ImpactReady™

UN Women
GERAAS 2013

*Global Evaluation Reports
Assessment and Analysis System*

META EVALUATION REPORT

Version 1.1

18 March 2014

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

Prepared by ImpactReady LLP
Joseph Barnes,
in collaboration with IOD PARC

Contents

Tables and Figures

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

This meta-analysis assesses 27 evaluation reports from 2013 that were submitted to UN Women's Global Evaluation Reports Assessment and Analysis System (GERAAS). GERAAS is a new approach to rating evaluation reports using UN Women, UN Evaluation Group (UNEG) and UN System Wide Action Plan (SWAP) standards and indicators of report quality. The purpose of GERAAS is to develop constructive lessons for future systemic strengthening of the evaluation function, and provide a baseline for the future.

24 evaluation reports from across the UN Women regions and 3 corporate evaluation reports were included in the analysis, representing at least 29 countries (some reports were multi-country, regional, or corporate). Each report was assessed against 8 parameters, including a specific parameter on Gender Equality and Human Rights. These parameters were further disaggregated into 45 guiding points. In addition to ratings, justifications and constructive feedback was provided for each parameter in an Executive Review. Finally, each report was given an overall classification:

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 cannot be used with confidence for learning, accountability, evidence generation or informed decision making.

In addition to UNEG standards, reports were assessed using the UN SWAP scorecard. This provides the required data on gender-responsive evaluation under ECOSOC Resolution 2007/331 and the Quadrennial Comprehensive Policy Review Resolution (A/RES/67/226). 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 13 UN SWAP Evaluation Performance Indicators.

It is important to note that GERAAS assesses the quality of reports (as a standalone document) and not of the evaluation as a process. The ratings provided are thus indicative of overall evaluation performance, but are not comprehensive and do not represent the actual use of evaluations in enhancing performance, learning and accountability in UN Women. The cost of additional layers of analysis to fully understand utilization (such as follow-up interviews) is currently considered prohibitive.

Reviews were undertaken by an independent consultant with prior experience of evaluation report quality assurance for UN agencies and donors. Fluent Spanish, Portuguese and French speakers with prior experience in evaluation quality assurance provided additional support to the consultant with non-English reports. A process was established to ensure consistency of ratings. In addition, the UN Women Evaluation exercised its 'right-to-respond' to the ratings of 7 reports, resulting in written responses from the reviewers and changes to the ratings of two reports.

2.2 Findings

The review process found that 85% of evaluation reports could be considered as satisfactory or above according to UN Women standards: implying that decision makers could be reasonably confident in their findings, insights and recommendations. Whilst many of these reports were found to have opportunities for improvements, 26% were found to be exceptional – exceeding UN Women Standards in multiple parameters.

■ Unsatisfactory   ■ Satisfactory   ■ Good   ■ Very Good

| 14.81% | 29.63% | 29.63% | 25.93% |

The most evaluation reports (8) were submitted from the Eastern and Southern Africa region, the least (1) from Arab States. No region had more than a single report rated as Not Satisfactory. All regions included reports with parameters rated as 'Good'. The majority (44%) of evaluations were national level evaluations, with over 83% of these rating as Satisfactory or above.

Evaluations considered all of the Strategic Plan Impact Areas, with the greatest consideration given to Impact Area 3 – EVAW (13 reports) – and Impact Area 1 – participation and leadership (11 reports). All impact areas were considered in at least 6 reports.

UN Women directly managed at least 67% of the evaluation reports. Of the 26% of reports that were identified as being jointly managed, the majority related to Joint Programmes with other UN agencies. It was noticeable that the Very Good reports tended to be associated with direct management by UN Women.

Half (52%) of the reports related to programme evaluations (including six joint programmes), 29% to projects and 11% to policies. More than half of all project evaluations rated Good (25%) or Very Good (38%), and 86% of programme evaluations were rated at least Satisfactory (with 50% Good or above).

The strongest parameters of performance in terms of the overall standard of reports were Parameter 1 (Object and Context) and Parameter 7 (Gender and Rights). Both of these rated 100% Satisfactory or above. However, the most frequent examples of excellence were found in the Findings section, with 33% of reports rating as Very Good for Parameter 4.

