UN SWAP standards.

**Evaluation quality at the decentralised levels continues to improve.** Within the regional architecture, decentralised evaluations have increased in number over time to reach 36 in 2016. Of the six UN Women regions, half (3) produced evaluations that completely meet the UNEG/UN Women standards (Europe and Central Asia; Eastern and Southern Africa; and Arab States). For each of the other three regions, a majority of reports are also rated as Good and Very Good. As per previous years, centralised (corporate evaluations managed by the Independent Evaluation Office) were rated as Very Good.

**Evaluation teams and designs remain relatively uniform.** The average team size for most evaluations (non-global decentralised evaluations) during 2016 was just over two people, most typically comprising of an international female and a national female evaluator. Global evaluations included an average of nearly 6 team members. The majority of evaluations are managed by UN Women staff, with only 2 jointly managed evaluations. There were no country-led evaluations. Most of the evaluations are primarily qualitative in nature. Few evaluations undertake a survey or any other method to collect first-hand quantitative data. This was also recognised in previous meta-evaluations, and is also a feature highlighted in meta evaluations of some other UN entities. Whilst it is a limitation, it also reflects the intersection of competing demands: the requirement for gender responsive methods, the level of resourcing available for evaluations, and the nature of the interventions being evaluated.

## 2.3 Conclusions

**The quality of evaluation reports in 2016 represents a multi-dimensional improvement that is the culmination of guidance and efforts to improve the credibility and utility of the evaluation function.** Over the course of the Strategic Plan 2014-2017, the evaluation function has demonstrated continuous improvement in coverage, quality, and gender responsiveness. All of these factors contribute to enhancing the credibility and utility of evaluation in support of management, accountability and learning.

**Scope remains to enhance the overall quality of decentralised evaluations regarding specific aspects of quality, but the number of these aspects is decreasing.** As with 2015, the aspects of evaluation reports controlled by the Terms of Reference (set by UN Women) remain strong, suggesting that the key to advancing quality is ensuring the application of proper methods, analysis and reporting by evaluators once they are commissioned.

**The coverage and scope of UN Women evaluations is increasing; whilst the range of designs and approaches used remain fairly narrow.** Whilst the increase in coverage of evaluations is welcome, it remains unevenly distributed. Along with the continuing similarity of evaluation designs, and the constraints that are leading to this concentration of approaches, this represents an area for consideration in the evaluation policy.

**The GERAAS system has contributed to UN Women having met the requirements of the UN SWAP evaluation performance indicator before the target date of 2017.** The increased emphasis within the Independent Evaluation Office on ensuring that the requirements for GEEW in evaluation were fully integrated, including through guidance but also through the regional evaluation architecture, is fully reflected in the significant improvements represented in the performance of 2016 evaluations. This illustrates the value of the accountability system established by the Independent Evaluation Office for improving the performance of evaluation.

## 2.4 Recommendations

**1: The Independent Evaluation Office is advised to focus the next round of guidance on improving lessons learned, comprehensive conclusions, and transparent recommendations.** The main areas requiring improvement within the GERAAS indicators for the 2016 portfolio of evaluations relate to the extended analysis of evidence required for the conclusions, recommendations and lessons learned. IEO is recommended to provide guidance, advice, examples and visibility to improving the quality of this analysis by evaluators.

**2: Regional Evaluation Specialists are recommended to ensure that the gains in gender responsive evaluation to UN SWAP standards are maintained and extended in 2017.** The decentralised evaluation portfolio has demonstrated substantive and significant improvements in the integration of gender responsive evaluation frameworks, methods, and analysis. This is a major achievement, and should remain a priority to ensure that it is maintained.

**3: The Independent Evaluation Office should review and enhance the GERAAS tool for the Strategic Plan 2018-21 period to include explicit weighting between parameters, disaggregate analysis of evaluation criteria, and reflect the UNEG 2016 norms and standards.** The consistency of the GERAAS tool has provided a longitudinal set of data that has given important insights into trends in the evaluation function. However, there is also scope to revise and further strengthen the robustness of the tool. There is also an opportunity to reflect the most recent updates to the UNEG norms and standards (2016), and to consider how the tool might adapt to a diversity in types of evaluations under renewed commitments to national evaluation capacity and leadership.

**4: UN Women is recommended to ensure continued support to enhancing the quality of the decentralised evaluation function.**
The establishment and embedding of the regional evaluation architecture is strongly correlated in GERAAS with a period of substantive and substantial improvement in the quality and coverage of evaluation. This, in turn, enhances the independence, credibility and utility of the evaluation function – and represents a recognised source of value for money in terms of organisational accountability, learning, and evidence-based decision making. Ensuring continued support to this capacity is vital to maintaining the performance of the evaluation function that is reflected in GERAAS trends.

# 3. Background

The purpose and role of evaluation in UN Women is to contribute to learning on best ways to promote gender equality and women's empowerment, enhance UN Women's accountability, and inform decision-making. By providing evidence-based information, evaluation contributes to UN Women's role to generate knowledge on what works to advance gender equality and women's empowerment.

The UN Women Evaluation Office (EO) provides leadership for the evaluation function throughout the organization, and leads the UN system on gender responsive evaluation and promotes accountability and evaluative evidence on UN gender equality results.

The UN Women Evaluation Policy came into effect in January 2013 and a new Strategic Plan (2014-2017) was endorsed in September 2013. A landmark System-Wide Action Plan (UN-SWAP) on gender equality and women's empowerment was also adopted that requires annual reporting against a performance indicator on gender-responsive evaluation.

Given the decentralized nature of the organization, the majority of the evaluations supported by UN Women are managed at a decentralized level. To address the organizational demands for ensuring good quality and credible evaluations particularly at decentralized level, the IEO has designed a Global Evaluation Reports Assessment and Analysis System (GERAAS) driven by similar good practices enforced by other UN entities and consistent with the UNEG Norms and Standards.

The system is believed to increase the application of sound approaches and methods to continuously improve the quality and credibility of evaluation methods and reports within the organization.

An independent firm was appointed to undertake both a meta-evaluation and meta-analysis of 2016 evaluation reports submitted to GERAAS, including the UN-SWAP scores[4]. A total of 36 reports were included.

---

[4] ImpactReady LLP

# 4. Purpose, Objectives & Scope

The Purpose of this meta-evaluation is to capture the quality of evaluation reports according to UN Evaluation Group standards. This is required to develop constructive lessons for future systemic strengthening of evaluation.

The Global Evaluation Report Assessment and Analysis System (GERAAS) has four main objectives:
1. **Improve the quality and utility of evaluation reports:** improve the use of evaluation reports by providing an objective assessment of the overall quality of the evaluation reports to Senior Managers and the Executive Board;
2. **Strengthen internal capacity on gender responsive evaluation:** promote sound evaluation design and methodology as well as consistent and quality reporting through building internal capacity on managing and quality assuring evaluations;
3. **Improve UN Women's performance and organizational effectiveness:** provide senior management with better understandings and insights into key UN women performance areas requiring attention; and
4. **Promote learning and knowledge management:** help promote organizational learning and knowledge management through capturing experiences and lessons learned from credible evaluations.

This assessment considers all 2016 reports submitted to the GERAAS system that were assessed, according to the UN Evaluation Group definition, to be evaluation reports (rather than reviews, evaluability assessments, baselines, studies, etc). It considers only the evaluation report, as presented on the UN Women GATE system (http://gate.unwomen.org) as a standalone document. The actual evaluation process or utilisation of the evaluation is currently considered outside the scope of this analysis. It should be recognised, therefore, that this report only provides a partial view in answer to the question *"what is the quality of evaluation in UN Women?"*

# 5. Methodology

GERAAS uses the UNEG evaluation reports standards as a basis for review and assessment of final evaluation reports, while ensuring specific standards relevant to UN Women. The Independent Evaluation Office (IEO) oversaw, coordinated and supported the review process.

Reports are independently rated and quality assured based on 8 parameters of quality and 39 indicators. Each indicator, each parameter and the overall report is rated on a four-point scale:

| Very good: | A 'very good quality' evaluation report is a report that has the features of being credible, addressing the evaluation questions, based on evidence, and, adheres to UNEG adapted UN Women Evaluation Report Standards. The report can be used with confidence and is considered a good example. |
|---|---|
| Good: | The report adheres to UNEG/UN Women evaluation standards, good analysis and credible recommendations. The report can be used with confidence. |
| Satisfactory: | The report meets requirements with regard to quality but some elements are missing or inadequately addressed. The report has useful information. |
| Unsatisfactory: | Reports rated unsatisfactory entail serious limitations and hence caution should be exercised when using the findings or recommendations for learning, accountability, evidence generation or informed decision making. |

Reports are rated, analyzed and synthesized through a 5-step process that has been refined over the course of GERAAS 2013-2016. This meta evaluation assesses final evaluation reports from fiscal year 2016 that were uploaded in the UN Women Global Accountability and Tracking of Evaluation System (GATE) by January 2017. It includes a rating of all reports using the UN SWAP evaluation performance indicator to the standards agreed by the UNEG working group on gender equality and human rights.

The Independent Evaluation Office will present the findings of the review at the Annual Session of the Executive Board and to the Senior Managers and the Global Evaluation Committee. The report is also to be shared with concerned HQ divisions, Regional Offices (RO) and Country Offices (COs) to improve the quality and utility of evaluations by highlighting the strengths, good practices and areas that require improvement.

The final report will be posted in the GATE System to allow access to the general public. This contributes to the transparency and credibility of UN Women when reporting on its performance. The accompanying meta-synthesis also serves as a useful repository of information on UN Women's operations at global, regional and country levels.

**A full explanation of the method is included in Annex 9.4.**

## 5.1 Limitations

**GERAAS does not measure the quality of evaluation processes.** It is designed to assess the quality of *reports* – considered a major output of evaluations – and it does so against a very specific and prescriptive set of UNEG standards for what an evaluation report should look like.

The benefit of this approach is **consistency against a set of standards** that are not only widely available but should also be provided to all evaluation teams prior to working for UN Women. A limitation of the approach is the reliance on a single source of information (the evaluation report) to develop a view on the utility of an evaluation.

The use of UNEG and UN Women standards also allows for comparison of reports across a wide range of budgets, time, and quality assurance mechanisms. The reports are assessed as a document, and thus, a project evaluation report that describes a methodology that is appropriate to a 'simple' evaluand and developed relevant conclusions can be compared to a corporate evaluation that describes a far more elaborate design and set of conclusions for a far more complex evaluand. The corporate report may be more detailed and complex – but it needs to be to reflect that nature of what is being evaluated at this level. Thus, **a 'Good' output-level report may not look like a 'Good' impact-level report,** but it may still meet the requirements of UN Women standards.