The Parameters with the greatest challenge were conclusions (Parameter 5) report structure (Parameter 8), with 22% of reports rating as Unsatisfactory. Recommendations (Parameter 6) is also an area of challenge, with fewer reports (45%) rating as Good in this section than any other. Many reports struggled to translate excellent findings into deeper insights for decision makers and relevant, actionable and targeted recommendations.

The review identified a wide range of good evaluation practices across all regions and levels of evaluation. Examples of these include:
☐ Use of evaluation reference groups to guide the inception-stage design, comment on emerging findings, and reflect on lessons learned;
☐ Including a process for obtaining free and informed consent as an ethical safeguard during data collection;
☐ Clear processes by which marginalised stakeholders were identified and engaged in participatory inception and validation phases;
☐ Several reports reconstructed the theories of change used by a programme, project or policy – and then tested these theories as part of the evaluation process;
☐ A report that used feminist theory to identify strategic challenges for the programme in terms of its operating environment and the patriarchical power systems within the institutions that the programme is seeking to influence.

UN Women is well suited to performing strongly in relation to UN SWAP criteria, as all the evaluated objects pertain to gender issues. Overall, the SWAP score (1.46) translated to a rating of **Meeting Requirements**. There was, however, variation across the criteria. The review observed the strong performance in criteria such as 'evaluation questions' and 'analysis' is often a consequence of evaluating a gender-focused object, rather than as a result of gender-specific questions. Despite this strong SWAP score, therefore, improving the gender-responsiveness of evaluation remains a priority for UN Women.

Preparation	1.43	Meeting Requirements
Methodology	1.77	Exceeding Requirements
Report and Use	1.01	Approaching Requirements
OVERALL	**1.46**	**Meeting Requirements**

2.3 Conclusions

Conclusions were developed based on the evidence presented in the findings, and have drawn on UN Women/UNEG standards for evaluation, evaluation reports and ethics in evaluation.

Conclusion 1: Evaluation Reports in UN Women perform well against UN Women, UNEG and UN SWAP standards despite the organisation having limited access to specialist evaluation resources. This is hypothesised as being an outcome of having regional evaluation expertise, good quality TORs, recruiting evaluators with credible expertise in gender responsive techniques, and focusing on evaluation objects that are inherently concerned with rights and gender equality.

Conclusion 2: The portfolio of evaluations is largely reliant on a single design and set of methods. Whilst these may be implemented with varying quality, the variance in GERAAS ratings comes primarily from the extent to which evaluation reports address issues of ethics, the participation of affected populations in the evaluation, and structured approaches to developing conclusions and recommendations.

Conclusion 3: Evaluations are addressing gender issues within the scope of the social enquiry process, but few are reporting validation and utilization processes that fully reflect gender-responsive evaluation techniques and human rights based approaches.

Conclusion 4: There is a good spread of evaluation experience across UN Women regions, with many examples of excellence – and the potential to learn from them – found everywhere.

Conclusion 5: Ensuring that all reports are logically structured is a priority for maximising existing strengths, especially strong Findings sections and frequent use of evaluation frameworks.

2.4 Recommendations

Recommendations are based on analysis of the findings and conclusions, rather than a participatory process.

Recommendation 1: Evaluation managers in UN Women can set their sights on performing well beyond the requirements of UN standards for gender and human rights; and systematically apply gender analysis to evaluation design, interpretation of findings, and development of conclusions.

UN Women can afford to, and need to, aim higher than the minimum standards, and learn from the best reports in its portfolio to push the boundaries of participation at all stages of the evaluation process. Evaluation commissioners in UN Women should thus continue to be guided by the UNEG standards, but also encouraged by the Executive Board and the Evaluation Office to commit the resources and time to deepening the use of participatory, gender-responsive, and empowerment evaluation.

Recommendation 2: Evaluation commissioners will benefit from maintaining and strengthening the existing quality of Terms of Reference and recruitment of evaluators with expertise in gender responsive evaluation.

Evaluation commissioners should remain focused on continuous improvement in the development of TORs and recruitment of evaluation teams. This should also be supported as a key consideration in systems strengthening initiatives undertaken by the Evaluation Office.