Given that the meta-evaluation is based only on a limited number of evaluation reports, it also has limitations connected to developing findings around methodological and participatory processes. The report aims to highlight where there is uncertainty, and makes transparent suggestions for explanations where the data does not support firm conclusions.

# 6. Findings

## 6.1 Overall Ratings and Feedback

### Finding 1: Evaluation coverage is increasing

The meta-evaluation assessed 36 reports from 2016, a substantive increase from the 27 reports rated from 2015 and 21 from 2014. Over the course of the 2014-2017 Strategic Plan, UN Women has completed 84 evaluations[5].

52 countries[6] with programme presence were explicitly covered by evaluations in 2016, reflecting the increasingly multi-country, regional, and global nature of UN Women evaluations[7]. Since 2013, a total of 82 countries including programme presence have been included in evaluations, with each country included in an average of 2.6 evaluations each (i.e. biennially).

12 countries including programme presence were included in evaluations in 2016 that have not previously be covered. Despite the overall improvement in coverage, however, it remains geographically uneven with 27 countries being covered only once in the past 4 years. The most covered countries since GERAAS began are Nepal (10), Jordan (8) and Georgia (7); the most covered countries in 2016 were Senegal (5 evaluations) and Nepal (4 evaluations).

*Figure 1: Coverage of evaluations in 2016 (orange) and since 2013 (blue)*
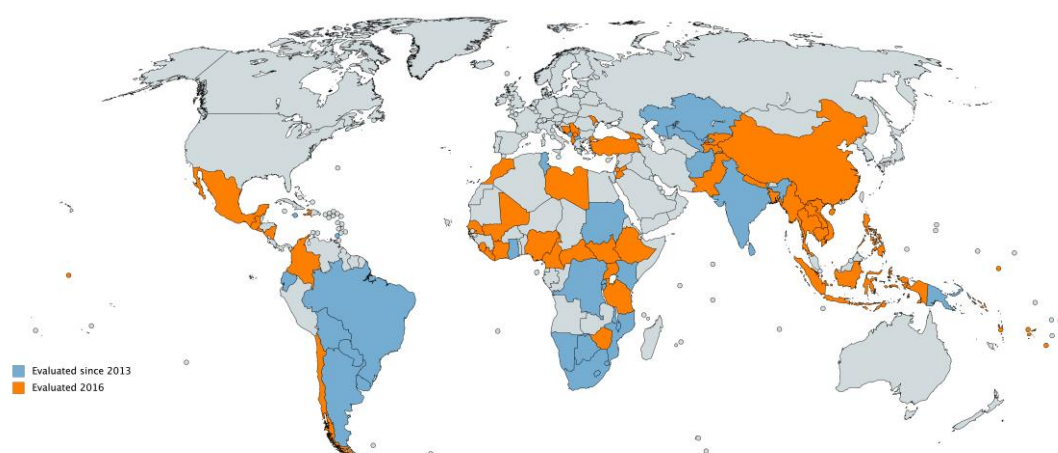


### Finding 2: Quality of evaluation reports is stabilising

Over the period of the Strategic Plan 2014-2017, no evaluation has been found to be unsatisfactory according to UNEG/UN Women standards. The level of reports rated Good or Very Good appears to have stabilised at around 80%, meaning that the majority of evaluations can be used with confidence. The remaining evaluation reports (17% rated Satisfactory in 2016) can still be used for management, accountability, and learning; but with awareness of the limitations that they face.

---

[5] 108 evaluations have been rated in the four years since GERAAS began, covering 2013-2016

[6] This figure includes countries covered through country case studies for corporate, regional and HQ evaluations

[7] Countries are only counted where they are explicitly covered as a case study in an evaluation, rather than implicitly as part of a desk review.

*Figure 2: Overall ratings of evaluation report quality 2013-2016*
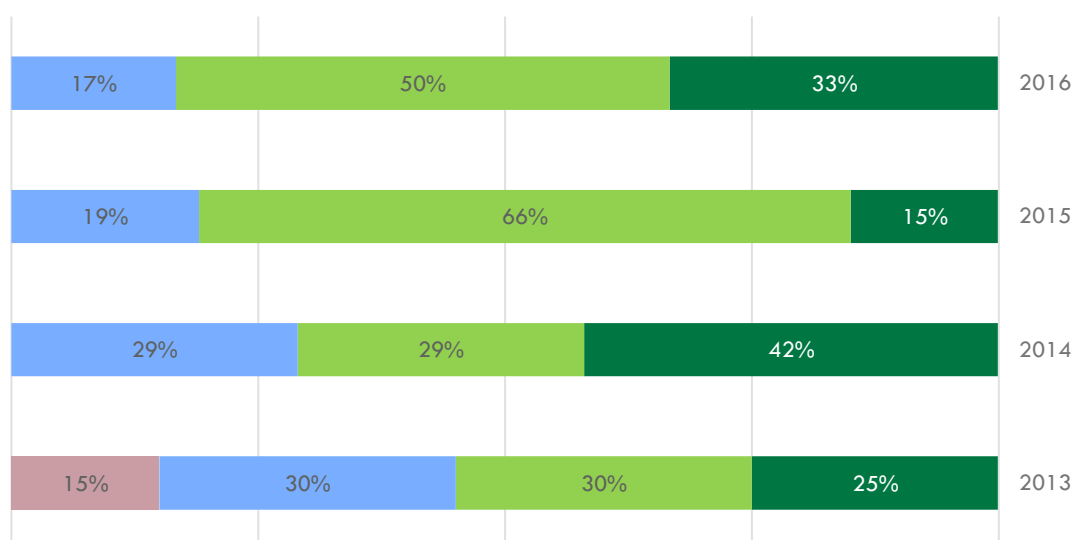


**Finding 3: Evaluations are becoming more strategic**

Project and programme evaluations continue to dominate the overall portfolio, although 2016 included 2 corporate, 3 global and 2 regional evaluations. There were also 2 Country Portfolio Evaluations. Of these strategic evaluations, 80% were rated Very Good, and 20% rated Good.

The portfolio of project and programme evaluations was also found to be more strategic than in previous years. No evaluation in 2016 was limited to the output level of analysis, with 64% attempting to measure outcome-level changes (the result of interventions by multiple actors) and the remaining 46% attempting to measure higher-level outcomes (the cumulative effects on people's lives and human rights). Both levels of assessment included a similar stratification of report quality.

Evaluations have also started to cover multiple impact areas under the Strategic Plan 2014-2017. The following table illustrates the coverage by evaluations and the quality of that coverage. The most evaluated impact area was Women's Economic Empowerment. The least evaluated areas were the Normative Framework and Women Peace and Security. Reviewers noted that 72% of evaluations covered issues related to National Planning and Budgeting (NPB), although this was mostly indirect rather than of specific NPB programmes.

> **Higher-level outcome:** Final results of a programme or policy on the intended beneficiaries. Reflects the cumulative effect of donor supported programmes of cooperation and national policy initiatives.
>
> **Outcome:** Effects from one or more programmes being implemented by multiple actors (UN Women and others), where the cumulative effect of outputs elicits results beyond the control of any one agency or programme
>
> **Output:** Causal effects deriving directly from programme activities, and assumed to be completely under programme control

*Table 1: Coverage of UN Women Impact Areas in 2016 evaluation reports*

| Impact Area | | Coverage in evaluations | Coverage rated Very Good or Good |
|---|---|---|---|
| 1 | Women's Political Participation | 64% | 87% |
| 2 | Women's Economic Empowerment | 69% | 92% |
| 3 | Ending Violence Against Women | 64% | 83% |
| 4 | Women, Peace and Security | 58% | 81% |
| 5 | National Planning and Budgeting | 72% (indirect) | 85% |
| 6 | Normative Framework | 53% | 84% |

### Finding 4: Evaluations meet UN SWAP standards for gender responsiveness

In late 2015, the Independent Evaluation Office published guidance on managing gender responsive evaluations designed to improve the performance of UN evaluation in the achievement of the UN SWAP Evaluation Performance Indicator (EPI). In 2016, an independent assessment of UN SWAP for UNEG found that the application of the EPI criteria by UN Women was consistent with the UNEG standards: something that was not the case for all entities.

The 2016 meta evaluation found a significant increase in performance of evaluation reports with regard to the UN SWAP EPI, with the highest recorded performance for UN Women since tracking of this indicator began.

Overall, UN Women evaluations were found to meet the UNEG requirements, with an average performance rating of 8.31 (out of a maximum score of 12). This is a substantive improvement over the 2015 average of 6.6 and the 2014 average of 7.5. This means that UN Women has met the agreed deadline for reaching UN SWAP standards.

The meta evaluation observes a broad correlation between overall quality of an evaluation report and performance regarding UN SWAP: as might be reasonably expected. The area with greatest opportunity for improvement – and the only criterion to rate on average as 'approaching requirements' – is the methods used by evaluations. These can be enhanced to be more gender responsive. On average, all other criteria now meet requirements.

*Table 2: UN SWAP average scores for 2016 evaluation reports*

| Scope | Questions | Methods | Analysis | Overall | Classification |
|---|---|---|---|---|---|
| 2.17/3 | 2.25/3 | 1.69/3 | 2.19/3 | 8.31/12 | Meets Requirements |

Three reports were rated as exceeding the UN SWAP standards, including both the corporate and the ECA regional evaluations of UN Women's contribution to UN system coordination, as well as the Final Evaluation of the Building a Constituency for Peace Project in Kyrgyzstan. However, other evaluations included innovations and excellence in particular aspects of the integrated of gender equality and human rights. For example, the Corporate Evaluation of the Regional Architecture of UN Women is particularly strong at describing the level of stakeholder engagement in the evaluation process, including the use of benchmarking to support the participatory formulation of the recommendations.
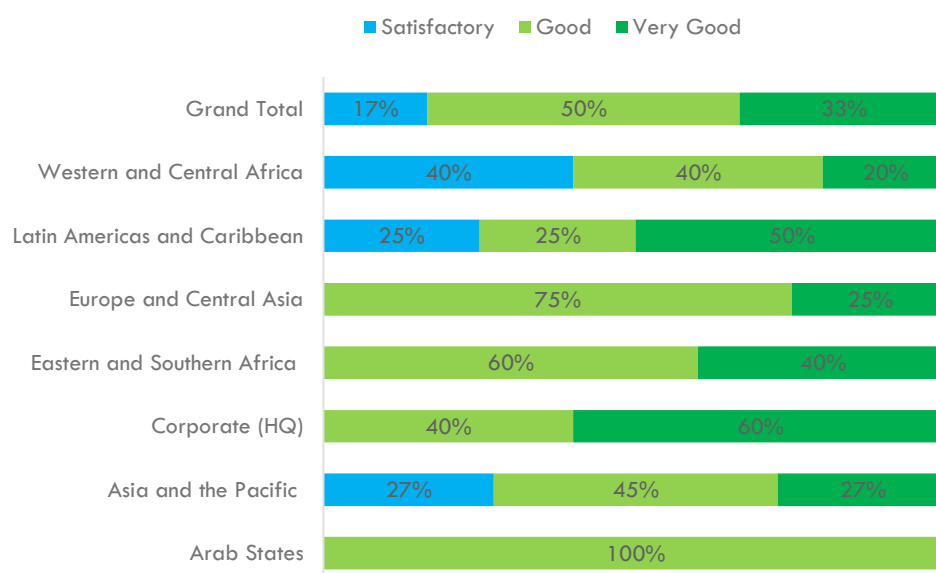
**Finding 5: Quality at the decentralised levels continues to strengthen**

As per previous years, centralised (corporate evaluations managed by the Independent Evaluation Office) were rated as Very Good. Decentralised global-level evaluations (those undertaken by HQ units, including multi-country evaluations) were all rated Good (4) or Very Good (2).

Within the regional architecture, decentralised evaluations have increased in number over time to reach 28 in 2016: 6 regional-coverage evaluations, 13 with national-coverage and 9 evaluations with coverage of a specific project area. The stratification of report quality is similar across all three of these levels of coverage.

Of the six UN Women regions, half (3) produced evaluations that completely meet the UNEG/UN Women standards (Europe and Central Asia; Eastern and Southern Africa; and Arab States). For each of the other three regions, a majority of reports are also rated as Good and Very Good: with overall improvements in Latin America and the Caribbean, and West and Central Africa.

*Figure 3: Regional disaggregation of overall GERAAS quality ratings for 2016 evaluations*

Legend: ■ Satisfactory  ■ Good  ■ Very Good

| Region | Satisfactory | Good | Very Good |
|---|---|---|---|
| Grand Total | 17% | 50% | 33% |
| Western and Central Africa | 40% | 40% | 20% |
| Latin Americas and Caribbean | 25% | 25% | 50% |
| Europe and Central Asia | | 75% | 25% |
| Eastern and Southern Africa | | 60% | 40% |
| Corporate (HQ) | | 40% | 60% |
| Asia and the Pacific | 27% | 45% | 27% |
| Arab States | | 100% | |

**Finding 6: Evaluation teams and designs remain relatively uniform**

The average team size for most evaluations (non-global decentralised evaluations) during 2016 was just over two people, most typically comprising of an international female and a national female evaluator. There was a 40% occurrence of a decentralised evaluation having an international and/or national male evaluator. Global evaluations included an average of nearly 6 team members. These comprised approximately 3 international female evaluators, and 1 each of international male, national female, and national male evaluators.

The vast majority of evaluations are managed by UN Women staff, with only 2 jointly managed evaluations include within the review (both rated good). There were no country-led evaluations.
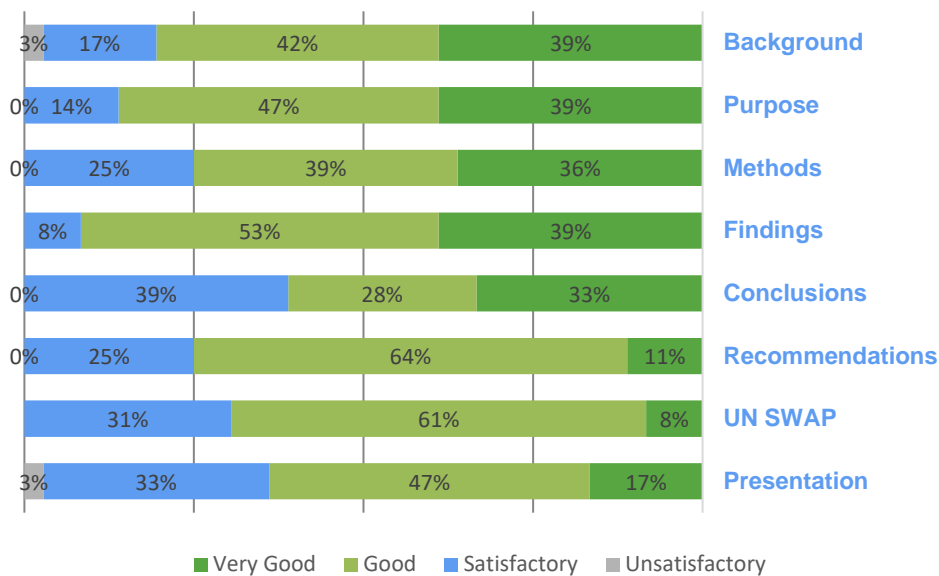
In terms of design, the reviewers' observed that most of the evaluations are primarily qualitative in nature (even where they state that they use mixed methods). Few evaluations undertake a survey or any other method to collect first-hand quantitative data. Some reports have cited a

small evaluation scope and limited evaluation funding as the reason for this. Even though the ToRs often explicitly call for quantitative data, many evaluations are not currently collecting this.

This observation of relative homogeneity in UN Women evaluation approaches was also recognised in previous meta-evaluations, and is also a feature highlighted in meta evaluations of some other UN entities. Whilst it is a limitation, it also reflects the intersection of competing demands: the requirement for gender responsive methods, the level of resourcing available for evaluations, and the nature of the interventions being evaluated.

## 6.2 Trends by Quality Assessment Parameters
*Figure 4: Disaggregation by GERAAS quality parameter of ratings for 2016 evaluations*



| | Very Good | Good | Satisfactory | Unsatisfactory |
|---|---|---|---|---|
| Background | 39% | 42% | 17% | 3% |
| Purpose | 39% | 47% | 14% | 0% |
| Methods | 36% | 39% | 25% | 0% |
| Findings | 39% | 53% | 8% | |
| Conclusions | 33% | 28% | 39% | 0% |
| Recommendations | 11% | 64% | 25% | 0% |
| UN SWAP | 8% | 61% | 31% | |
| Presentation | 17% | 47% | 33% | 3% |

**Parameter 1: Object and Context of the Evaluation**
The 2016 portfolio of evaluation reports is strong regarding the context and background section. In particular, reports consistently present a comprehensive contextual analysis of the intervention (social, cultural, political and economic background) alongside a full presentation of the implementation status of the evaluation object. Stakeholder analysis and a description of the object of the evaluation itself is also consistently strong. All of these dimensions meet UNEG/UN Women standards in over 90% of reports.

The main area of challenge for reports is the presentation of the logic model for interventions. 69% of reports do a Good or Very Good job of presenting the documented or reconstructed theories of change for interventions. However, the remaining 31% either miss an explicit presentation of the logic model, or only include a brief reference to the relevant results framework without discussing the relevance of this to the evaluation purpose. Given the blanket use of theory-based evaluation designs[8], this aspect of reports needs to be strengthened.

Examples of Very Good reports include:
1. **The WCAR Thematic Evaluation on Women's Leadership and Political Participation**, which covers the international and national norms guiding WPP, the current situation of

---

[8] An alternative evaluation design is "goal-free evaluation", but no report in the GERAAS system has intentionally applied this approach.

WPP in WCAR, and the key factors inhibiting full WPP. The current programme's theory of change is thoroughly examined.

2. **The Final Evaluation of the Knowledge Gateway on Women's Economic Empowerment Project (Empower Women),** which includes information on the reach of the knowledge gateway, the key beneficiaries and stakeholders, and the budget. The evaluation reconstructed a theory of change (ToC) in detail since the original ToC had evolved considerably since it was first established.

3. **The Evaluation of UN Women-Sida Strategic Partnership Framework 2011–16**, which includes thorough information on the context in which the object of evaluation operated and provides a good understanding of the environment in which the SPF was implemented. The scale, complexity and importance of the SPF within UN Women is described in detail.

## Parameter 2: Purpose, Objectives and Scope

This parameter of the report is largely controlled by the Terms of Reference, and is a strong aspect of reports overall, with at least 90% of reports rating as Good or Very Good for each indicator. The purpose (that change that the evaluation seeks to contribute to) is the strongest indicator, with no report rating unsatisfactory. This is a positive observation in terms of utility of evaluation. The main area for strengthening is to ensure that human rights is consistently mainstreamed into the evaluation framework for all reports.

Examples of Very Good reports include:
1. The **Evaluación Final del Programa Ampliando las Oportunidades Económicas para las Mujeres Rurales Emprendedoras en América Latina - Programa BEO,** which included additional and particularly detailed work explaining the objectives of the evaluation and the evaluation questions in the annexes. The report also explains in detail how gender and human rights issues were included in the objectives and scope of the evaluation.

2. The **Rapport Final d'Evaluation du Programme Global «Femmes, Paix et Sécurité» - Mali,** which described in detail the temporal and geographic scope, the specific objectives of the evaluation, and how the evaluation has addressed the gender and human rights dimensions.

3. The **Final Evaluation of the Global Programme on 'Increasing Accountability in Financing for Gender Equality',** which is particularly strong at describing who will use the evaluation information and for what purposes.

## Parameter 3: Methodology

Evaluation reports are strengthening in terms of describing the methodological approach. There is now consistently good quality description of the main data sources and methods to ensure data quality. Overall, no report was Unsatisfactory with regard to the methods section, and 75% were Good or Very Good. 83% of 2016 evaluation reports fully described a gender responsive method, which is a strong underpinning of a high quality UN SWAP rating.

The main area for greater consistency is with regard to the presentation of data collection methods, whilst the only area that had a substantive (14%) of reports rated as Unsatisfactory was the description of ethics considerations – although even this represents a positive trend over time.

Examples of Very Good reports include:
1. The **Final Evaluation of the Project "Strengthening Implementation of the Women, Peace and Security Agenda in Nepal (SIWPSAN)",** which clearly described data

collection and involved quantitative and qualitative methods. The evaluation describes the way each phase was developed to include Human Rights (HR) and gender responsive features. Similarly, the evaluation thoroughly discusses the ethical safeguards used in the evaluation process, citing the UNEG guidelines.