Recommendation 3. The UN Women Evaluation Office will help fill a gap by providing simple model guidance on methodological limitations, ethical considerations and utilization approaches in mixed methods participatory evaluations.

It is recommended to the Evaluation Office to produce the following capacity-strengthening products written for non-specialist evaluation managers:

1. **For the ToR stage:** a brief on the wider range of evaluation designs that may be appropriate or adapted to UN Women's context;
2. **For the inception stage:** a quick reference guide on epistemological/ontological limitations and ethical issues faced by theory-based mixed method designs;
3. **For the implementation stage:** model *free and informed consent* protocols;
4. **For the reporting stage:** a brief on ideas and expectations for the dissemination, follow-up and utilisation of evaluation outputs (including reports).

Recommendation 4: By focusing on developing a learning community among evaluation managers in UN Women, the Evaluation Office can maximise the proven capacity of non-specialist staff managers to deliver high quality evaluation reports.

The Evaluation Office should explore opportunities to nurture a learning community for evaluation within UN Women to share experience and inspiration across regions. As a contribution to this knowledge community, it is also recommended that the Evaluation Office continue to refine GERAAS, including enhancing the consistency of budget information available on GATE to allow for analysis of quality according to the resource-scale of evaluations and evaluands.

2.5 Lessons Learned

The overall achievement of 85% of reports satisfactorily meeting UNEG standards and strong performance in UN SWAP would appear to suggest that **designing evaluations with a focus on gender and human rights can also have wider beneficial effects** on other aspects of evaluation quality, even in situations where specialist resources are (relatively) constrained. Our hypothesis is that the recruitment of evaluation teams based on credible GEWE experience may be leading to self-selection of evaluators with a greater range of skills and commitment to professional development. UN Women's commitment to empowerment and human rights based approaches is thus likely to be advantageous to pursuing wider aspects of technical excellence in evaluation.

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 EO 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 evaluator was appointed to undertake both a meta-analysis and meta-synthesis of 2013 evaluation reports submitted to GERAAS, including the UN-SWAP scores. The selected evaluator, Joseph Barnes, leads a professional evaluation and design partnership – ImpactReady – based in the UK, has previously supported the design and implementation of a similar system for another UN agency, and has led or quality assured evaluations for over 20 development organisations. Non-English reports were assessed in partnership with IOD PARC, an experienced performance assessment consultancy.

4. Purpose, Objectives & Scope

The Purpose of this meta-analysis is to capture the quality of evaluation reports – according to UN standards – from a critical year of transition in UN Women's evaluation capability. This is required to develop constructive lessons for future systemic strengthening of evaluation, and provide a baseline for future analyses based on the GERAAS methods.

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 2013 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 (with the exception of the management response). The actual evaluation process or utilisation of the evaluation is 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 Evaluation Office oversaw, coordinated and supported the review process.

This report assesses final evaluation reports uploaded in the UN Women Global Accountability and Tracking of Evaluation System (GATE) by 15 January 2014. The Evaluation Office 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. Two reports were removed from the final analysis as they were considered to be reports other than evaluation reports. Written comment was provided to the Independent Evaluation Office justifying this view.

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.

The IEO 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.

5.1 Review 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:

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.
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	*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. *Conclusions* present reasonable judgments based on findings and substantiated by evidence, providing insights pertinent to the object and purpose of the evaluation.

sound analysis	*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. *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.

5.2 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 secondary reviews by the main reviewer to quality assure consistency, and c) using reviewers who have previously undertaken similar meta-analyses (including UN SWAP) for UNICEF and WFP.

Table 1: Languages and reviewers

Language	Reports	First round	Second round
English	21	Reviewer 1	Reviewer 1
French	3	Reviewer 2	Reviewer 1
Spanish	2	Reviewer 3	Reviewer 1
Portuguese	1	Reviewer 3	Reviewer 1

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.

Figure 1: Meta-evaluation process

Filtering by EO

| IEO checks that all submitted reports are evaluations | **2** reports removed from the review as considered outside the scope of 2013 |

Filtering by Independent Reviewer

| Reviewer checks that all submitted reports are evaluations | 2 reports (one baseline study, one training assessment) removed and feedback provided |

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 re-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 | 7 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 findings, conclusions and recommendations from reports meeting UN Women standards | See separate meta-analysis report. |

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 45 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.