2. The **Zimbabwe Mid Term Evaluation of the Joint Programme on Prevention of Gender Based Violence (JPGBV) Against Young Women and Adolescent Girls,** which uses methods that are appropriate for analysing gender and human-rights considerations. The data sources, the sampling frame as well as the rationale for the sample, and the evaluation design's ethical safeguards are described in detail.

3. The **LACR Evaluación Regional de Acceso a la Justicia como mecanismo de prevención para acabar con las violencias contra las mujeres,** which describes with great clarity the methodology used and presents a systematic justification for the decisions that were made regarding the methodology. The evaluation clearly describes the triangulation of the information obtained and explains the limitations of the methodology used, accompanying them with mitigation strategies used in each case.

## Parameter 4: Findings

In general, most reports have findings sections that systematically analyse the evidence to address the evaluation criteria and respond to the evaluation questions. A full 97% of reports were rated as Good or Very Good in terms of including robust evidence-based findings. This is a positive observation in terms of the credibility of evaluations and their use for decision making.

There are two main areas where continued strengthening is required. The first is to ensure more consistent discussion around the gaps and limitations in the data and evidence. Only 61% of reports rate Good or Very Good for this indicator, a key aspect of transparency. The other area for strengthening is to more consistently discuss unexpected outcomes of interventions – both positive and negative in order to support learning.

The analysis of the findings section may benefit in the future from disaggregated rating of the quality of the evaluation criteria, since observation comments suggest that efficiency and sustainability are not being addressed to the same level of quality as effectiveness and relevance. More specific and comparable data on this would therefore be useful.

Examples of Very Good reports include:
1. The evaluation of **Strengthening a Participatory, Evidenced Based Formulation of a Comprehensive Action Plan to End Violence Against Girls in Cambodia,** in which all findings reflect a systematic and appropriate analysis and interpretation of the presented data, using a structured ranking system. Gaps and limitations are addressed and the reasons for accomplishments and failures are identified.

2. The **Moldova Final Evaluation of the Joint Integrated Local Development Programme (JILDP),** in which findings are clearly stated and are presented around the evaluation criteria. Evaluation questions are identified for each criterion to help guide the reader. The report is particularly strong in discussing factors supporting or hindering programme success.

3. The **Sénégal Evaluation finale du Projet de «Renforcement des capacités des femmes parlementaires pour une application effective des engagements sur l'égalité des sexes»,** which presents findings that demonstrate systematic analysis and interpretation of the data. A table is presented summarizing the findings according to the evaluation criteria. The limitations of the data to meet the criteria are addressed as well as the reasons for successes and failures.

**Parameter 5: Conclusions and Lessons Learned**

Conclusions are mostly a strong aspect of reports, with 97% of evaluations being Good or Very Good in regard to having evidence-based conclusions that add additional insights to the findings. There is a tendency to dwell on the positive strengths of interventions, with 25% of reports having scope to present a more balanced view of strengths and weaknesses. In general, however, most evaluations (83%) help to identify solutions to the challenges identified in the findings.

By comparison, lessons learned sections are much weaker, with 47% of reports being of insufficient quality in this regard. Where lessons are included, many evaluations struggle to abstract them to a generalised level, presenting intervention-specific and detailed insights rather than explaining how these can be adapted to other contexts. As one of the weakest aspects of 2016 evaluation reports, lessons learned is a potential focus area for future guidance.

Examples of Very Good reports include:
1. The **Uganda Joint Programme on Peace Building and Enhancing Protection Systems, Gender Promotion Initiative End of Programme Evaluation,** in which conclusions provide solutions to issues that may be encountered and they address both the strengths and areas of improvement of the object of evaluation. Similarly, the report presents Lessons Learned which are for the most part correctly generalised.
2. The **Corporate evaluation of UN Women's contribution to UN system coordination on gender equality and the empowerment of women,** in which each conclusion specifies from which findings it was drawn from, therefore strengthening the links between conclusions and evidence. Lessons learned are exceptionally well stated and provide excellent information that can inform similar programming in other contexts.
3. The **Final Evaluation of the Knowledge Gateway on Women's Economic Empowerment Project (Empower Women),** in which conclusions provide useful insights directly related to the key evaluation questions, present both the strengths and weaknesses of the evaluated object, and provide insights into important problems or areas for improvement.

**Parameter 6: Recommendations**

Reports were found to consistently present relevant and evidence-based recommendations that are actionable. This is a key factor for the utility of evaluations. In terms of the way recommendations are written, there remains scope to improve the specific targeting and prioritisation of recommendations.

Very few evaluation reports describe the process of how recommendations were developed (only 36%). This is required for transparency; and should refer to the role of evaluation reference groups, or other participatory processes in ensuring the relevance and utility of recommendations.

Examples of Very Good reports include:
1. The **Ethiopia Country Portfolio Evaluation,** in which recommendations reflect a deep understanding of the commissioning organization and potential constraints. The report duly describes the process followed in developing the recommendations, including the level and type of consultation with stakeholders.
2. The **UN Women South Sudan Country Office Mid-Term Programme Evaluation,** in which the recommendations are relevant to the object and purpose of the evaluation. They are supported by evidence, identify the target group for each recommendation, and are appropriately prioritised.

3.  The **UN Women Evaluation of the Regional Architecture of UN Women,** which is exceptionally strong at describing the process followed in developing the recommendations (including stakeholder engagement), ensuring that the recommendations are relevant to the object of evaluation, and that they are actionable and applicable to the context in which the organization operates.

**Parameter 8: The Report Structure**

The substantive aspects of reports have improved in quality at a faster rate than the presentational aspects, which is positive; but ensuring good quality structure and presentation is also an important contributing factor to utility and credibility. The most prominent area for improving the documentation of evaluations is regarding the comprehensiveness and quality of the annexes, where 36% of reports were found to require additional material.

On a related note, in response to comments from Regional Evaluation Specialists, the meta evaluation undertook a basic comparative assessment between the overall ratings and the unweighted average of parameter ratings. This suggested that the presentation of evaluation reports has a degree of correlation with overall ratings: making the case for future versions of the GERAAS tool to introduce explicit weighting between parameters in order to ensure emphasis on the substantive aspects of evaluation quality.

# 7. Conclusions

These conclusions have been developed by the reviewers based on the evidence presented in the findings, and have drawn on UN Women, UNEG and UN SWAP standards for evaluation, evaluation reports and ethics in evaluation. The conclusions are reliant on feedback from the UN Women Evaluation Office for validation.

**Conclusion 1: The quality of evaluation reports in 2016 represents a multi-dimensional improvement that is the culmination of guidance and efforts to improve the credibility and utility of the evaluation function**

The four years of GERAAS meta evaluations 2013-2016 reveals a comprehensive improvement in the quality of evaluation reports. Whilst the overall quality ratings of evaluation reports have stabilised since 2015, this headline hides multiple improvements found within the GERAAS indicators. For example, ethics and description of methods are both aspects that have seen substantive improvements.

The meta evaluation notes that the continuous development of the evaluation function in UN Women is illustrated by the establishment of various systems and guidance (see table below). Whilst there is not evidence of direct causation, these interventions are correlated with subsequent improvements that are captured in GERAAS.

Over the course of the Strategic Plan 2014-2017, therefore the evaluation function can be said to have demonstrated continuous improvement in coverage, quality, and gender responsiveness. All of these factors contribute to enhancing the credibility and utility of evaluation in support of management, accountability and learning.

*Table 3: Correlation between evaluation function activities and quality as captured by GERAAS*

| Year | Intervention | Subsequent change |
|------|--------------|-------------------|
| **2013** | Establishment of GERAAS system | Unsatisfactory reports eradicated in 2014 |
| **2014** | RES in place, and Transform Magazine/comms strategy | Increasing levels and coverage of evaluation in 2016 |
| **2015** | Country Portfolio Evaluation approach piloted | Greater coverage of Strategic Plan impact areas in 2015 and 2016 |
| **2015** | Guidance on Managing Gender Responsive Evaluation | UN SWAP evaluation performance indicator met in 2016 |

**Conclusion 2: Scope remains to enhance the overall quality of decentralised evaluations regarding specific aspects of quality, but the number of these aspects is decreasing**

The diversity of decentralised evaluation is increasing, with more regional evaluations in 2016 and the continued roll-out of country portfolio evaluations. Whilst the proportion of Satisfactory evaluations (i.e. evaluations reports that need improvement to fully meet UN Women standards) remains similar to 2015, the geographical spread and the range of aspects affecting these reports are both decreasing.

As with 2015, the aspects of evaluation reports controlled by the Terms of Reference remain strong, suggesting that the key to advancing quality is ensuring the application of proper methods, analysis and reporting by evaluators once they are commissioned. Whilst an increase in the number of evaluations reduces the potential for the regional evaluation specialists to do this follow-up directly, there are some areas of concentration that may help to focus resources. These especially include: 1) the balance and comprehensiveness of conclusions, 2) the transparent

development of recommendations, and 3) the understanding of evaluators about UNEG requirements for lessons learned.

**Conclusion 3: The coverage and scope of UN Women evaluations is increasing; whilst the range of designs and approaches used remain fairly narrow**
The evaluation portfolio for 2016 represents a major increase in coverage of both countries and thematic areas. This may reflect several trends, including the programme cycle (nearing the end of the 2014-2017 Strategic Plan), and the uptick in regional, multi-country and country portfolio evaluations.

Whilst the increase in coverage of evaluations is welcome, it remains unevenly distributed with some countries (such as Nepal, Jordan, Georgia and India) covered by multiple evaluations each year, whilst others have only been covered by a single evaluation during the Strategic Plan. Whilst the decentralised nature of the UN Women evaluation function makes it a challenge to address this coverage directly, it does represent an important policy consideration for the next Strategic Plan period.

Similarly, the meta evaluation notes the continuing similarity of evaluation designs, and the constraints that are leading to this concentration of approaches. Nevertheless, the predominance of qualitative approaches and the limited generation of primary quantitative data to triangulate qualitative data (with the exception of online data such as that gathered for the evaluation of the Knowledge Gateway) potentially represents a vulnerability to the credibility of evaluations. This also represents an area for consideration in the evaluation policy.

**Conclusion 4: The GERAAS system has contributed to UN Women having met the requirements of the UN SWAP evaluation performance indicator before the target date of 2017**
The 2015 meta evaluation found that UN Women evaluations for that year were approaching requirements for UN SWAP (in 2014 they rated as having just met the requirements of the evaluation performance indicator). This finding led to an increased emphasis within the Independent Evaluation Office on ensuring that the requirements for GEEW in evaluation was fully integrated, including through guidance but also through the regional evaluation architecture. This effort is fully reflected in the significant improvements represented in the performance of 2016 evaluations; and represents the value of the accountability system established by the Independent Evaluation Office for improving the performance of evaluation.