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 cannot be used with confidence 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.

5.3 Changes Made From Previous Years
Unlike in previous years, GERAAS uses a qualitative-based approach to assess report quality. This acknowledges the wide variety of contexts in which evaluation reports are produced; and by doing so seeks to focus on developing constructive insights into strengthening the evaluation function, rather than the minutiae of a quantitative 'score'. It also aims to reflect the real-world use of evaluation reports by

decision makers, who are more interested in the overall coherence of a report rather than the aggregate 'score' of whether particular elements are present or not. The review process and the meta-analysis use the ratings system as a departure point for identifying constructive guidance for future reports, more than a final 'number' that defines a report's quality.

5.4 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. Furthermore, the recent Quadrennial Comprehensive Policy Review Resolution (A/RES/67/226) adopted by the General Assembly in December 2012 requests 'the Joint Inspection Unit to undertake a system-wide evaluation of the effectiveness, value added and impact of the System-wide Action Plan on Gender Equality and the Empowerment of Women as a tool for performance monitoring and accountability for submission to the General Assembly following its full implementation.' Accordingly, this requires that systems to report against this performance indicator are developed and in place so that progress can be shown by UN entities and as an input to the evaluation.

The ultimate goal is that all UN system entities "meet requirements" related to this 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 13 UN SWAP Evaluation Performance Indicators:

Fully Met	2.0
Partially Met	1.0
Missing	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:

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

It was noted that, despite the definitions provided under the SWAP, the interpretation of ratings can still accommodate a wide range. It was also noted that the underlying logic of the ratings criteria and definitions are more aligned to a traditional 'project' evaluation, rather than an evaluation of a GEEW themed policy (for example). The following general assumptions have thus been used:

1. An evaluation object that is themed on GEEW 'converts' the standard OECD evaluation criteria to GEEW-sensitive, so long as they are discussed as such in the main report;
2. Evaluation of GEEW aspects of an evaluation is a proxy indicator of an evaluability assessment having taken place (even if it is not in the report);
3. Evaluation of GEEW aspects of an evaluation is a partial proxy indicator of a gender-specialist team being recruited, even if no details are provided in the report;
4. The use of participatory tools – especially focus group discussions with disaggregated groups of beneficiaries – is an indicator of a gender-sensitive methodology within the resource-window available to the evaluation function in UN Women.

5.5 Limitations

GERAAS does not measure the quality of evaluations. 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.

This leads to **some known cases of over-rating and under-rating**. For example, where an evaluation contains all the required elements – but has departed from approaches agreed at the inception stage – the report is likely to be rated more highly than the evaluation manager might expect. Conversely, where the report contains limited information but is an output of an extensive and useful process – such as might be expected from Collaborative Outcomes Reporting Technique, for example – the rating is likely to under-rate the evaluation compared to its 'felt-utility'.

Nevertheless, addressing these wider aspects of utility – through, for example, interviews with evaluation users – introduces a much higher degree of complexity,

subjectivity, and potential for bias that requires substantially more time and resources than is available to GERAAS. Within these real-world constraints, therefore, the approach taken to GERAAS attempts to maximise consistency of rating across the portfolio of evaluations. Furthermore, where decisions on final ratings are borderline, the reviewers have sought to 'reward' positive aspects of reports, rather than seek out gaps.

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-analysis is based only on a limited number of evaluation reports, it also has limitations connected to developing findings around methodological and participatory processes. **Some data is limited**: for instance, in connecting particular themes of EVAW, WEE, Global Norms to the quality of evaluations. The report aims to highlight where there is uncertainty, and makes transparent suggestions for explanations where the data does not support firm conclusions.

A reviewer that was fluent in the language of a report first rated non-English reports. This introduces that **potential for bias based on languages**. However, the main reviewer also reviewed these reports a second time and ratings were only upgraded (not downgraded) by this second review. Reviewers were also selected who have previous experience with evaluation report quality assurance for the UN system.