# 8. Recommendations

As with the conclusions, the reviewers have developed these recommendations based on the findings and conclusions, rather than a participatory process. They can be expected, therefore, to be subject to further validation with the UN Women Independent Evaluation Office. In accordance with the constructive feedback given to evaluation managers, the meta-evaluation has given an indication of the urgency, potential impact, and effort required for each recommendation, to aid with prioritisation.

**Recommendation 1: The Independent Evaluation Office is advised to focus the next round of guidance on improving lessons learned, comprehensive conclusions, and transparent recommendations**

The main areas requiring improvement within the GERAAS indicators for the 2016 portfolio of evaluations relate to the extended analysis of evidence required for the conclusions, recommendations and lessons learned. IEO is recommended to provide guidance, advice, examples and visibility to improving the quality of this analysis by evaluators. In the case of recommendations, this may be through demanding greater transparency; but in the case of lessons learned a more fundamental alignment of understanding and expectations is required to meet the UNEG standards.

**Recommendation 2: Regional Evaluation Specialists are recommended to ensure that the gains in gender responsive evaluation to UN SWAP standards are maintained and extended in 2017**

The decentralised evaluation portfolio has demonstrated substantive and significant improvements in the integration of gender responsive evaluation frameworks, methods, and analysis. This is a major achievement, and should remain a priority to ensure that it is maintained. Furthermore, there is still scope to continue to improve the quality of evaluations regarding GEEW. Given the correlation between high quality methods sections and strong UN SWAP ratings for methods, focusing on the quality and application of gender responsive methods is likely to have mutual benefit for other GERAAS indicators.

**Recommendation 3: The Independent Evaluation Office should review and enhance the GERAAS tool for the Strategic Plan 2018-21 period to include explicit weighting between parameters, disaggregate analysis of evaluation criteria, and reflect the UNEG 2016 norms and standards**

The consistency of the GERAAS tool has provided a longitudinal set of data that has given important insights into trends in the evaluation function. However, there is also scope to revise and further strengthen the robustness of the tool, including taking into account the recent experiences of other UN entities in light of the wider discussion about system-wide consistency in evaluation. In concrete terms, the implicit equal weighting between evaluation parameters should be replaced with an explicit weighting system to enhance clarity for both reviewers and the evaluation specialists in UN Women. There is also an opportunity to reflect the most recent updates to the UNEG norms and standards (2016), and to consider how the tool might adapt to a diversity in types of evaluations under renewed commitments to national evaluation capacity and leadership

**Recommendation 4: UN Women is recommended to ensure continued support to enhancing the quality of the decentralised evaluation function, and to review the evaluation policy for 2018-2021**

The establishment and embedding of the regional evaluation architecture is strongly correlated in GERAAS with a period of substantive and substantial improvement in the quality and coverage of evaluation. This, in turn, enhances the independence, credibility and utility of the evaluation function – and represents a recognised source of value for money in terms of organisational accountability, learning, and evidence-based decision making. In the context of uncertain financial and political support for GEEW (and multilateral action), ensuring continued support to this capacity is vital to maintaining the performance of the evaluation function that is reflected in GERAAS trends.

At the same time, the meta evaluation has noted the uneven geographical distribution of evaluations and the conceptual concentration of evaluation designs. It is beyond the scope of the meta evaluation to provide detailed insights, evidence or recommendations on the aspects of evaluation policy that are contributing to these patterns. There is, however, a sufficiently strong case to recommend that the Independent Evaluation Office review the evaluation policy to ensure continued alignment with the new Strategic Plan 2018-2021.

# 9. Annexes

## 9.1 Terms of Reference

**Background**

UN-Women is dedicated to the achievement of gender equality and the empowerment of women. The mandate and functions of UN-Women call for the promotion of organizational and UN system accountability on gender equality through evaluation, strengthening evaluation capacities and learning from evaluation, and developing systems to measure the results and impact of UN-Women with its enhanced role at the country, regional and global levels.

The purpose and role of evaluation in UN Women is to contribute to learning on best ways to promote gender equality and women's empowerment, enhance UN Women's accountability, and inform decision-making. By providing evidence-based information, evaluation contributes to UN Women's role to generate knowledge on what works to advance gender equality and women's empowerment.

Given the decentralized nature of the organization, the majority of the evaluations supported by UN Women are managed at a decentralized level. On average, 30 evaluations get carried out by UN-Women world-wide each year. Therefore, UN-Women IEO is giving increased emphasis to strengthening support for decentralized evaluations.

To address the organizational demands for ensuring good quality and credible evaluations particularly at decentralized level, the IEO has designed a Global Evaluation Reports Assessment and Analysis System (GERAAS) driven by similar good practices enforced by other UN entities and consistent with the UNEG Norms and Standards. The system is serving as a key instrument to increase the application of sound approaches and methods to continuously improve the quality and credibility of evaluation methods and reports within UN-Women. As part of this process, the IEO assesses the quality of corporate and decentralized evaluations on a yearly basis and made available assessment findings to senior managers, programme units, and the UNW Executive Board.

In addition to the quality assessment of individual reports, the GERAAS system requires a meta-analysis of evaluations to capture the key insights from evaluation reports – rated satisfactory or above according to UN Women standards. This is required to develop constructive lessons for future systemic strengthening of programming, organizational effectiveness and the evaluation function. Whereas the meta-evaluation provides a rating of the quality of evaluation reports according to UN Women standards, meta-analysis synthesizes the key findings, conclusions, and recommendations of the body of evaluation reports that meet UN Women quality requirements.

The IEO is seeking to establish a long term agreement with a well-established firm to conduct a meta-evaluation and meta-analysis of final evaluation reports[9] on a yearly basis. The selected firm will review final evaluation reports (on average 30 final reports[10]), rate them against UNEG/UNWOMEN standards, write an executive feedback to be sent to the CO concerned, make analysis of trends, key weaknesses and strengths of UN-Women managed evaluation reports and produce a meta-analysis report by synthesizing the recurrent findings, recommendations, conclusions, and lessons learned of all evaluation reports completed in a given year.

**Purpose & Methodology**

In general, the GERAAS has four main objectives:

---

[9] Majority of evaluation reports are in English language but some are also available in Spanish, French and Portuguese.

[10] The number of reports varies from one year to another. The cost required will be adjusted accordingly.

1.      *Improve the quality and utility of evaluation reports*: improve the use of evaluation reports by providing an objective assessment of the overall quality of the evaluation reports to Senior Managers and the Executive Board;

2.      *Strengthen internal capacity on gender responsive evaluation*: promote sound evaluation design and methodology as well as consistent and quality reporting through building internal capacity on managing and quality assuring evaluations;

3.      *Improve UN Women's performance and organizational effectiveness*: provide senior management with better understandings and insights into key UN women performance areas requiring attention; and

4.      *Promote learning and knowledge management*: help promote organizational learning and knowledge management through capturing experiences and lessons learned from credible evaluations.

**Key components of the consultancy**
The consultancy will have **two major components -** *Meta-Evaluation and Meta-Analysis*:

**Meta- evaluation** including assessment of the quality of individual evaluation report and provision of executive feedback to commissioning offices
**Meta Evaluation**
The Purpose of the meta-evaluation is to capture the quality of evaluation reports – according to UNEG-UNW standards. This is required to develop constructive lessons for future systemic strengthening of evaluation, and to allow possible trend analysis to examine changes in the quality and credibility of evaluations managed by the IEO and by all decentralized offices including HQ divisions. This meta-analysis summarizes key trends (by region, type, scope, results, thematic areas, stage, management etc), weaknesses and strengths, as well as lessons learned and good practices emerging from the review of the evaluation reports.

**Assessment of the quality of individual reports and scoring**
This comprises an assessment of the quality of individual evaluation reports against eight UN Women-adapted UNEG Parameters (Object and Context of Evaluation; Evaluation Purpose, Objectives and Scope; Evaluation Methodology; Findings; Conclusions and Lessons Learned; Recommendations; Gender and Human Rights Considerations; and the report structure).

**Executive Feedback to commissioning offices**
Inherent within the GERAAS is provision of executive feedback to commissioning offices about the quality of evaluation reports they managed. This is mainly designed to strengthen internal evaluation capacity by providing practical recommendations to improve future evaluations and to inform their own assessment of the performance of external consultants who might be hired for future evaluations.

**UN SWAP**
One of the expected deliverables of the assignment under the meta-evaluation is an individual evaluation report scoring using the UN SWAP Evaluation Performance Indicator. The UN SWAP Evaluation Scorecard is a reporting tool organized around 13 scoring criteria which are articulated around 3 headings that capture the overall elements related to mainstreaming gender equality throughout the evaluation process. It is a requirement for all UN entities to use the Scorecard to assess each evaluation report using the standard rating system for each criterion. This only requires quantitative scoring against set of established criteria and the tools has been integrated as part of the GERAAS methodology attached.

**3.1.1 Methodology for Meta-Evaluation**
The quality assessment uses the UNEG evaluation reports standards as a basis for review and assessment of final evaluation reports, while ensuring specific standards relevant to UN Women. The tools to be used for the quality assessment and scoring of the individual evaluation reports are annexed to this Terms of Reference.

The meta-analysis will consider only the final evaluation reports submitted to the UN Women Global Accountability and Tracking of Evaluation System – the GATE system. Only those reports classified as 'evaluation' will be subject to the meta-evaluation (rather than reviews, evaluability assessments, baselines, studies, etc).

**3.2: Meta-analysis**
The other main output of this consultancy is a synthesis of the evaluation reports. This requires analysis and synthesis of the findings, conclusions and recommendations including lessons learned and good practices presented in all evaluation reports rated 'Satisfactory' and above as part of the meta-evaluation. The synthesis of this information supports the use of evaluation findings by UN Women.

**Methodology for Meta-analysis**
The meta-synthesis aggregates the recurrent findings, conclusions, lessons learned, good practices and recommendations that have come out of evaluations every year. The meta-analysis is poised to provide a basis to better understand UN Women interventions around the UNEG criteria (relevance, effectiveness, efficiency, sustainability and impact) - whenever these criteria were covered by the evaluations and gender equality. Where possible, it also provides further analysis on the progress made against the six UN Women impact areas and the key operational effectiveness and efficiency priorities. The principles and approaches stipulated in the Strategic Plan such as capacity development; alignment with national development plans and strategies; inclusiveness; advocacy and knowledge generation; south-south and triangular cooperation; as well as lessons learned on aligning practice with normative guidance are also the basis to do the meta-analysis.