6. Findings

6.1 Overall Ratings and Feedback

The review process found that 85% of evaluation reports could be considered as satisfactory or above according to UN Women standards: implying that decision makers could be reasonably confident in their findings, insights and recommendations. Whilst many of these reports were found to have opportunities for improvements, 26% were found to be exceptional – exceeding UN Women Standards in multiple parameters.

Figure 2: Overall ratings for all 27 reviewed reports
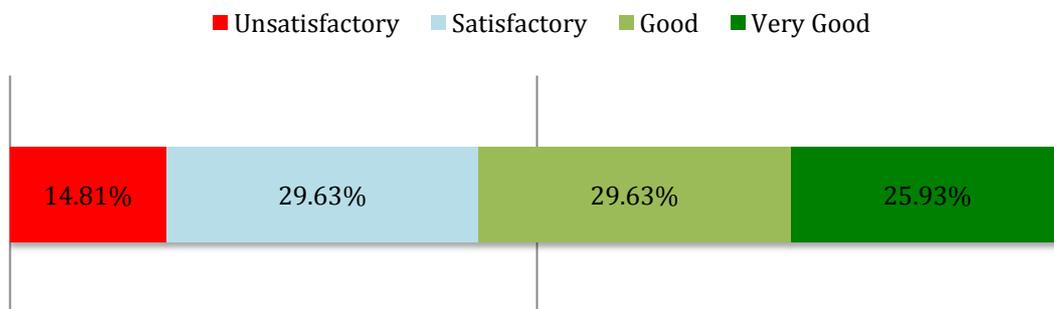
The review process revealed a number of patterns in the evaluation reports that were considered:

1. All evaluations choose to use very similar designs and methods: triangulating primary qualitative data (interviews and focus groups discussions) with secondary quantitative data (monitoring data, project data, financial reports). This design is generally well suited to the objects and purposes of the evaluations, but is interesting to note that as an overall portfolio there is a particular reliance on one type of evidence and its associated limitations;
2. There are only one or two evaluation reports submitted from each office. This may explain the strong performance of evaluations in comparison with the level of resourcing available to UN Women as an agency;
3. The objects of the evaluations and the evaluation questions are well suited to performing well according to the human rights and gender aspects of the UNEG standards and UN SWAP (i.e. they are inherently gender aware) – please see section 6.4.8 for a more detailed discussion of this; and
4. Partly because of the similar designs, the evaluation reports that rated Very Good tended to be differentiated by being very strong in relation to gender equality, participation and ethics.

6.2 Overall Regional Trends

The most evaluation reports (8) were submitted from the Eastern and Southern Africa region, the least (1) from Arab States. No region had more than a single report rated as Not Satisfactory.

Figure 3: Number of overall reports and ratings per region
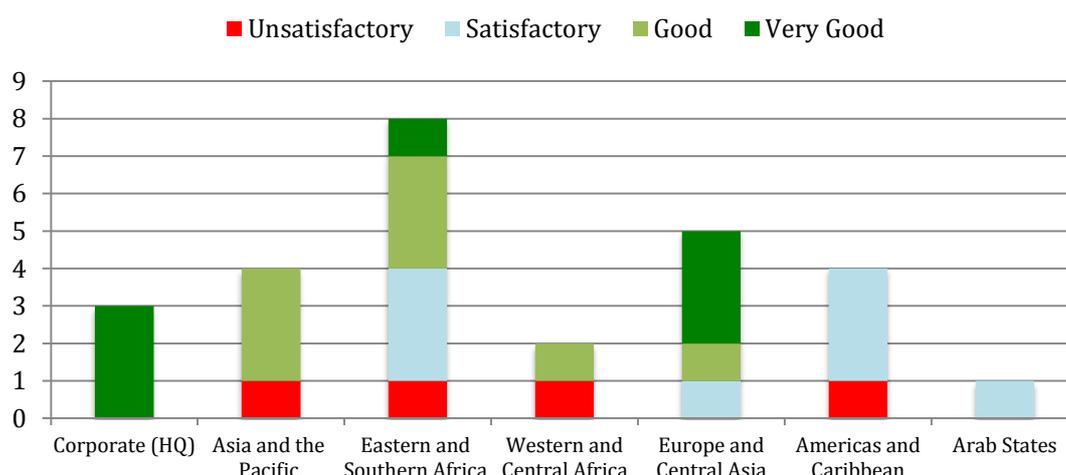
Europe and Central Asia was the only region to include a Country Programme Evaluation, Americas and the Caribbean was the only region to include an evaluation of a Pilot, and HQ provide the three Policy evaluations. The rest of the reports were programme and project evaluations.