Only reports rated satisfactory and above by the quality assessment are used for meta-analysis. Unlike the qualitative assessment which assesses and produces separate report for each individual evaluation report, one synthesis report will be produced. This helps to paint a global perspective of UN Women interventions at different levels and facilitate better understanding and insight on what works to advance gender equality and women empowerment.

**Expected Deliverables**
The main expected deliverables of the exercise will be
**Meta-evaluation report** - summarizing key trends (by region, type, scope, results, thematic areas, stage, management etc), weaknesses and strengths, as well as lessons learned and good practices emerging from the review of the evaluation reports.
Assessment of the quality of individual reports and scoring using the tool to be provided
Executive Feedback to commissioning offices using the tool to be provided
Scoring against the UN SWAP defined scoring criteria using the tool to be provided
**Meta-Analysis** report– analysis and synthesizes the findings, conclusions, recommendations and lessons learned presented in the evaluation reports rated satisfactory and above

**Management of the consultancy**
The UN-Women Independent Evaluation Office is responsible for the management of the evaluation. The IEO will provide support to assure the quality of the draft and final products as well as facilitating administrative and other backstopping support.

The selected firm will assume sole responsibility in ensuring the consistency, quality and timely delivery of expected products, and overall coordination with UN-Women Independent Evaluation Office.

**Required Qualifications**
Excellent and proved knowledge of evaluation methodologies and approaches
Proven experience with meta-evaluation and meta-analysis of evaluation reports, preferably with UN agencies
Experience and background in gender equality/gender analysis and gender responsible evaluations
Proven practical professional experience in designing and conducting major evaluations

Excellent analytical and writing skills in English required. Working language of Spanish, French or Portuguese as asset

Familiarity with UNEG evaluation standards is an asset

Knowledge and expertise of other or similar quality assurance systems will also be an asset

**Duration of contract**

The is a yearly contract renewable up to four years upon satisfactory performance.

## 9.2 List and Ratings of Evaluation Reports Reviewed

| | Country | Evaluation title | UN SWAP | GERAAS |
|---|---|---|---|---|
| 1 | Bangladesh | Final evaluation of Gender and Climate Change project | Meets Requirements | Satisfactory |
| 2 | Cambodia | Final Evaluation of EVAW Programme | Meets Requirements | Very Good |
| 3 | China | MID-TERM EVALUATION OF THE CHINA GENDER FUND | Approaching Requirements | Good |
| 4 | Regional Office for Asia and the Pacific (Thailand) | Final Evaluation of Regional Programme on Improving Women's Human Rights in South East Asia – CEDAW SEAP | Meets Requirements | Good |
| 5 | Regional Office for Asia and the Pacific (Thailand) | Final evaluation of SDC funded part of Regional Programme on Empowering Women Migrant Workers in Asia (Phase III EWMWA 2012-2015) | Meets Requirements | Good |
| 6 | Regional Office for Asia and the Pacific (Thailand) | Endline Evaluation Report Leveraging Technical Tools, Evidence and Community Engagement to Advance the Implementation of Laws and Provision of Services to Women Experiencing Violence in South-East Asia | Approaching Requirements | Good |
| 7 | Pakistan | End of Term Evaluation of the Project 'Towards Gender Equality Women's Economic Empowerment Home Based Workers, Phase II' (2012-2015) | Meets Requirements | Very Good |
| 8 | Pakistan | Women Leadership and Social Reconstruction Program | Meets Requirements | Good |
| 9 | Multi-Country Office for the Pacific (Fiji) | Mid Term Evaluation of Pacific Regional EVAW Facility Fund | Meets Requirements | Satisfactory |
| 10 | Nepal | Final evaluation of the "Strengthening implementation of women, peace and security agenda in Nepal" (SIWPSAIN) project | Meets Requirements | Very Good |
| 11 | Nepal | Final evaluation of the "Localizing women, peace and security agenda in central Terai districts of Nepal" (GPI-CT) | Approaching Requirements | Satisfactory |

| 12 | Ethiopia | Country Portfolio Evaluation at end of Ethiopia SN 2014/2015-2016 | Meets Requirements | Very Good |
| 13 | Tanzania | Country Portfolio Evaluation | Meets Requirements | Good |
| 14 | South Sudan | Mid-Term Evaluation of UNW South Sudan Strategic Note 2014 - 2016 | Meets Requirements | Very Good |
| 15 | Uganda | Peace Building and Enhancing Protection systems (Gender Promotion Initiative) Final Evaluation | Meets Requirements | Good |
| 16 | Zimbabwe | Mid Term Evaluation of the Joint Programme on Prevention of Gender Based Violence (JPGBV) Against Young Women and Adolescent Girls | Meets Requirements | Good |
| 17 | Guatemala | BEO Programme Evaluation | Meets Requirements | Very Good |
| 18 | Regional Office for Americas and the Caribbean (Panama) | Access to Justice as a prevention mechanism to EVAW Regional Evaluation | Meets Requirements | Very Good |
| 19 | Chile | Informe Final de Evaluación del Programa Fortalecimiento de Organizaciones de la Sociedad Civil que promueven la Igualdad de Género en Chile 2013-2016 ONU MUJERES | Approaching Requirements | Satisfactory |
| 20 | El Salvador | EVALUACIÓN FINAL DEL PROYECTO: MUJERES Y POLÍTICAS MUNICIPALES A FAVOR DE LA IGUALDAD EN EL SALVADOR | Meets Requirements | Good |
| 21 | Regional Office for Europe and Central Asia (Turkey) | UN Women´s Contribution to UN Coordination on GEEW in ECA | Exceeds Requirements | Very Good |
| 22 | Moldova | Joint Integrated Local Development Programme Evaluation | Approaching Requirements | Good |
| 23 | Multi-Country Office for Central Asia (Kazakhstan) | Final Evaluation of the UN Women Project Empowering abandoned women from migrants families in Tajikistan | Meets Requirements | Good |
| 24 | Kyrgyzstan (CO) | Evaluation of PBF project "Building a Constituency for Peace" | Exceeds Requirements | Good |

| 25 | Regional Office for Arab States (Egypt) | Thematic Evaluation on Women's Political Participation | Meets Requirements | Good |
|---|---|---|---|---|
| 26 | Jordan | Promoting Social Cohesion through Women's Economic Empowerment and Protection Initiatives in Irbid and Zarqa | Meets Requirements | Good |
| 27 | Senegal | Final evaluation of the project « d'appui à la promotion des droits des femmes et des filles (PAPDFF) » [Support to the promotion of women's and girl's rights] with the Ministry of Justice | Approaching Requirements | Satisfactory |
| 28 | WCARO | Thematic Evaluation on Women's Leadership and Political Participation (2011-2015) | Approaching Requirements | Satisfactory |
| 29 | Senegal | Evaluation finale du Projet de « Renforcement des capacités des femmes parlementaires pour une application effective des engagements sur l'égalité des sexes » | Meets Requirements | Very Good |
| 30 | Mali | Women Leadership and Participation in Peace Security and Humanitarian Action Final Evaluation | Approaching Requirements | Good |
| 31 | Mali | Support to women/girls affected by conflict and participation of women in the process of consolidation of peace Midterm Evaluation | Approaching Requirements | Good |
| 32 | Corporate evaluation on coordination | UN Women contribution to the United Nations system coordination | Exceeds Requirements | Very Good |
| 33 | Corporate evaluation on Regional Architecture | Corporate evaluation on Regional Architecture | Meets Requirements | Very Good |
| 34 | Policy Division | Final Evaluation of the Knowledge Gateway on Women's Economic Empowerment Project (Empower Women) | Meets Requirements | Very Good |
| 35 | Policy Division | Financing for Gender Equality | Approaching Requirements | Good |
| 36 | Programme Division | UN Women Sida Strategic Partnership Framework 2011-2016 | Approaching Requirements | Good |

## 9.3 Report Review Format

| PART I: BASIC INFORMATION AND CLASSIFICATION OF THE EVALUATION | | | |
|---|---|---|---|
| | **Response** | | |
| **Title of the Evaluation Report** | | | |
| **Report sequence number** | | **Year of the Evaluation Report** | 2013 |
| **Region** | | **Country(is)** | |
| **Programme Implementation Period** | | **TORs Present** | |
| **Duration of the Evaluation** | | | |
| **Project/Programme Budget** | | | |
| **Evaluation Budget** | | | |

| Classification of Evaluation Report | | Comments |
|---|---|---|
| **Geographical** (*Coverage of the programme being evaluated & generalizability of evaluation findings*) | | |
| **Management of Evaluation**(*Managerial control and oversight of evaluation decisions*) | | |
| **Type of intervention evaluated** | | |
| **Result** (*Level of changes sought, as defined in results framework refer to substantial use of highest level reached*) | | |
| **UN Women Strategic Plan Correspondence** | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| **Stage of Evaluation** | | |

| PART II: THE EIGHT KEY PARAMETERS | | | | |
|---|---|---|---|---|
| **PARAMETER 1: OBJECT AND CONTEXT OF THE EVALUATION** | | | **Guiding Question** | **Constructive feedback for future reports** *Including how to address weaknesses and maintaining good practice* |
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | Does the report present a clear & full description of the 'object' of the evaluation? | |
| **Object and Context of the Evaluation** | | | | |
| 1.1 The **logic model and/or the expected results chain** (inputs, outputs and outcomes) of the object is clearly described. | | | | |
| 1.2 The **context** includes factors that have a direct bearing on the object of the evaluation: social, political, economic, demographic, institutional. This also includes explanation of the contextual gender equality and human rights issues, roles, attitudes and relations. | | | | |
| 1.3 **The scale and complexity of the object of the evaluation are clearly described** (the number of components, the geographic context and boundaries, the purpose, goal and organization/management of the object and the total resources from all sources including humans and budgets). | | | | |
| 1.4 The **key stakeholders** involved in the object implementation, including the implementing agency(s) and partners, other stakeholders and their roles. | | | | |
| 1.5 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. | | | | |
| **PARAMETER 2: PURPOSE, OBJECTIVES AND SCOPE** | | | **Guiding Question** | **Constructive feedback for future reports** Including how to address weaknesses and maintaining good practice |
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation? | |
| **Purpose, objectives and scope** | | | | |
| 2.1 **Purpose of evaluation:** is clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information is needed, how the information will be used. | | | | |
| 2.2 **Evaluation Objectives:** A clear explanation of the evaluation objectives including main evaluation questions is provided. | | | | |
| 2.3 **Evaluation Scope:** The scope of the evaluation is described including justification of what the evaluation covers and did not cover | | | | |
| 2.4 **Evaluation Criteria:** The report describes and provides an explanation of the chosen evaluation criteria, performance standards, or other criteria used by the evaluators. | | | | |
| 2.5 **Gender and Human Rights:** Evaluation objectives and scope include questions that address issues of gender and human rights. | | | | |
| **PARAMETER 3: METHODOLOGY** | | | **Guiding Question** | **Constructive feedback for future reports** Including how to address weaknesses and maintaining good practice |
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | Is the methodology used for the evaluation clearly described and is the rationale for the methodological choice justified? | |
| **Methodology** | | | | |

| GUIDING POINTS/EXPLANATORY NOTE | RATING | Remarks | Guiding Question | Constructive feedback for future reports |
|---|---|---|---|---|
| 3.1 **Methodology:** The report presents transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve evaluation purposes and objectives. | | | | |
| 3.2 **Data Collection:** The report describes the data collection methods and analysis, the rationale for selecting them, and their limitations. Reference indicators and benchmarks are included where relevant. | | | | |
| 3.3 **Data Sources:** The report describes the data sources, the rationale for their selection, and their limitations. 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 limits. | | | | |
| 3.4 **Sampling Frame:** The report describes the sampling frame – area and population to be represented, rationale for selection, mechanics of selection, numbers selected out of potential subjects, and limitations of the sample. | | | | |
| 3.5 **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. | | | | |
| 3.6 **Data Quality:** The report presents evidence that adequate measures were taken to ensure data quality, including evidence supporting the reliability and validity of data collection tools (e.g. interview protocols, observation tools, etc.) | | | | |
| 3.6 **Gender and Human Rights considerations:** The methods employed are appropriate for analyzing gender and rights issues identified in the evaluation scope. | | | | |
| 3.7 **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, privacy and confidentiality considerations. | | | | |

| PARAMETER 4: FINDINGS | | | Guiding Question | Constructive feedback for future reports |
|---|---|---|---|---|
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | **Are the findings clearly presented, relevant and based on evidence and sound analysis?** | Including how to address weaknesses and maintaining good practice |
| **Findings** | | | | |
| 4.1 Findings respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report and are based on evidence derived from data collection and analysis methods described in the methodology section of the report. | | | | |
| 4.2 Reported findings reflect systematic and appropriate analysis and interpretation of the data. | | | | |
| 4.3 Reported findings address the evaluation criteria (such as efficiency, effectiveness, sustainability, impact and relevance) and questions defined in the evaluation scope. | | | | |
| 4.4 Findings are objectively reported based on the evidence | | | | |
| 4.5 Gaps and limitations in the data and/or unanticipated findings are reported and discussed. | | | | |
| 4.6 Reasons for accomplishments and failures, especially continuing constraints, were identified as much as possible | | | | |

| PARAMETER 5: CONCLUSIONS AND LESSONS LEARNED | | | Guiding Question | Constructive feedback for future reports |
|---|---|---|---|---|
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | **Are the conclusions clearly presented based on findings and substantiated by evidence?** | Including how to address weaknesses and maintaining good practice |
| **Conclusion and Lessons Learned** | | | | |
| 5.1 Conclusions present reasonable judgments based on findings and substantiated by evidence, and provide insights pertinent to the object and purpose of the evaluation. | | | | |
| 5.2 The conclusions reflect reasonable evaluative judgments relating to key evaluation questions. | | | | |
| 5.3 Conclusions are well substantiated by the evidence presented and are logically connected to evaluation findings. | | | | |
| 5.4 Stated conclusions provide insights into the identification and/or solutions of important problems or issues pertinent to the prospective decisions and actions of evaluation users. | | | | |
| 5.5 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. | | | | |
| 5.6 Lessons Learned: When presented, lessons drawn represent contributions to general knowledge. They may refine or add to commonly accepted understanding, but should not be merely a repetition of common knowledge. Lessons presented suggest how they can be applied to different contexts and/or different sectors. | | | | |

| PARAMETER 6: RECOMMENDATIONS | | | Guiding Question | Constructive feedback for future reports |
|---|---|---|---|---|
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | **Are the recommendations relevant to the object and purpose of the evaluation and clearly** | Including how to address weaknesses and |
| **Recommendations** | | | | |

| | | | presented in a priority order? | maintaining good practice |
|---|---|---|---|---|
| 6.1 Recommendations are supported by evidence and conclusions, and were developed with the involvement of relevant stakeholders. | | | | |
| 6.2 The report describes the process followed in developing the recommendations including consultation with stakeholders. | | | | |
| 6.3 Recommendations are relevant to the object and purposes of the evaluation. | | | | |
| 6.4 Recommendations clearly identify the target group for each recommendation. | | | | |
| 6.5 Recommendations are clearly stated with priorities for action made clear. | | | | |
| 6.6 Recommendations are actionable and reflect an understanding of the commissioning organization and potential constraints to follow up. | | | | |

| **PARAMETER 7: GENDER AND HUMAN RIGHTS** | | | **Guiding Question** | **Constructive feedback for future reports** Including how to address weaknesses and maintaining good practice |
|---|---|---|---|---|
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | **Are gender and human rights perspectives integrated and well addressed in the process of the evaluation as well as in the evaluation report?** | |
| **GENDER AND HUMAN RIGHTS** | | | | |
| 7.1 GEEW is integrated in the evaluation **scope** of analysis and **indicators** are designed in a way that ensures GEEW-related data will be collected | | | | |
| 7.2 Evaluation **criteria and evaluation questions** specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved. | | | | |
| 7.3 A gender-responsive evaluation **methodology,** methods and tools, and data analysis techniques are selected. | | | | |
| 7.4 The evaluation findings, conclusions and recommendation reflect a **gender analysis** | | | | |

| **PARAMETER 8: THE REPORT STRUCTURE** | | | **Guiding Question** | **Constructive feedback for future reports** Including how to address weaknesses and maintaining good practice |
|---|---|---|---|---|
| **GUIDING POINTS/EXPLANATORY NOTE** | **RATING** | **Remarks** | **Is the report well structured, logical, clear and complete?** | |
| **THE REPORT STRUCTURE** | | | | |
| 8.1 **Report is logically structured** with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations). | | | | |
| 8.2 **The title page and opening pages** provide key basic information A. Name of the evaluation object B. Timeframe of the evaluation and date of the report C. Locations (country, region, etc.) of the evaluation object D. Names and/or organizations of evaluators E. Name of the organization commissioning the evaluation 6. Table of contents which also lists Tables, Graphs, Figures and Annexes G. List of acronyms. | | | | |
| 8.3 The Executive Summary is a stand-alone section that includes A. Overview of the evaluation object B. Evaluation objectives and intended audience C. Evaluation methodology D. Most important findings and conclusions E. Main recommendations | | | | |
| 8.4 **Annexes** increase the credibility of the evaluation report. They may include, inter alia: A. TORs B. List of persons interviewed and sites visited. C. List of documents consulted D. More details on the methodology, such as data collection instruments, including details of their reliability and validity E. Evaluators biodata and/or justification of team composition F. Evaluation matrix G. results framework | | | | |
| **Additional Information** | | | | |
| Assess the extent to which the evaluation successfully addresses the ***Terms of Reference:*** If the report does not include a ToR then a recommendation should be given to ensure that all evaluations include the ToR in the future. Some evaluations may be flawed because the TORs are inappropriate, too little time etc. Or, they may succeed despite inadequate TORs. This should be highlighted. | | | | |
| Identify aspects of ***good practice*** of the evaluation | | | | |

| **PART III: THE OVERALL RATING** | | | | |
|---|---|---|---|---|
| The overall rating should be given largely based on the assessment given against the eight key parameters and the guiding points under PART II. Following are some of the key guiding questions to inform the overall rating which posits the degree to which the report could generally be used with confidence. | | | | |
| **Key Guiding Questions** | **Rating** | **Remarks** | **Overall rating** | **Is this a credible report that addresses the evaluation purpose and objectives based on evidence, and that can therefore be used with confidence?** |
| **The extent to which each of the eight parameters of the evaluation, taken on their own, provide sufficient credibility so that they be used with confidence.** | | | | |
| **The extent to which the eight parameters of the evaluation hold together in a logically consistent way that allows the confidence to act.** | | | | |
| | **ToRs** | | | |

X

| | | | | |
|---|---|---|---|---|
| **Describe any reason(s)  that might explain the overall performance or particular aspects of this evaluation report.** This is a chance to note mitigating factors and/or crucial issues apparent in the review of the report. | **Other** | | | |

## 9.4 GERAAS Method

IEO undertook an initial screening of all reports uploaded in the GATE system to a) decide whether the report can be classified as an evaluation as per UNEG definition and b) to ensure that the evaluation is managed or jointly managed by UN Women. The independent assessor undertook a secondary screening.

Included within GERAAS is the provision of executive feedback to commissioning offices about the quality of evaluation reports they managed.  This is mainly designed to strengthen internal evaluation capacity by providing practical recommendations to improve future evaluations and to inform their own assessment of the performance of external consultants who might be hired for future evaluations.

### Revew of Evaluation Reports

The full review-process is illustrated in Figure 1 (see below). An evaluation report is assessed as 'good quality' when it is a credible report that addresses the evaluation purpose and objectives based on evidence, and therefore **can be used with confidence.**

That is to say, a *good* or *very good* evaluation report will provide a clear and complete assessment of the object of the evaluation, based on evidence compiled and analyzed in accordance with UN Women-adapted UNEG standards, such that its conclusions and recommendations can be deemed to be credible and are thus a sound basis for decision-making.

A *satisfactory* report is a report that almost meets requirement with regard to quality but some elements are missing or inadequately addressed. The report has useful information that can be used with confidence. *Unsatisfactory* report do not yet meet multiple or critical standards.

Evaluation reports are reviewed using the UN Women-adapted UNEG Evaluation report standards Matrix to assess the following core elements:

| Clear and full description of the 'object' of the evaluation | The report describes the object of the evaluation including the results chain, meaning the 'theory of change' that underlies the programme being evaluated. This theory of change includes what the programme was meant to achieve and the pathway (chain of results) through which it was expected to achieve this. |
|---|---|
| | The context of key social, political, economic, demographic, and institutional factors that have a direct bearing on the object is described. For example, the partner government's strategies and priorities, international, regional or country development goals, strategies and frameworks, the concerned agency's corporate goals and priorities, as appropriate. |
| The evaluation's purpose, objectives and scope are fully explained | The purpose of the evaluation is clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information is needed, how the information will be used. The report provides a clear explanation of the evaluation objectives and scope including main evaluation questions and describes and justifies what the evaluation did and did not cover. The report describes and provides an explanation of the chosen evaluation criteria, and/or other criteria used by the evaluators. |
| Appropriate and sound methodology | The report presents transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve evaluation purposes. |
| | The report presents a sufficiently detailed description of methodology in which methodological choices are made explicit and justified and in which limitations of methodology applied are included. The report gives the elements to assess the appropriateness of the methodology. Methods as such are not 'good' or 'bad', they are only so in relation to what one tries to get to know as part of an evaluation. Thus this |

| | standard assesses the suitability of the methods selected for the specifics of the evaluation concerned, assessing if the methodology is suitable to the subject matter and the information collected are sufficient to meet the evaluation objectives. |
|---|---|
| Findings, conclusions, recommendations and lessons learned are based on evidence and sound analysis | *Findings* respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report. They are based on evidence derived from data collection and analysis methods described in the methodology section of the report.<br><br>*Conclusions* present reasonable judgments based on findings and substantiated by evidence, providing insights pertinent to the object and purpose of the evaluation.<br><br>*Recommendations* are relevant to the object and purpose of the evaluation, are supported by evidence and conclusions, and were developed with involvement of relevant stakeholders. Recommendations clearly identify the target group for each recommendation, are clearly stated with priorities for action, are actionable and reflect an understanding of the commissioning organization and potential constraints to follow up.<br><br>*Lessons learned* are grounded in the evidence arising from the evaluation, but provide insights that are relevant beyond the specific scope of the projects, programs or policies evaluated. Lessons learned highlight elements of preparation, planning, design or implementation that can be expected to have positive or negative effects on performance, outcome, or impact. |
| Gender and human rights perspectives integrated and well addressed | The report illustrates the extent to which the design and implementation of the object, the assessment of results and the evaluation process incorporate a gender equality perspective and human rights based approach. Gender sensitive and human rights-based language is used throughout, and data collection and analysis methods are gender equality and human rights responsive. |
| Well structured, logical and clear report | The report is logically structured with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations). It reads well and is focused. |

**Meta-evaluation**

The review process consists of five main parts. There were three reviewers involved in the process in order to ensure that a person fluent in each language assessed relevant reports. Consistency was ensured through a) a detailed briefing, b) using quality assurance by the main reviewer to quality assure consistency, and c) responding to comments/challenges by the Independent Evaluation Office.

To avoid real or perceived conflict of interests, the second and third reviewers were independent of ImpactReady and were commissioned to undertake all reviews overseen by Regional or Country Offices in which ImpactReady is working.

**PART 1: BASIC INFORMATION AND CLASSIFICATION OF THE EVALUATION**

The Evaluation Assessment Matrix is a simple tool designed to capture or provide a snap shot of the key aspects of the evaluation and the evaluation report. This comprises basic information such as title, region/country, type, costs, geographic and thematic coverage, stage/timing and management of the evaluation.

**PART 2: ASSESSMENT OF THE QUALITY AGAINST EACH OF THE EIGHT 'PARAMETERS'**

The final review template is composed of *8 Parameters* (Object and Context of Evaluation; Evaluation Purpose, Objectives and Scope; Evaluation Methodology; Findings; Conclusions and Lessons Learned; Recommendations; Gender and Human Rights Considerations; and the report structure). The Eight Parameters are further defined by 39 Guiding Points.

Qualitative and rated feedback on the Eight UN Women-adapted UNEG *Parameters* are considered and provided independently. The assessment follows guiding points that are designed to inform a qualitative story on the level of each of the eight parameters to be reviewed, noting any points that will subsequently inform the reviewer's reflection on areas for future improvement in evaluation practice (to be captured as part of the 'Overall Rating' step for each report.

*Figure 1: Meta-evaluation process*

## Filtering by EO

| | |
|---|---|
| IEO checks that all submitted reports are evaluations | 0 reports removed from the review as considered outside the scope of 2014 |

## Filtering by Independent Reviewer

| | |
|---|---|
| Reviewer checks that all submitted reports are evaluations | No reports removed |

## First Review

| | |
|---|---|
| Reports read and rated according to the review format (see annexes) | Reviews rated by a person fluent in the language of the report |

## Second Review

| |
|---|
| Reports assessed by main reviewer (via machine translation if required) and reviews updated if considered appropriate |

## Feedback

| | |
|---|---|
| Independant Evaluation Office offered 'right-to-respond' to specific reviews with comments | When reviews challenged, written feedback provided and reviews update if considered appropriate |

## Meta-evaluation draft report

| | |
|---|---|
| Quantitative analysis of report ratings using Excel | Qualitative analysis of review comments matched to quantitative results |

## Comments

| | |
|---|---|
| Two rounds of comments from Independant Evaluation Office and UN Women stakeholders | Comments considered and final report edited to reflect these |

## Meta-analysis draft report

| | |
|---|---|
| Qualitative assessment of lessons learned from reports meeting UN Women standards | See separate meta-analysis report. |

One key question for each of the eight parameters was answered to serve as a starting point for the reviewer to do the overall analysis on basis of the explanatory note provided for each parameter. Each parameter is also rated overall against a 4-point rating system. Clear explanatory descriptions are provided to guide and ensure consistency in the rating of each parameter.

| Very good |
| Good |
| Satisfactory |
| Unsatisfactory |

Based on the overall rating for the Parameter and the analysis of guiding points, feedback has been provided for the commissioning office on how to improve future evaluation reports. This includes ways to address weaknesses and to maintain good practice identified.

Focused feedback on the assessment of each Parameter has been formatted for Senior Management, including strengthens and weaknesses and followed by justification for the rating.

## PART 3: ASSESSMENT AND RATING OF THE OVERALL QUALITY OF THE ENTIRE INDIVIDUAL EVALUATION REPORT

The overall rating or final judgment on the quality of the evaluation report has been largely informed by the assessment provided against the eight key parameters. Guiding points were also provided to inform the overall rating.

The overall rating and the assessment of the strengths and weaknesses of the evaluation report gives an indication of the relative reliability of the results and determines **the extent to which the report can be used with confidence** to feed into future programming and to serve other purposes.  Accordingly, the reviewer has provided an overall rating for the report making use of the 4 point rating system.

| | |
|---|---|
| **Very good:** | A 'very good quality' evaluation report is a report that has the features of being credible, addressing the evaluation questions, based on evidence, and, adheres to UNEG adapted UN Women Evaluation Report Standards. The report can be used with confidence and is considered a good example. |
| **Good:** | The report adheres to UNEG/UN Women evaluation standards, good analysis and credible recommendations. The report can be used with confidence. |
| **Satisfactory:** | The report meets requirements with regard to quality but some elements are missing or inadequately addressed. The report has useful information. |
| **Unsatisfactory:** | Reports rated unsatisfactory entail serious limitations and hence caution should be exercised when using the findings or recommendations for learning, accountability, evidence generation or informed decision making. |

To ensure consistency in the rating of each parameter and the overall report, the reviewer undertook the review based on what is written in the evaluation report.  In line with GERAAS specifications, methodology, findings, conclusions and recommendations were given more

prominence to inform the overall rating. These are considered by UN Women to be the bedrock of a good quality report.

Qualitative feedback was provided regarding the coherence and credibility of the entire report using an Executive Review Template.

**PART 4: GENDER AND HUMAN RIGHTS CONSIDERATIONS**
The Independent Evaluation Office places considerable attention to ensure that evaluations managed by UN Women are Gender and Human Rights Responsive. This aspect has been treated in the GERAAS as a standalone Parameter for assessment of evaluation report as well as integrated in other parameters/guiding points where deemed appropriate.

**PART 5: META-EVALUATION**
This meta-analysis summarizes key trends (by region, type, scope, results, stage, management etc), weaknesses and strengths, as well as lessons learned and good practices emerging from the review of the evaluation reports.

**PART 5: META-ANALYSIS**
The other main output of the GERAAS is a synthesis of the evaluation reports. The same reviewer has analyzed and synthesized the findings, conclusions and recommendations including lessons learned and good practices presented in all evaluation reports rated Satisfactory and above. The synthesis of this information supports the use of evaluation findings by UN Women.

**Changes Made From Previous Years**
The main change from GERAAS 2015 is the provision of qualitative feedback in the original language of the report.

**UN SWAP**
ECOSOC Resolution 2007/331 requests the United Nations system, including United Nations agencies, funds and programmes within their organizational mandates, to strengthen institutional accountability mechanisms, including through a more effective monitoring and evaluation framework for gender mainstreaming based on common United Nations evaluation standards.

UN entities are expected to meet UN SWAP performance standards by 2017, with an extended timeframe to 2019 for those entities with a mainly technical focus. The ultimate goal is that all UN system entities "meet requirements" related to the Evaluation Performance Indicator. However, achieving this is only considered a starting point, and UN entities should continually strive to "exceed requirements" if the UN system is to truly benefit from gender responsive evaluation practice.

It is expected that the act of monitoring and reporting against this indicator will provide constructive momentum for reviewing progress made and reflecting on continuing challenges so as to improve performance over time, at both the level of the individual entity and the UN system.

The use of the UN SWAP Evaluation Scorecard provides a basis for harmonising the meta-reviews/evaluations conducted by different entities by assigning an overall aggregate score based on 4 UN SWAP Evaluation Performance Indicators:

| Gender mainstreaming | Score |
|---|---|
| **Fully integrated** | 3.0 |
| **Satisfactorily integrated** | 2.0 |

| | |
|---|---|
| **Partially integrated** | 1.0 |
| **Not at all integrated** | 0.0 |

Each report was rated using the UN SWAP scorecard, which was integrated into the GERAAS format. In rare cases, some criteria of the scorecard may "not be applicable" and clear justification was provided for entities reporting as such.

Average Score for each evaluation and the overall meta-analysis was calculated based on the UN SWAP protocol:

| Rating | Range |
|---|---|
| **Exceeding Requirements** | 1.76 and above |
| **Meeting requirements** | 1.26 - 1.75 |
| **Approaching requirements** | 0.51 - 1.25 |
| **Missing requirements** | 0 - 0.50 points |