All regions included reports with parameters rated as 'Good' (see Figure 4, below). The regions with more non-English reports had a higher occurrence of Satisfactory ratings. It is important to note, however, that most of these reports did still meet UN Women standards (see also the methods discussion under 5.5). This pattern follows a similar finding from other UN agencies[1] (undertaken by different reviewers) and it has been hypothesised that it could be linked to the availability of guidance material of a 'better fit' between the logic and structure of the UNEG standards with the logic and structure of the English language compared to Latin-based languages.

Figure 4: Percent of all parameter ratings rated per region

---
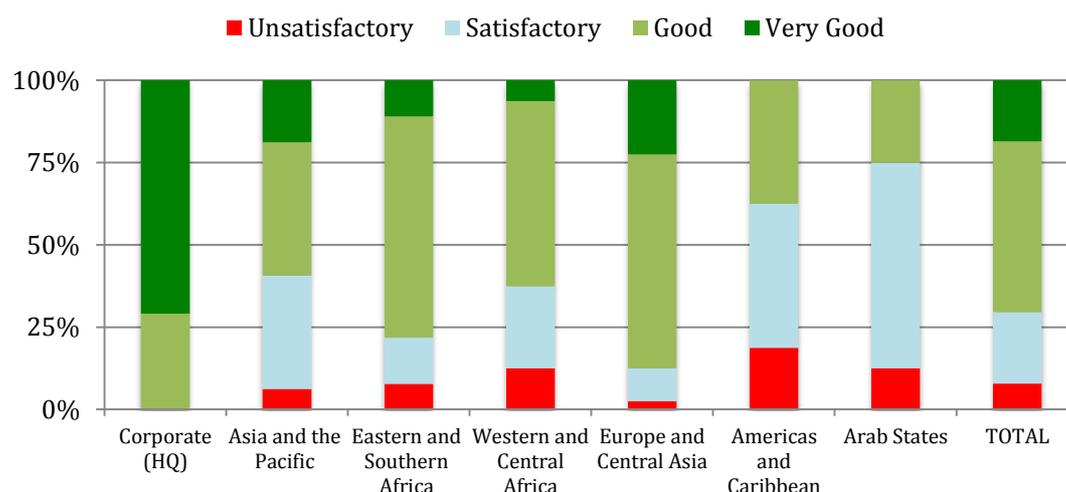
[1] E.g. UNICEF Global Evaluation Report Oversight System

6.3 Trends by Type and Scope of Evaluation

6.3.1 Geography

Excluding the case studies in corporate evaluations, at least 29 countries were included in evaluation reports. The majority (44%) of evaluations were national level evaluations, with over 83% of these rating as Satisfactory or above. A higher proportion of regional and multi country evaluations were associated with Europe and Central Asia, and Americas and the Caribbean. In relative terms, the highest performing evaluation reports were from global evaluations (11% of reports), and the weakest evaluations were sub-national in scope (19% of reports).

It was noted that, at the country level, reports with a national scope were rated significantly higher (66% Good or above) than sub-national evaluations. Regional and multi-country evaluations had a more mixed performance than national evaluations, although with a high proportional occurrence (43%) of Very Good reports. The available data is inconclusive in explaining these variations.

6.3.2 Management of Evaluation

UN Women directly managed at least 67% of the evaluation reports. Of the 26% of reports that were identified as being jointly managed, the majority related to Joint Programmes with other UN agencies (it is not clear from the reports how many were led by UN Women). It is thus relevant to apply the same UNEG standards to these reports. Management arrangements were not clear for the remaining reports.

Overall quality of the reports was very similar for both management arrangements (perhaps reflected the common UNEG standards noted above). However, it was noticeable that the Very Good reports tended to be associated with direct management by UN Women. This suggests that the organisation has good evaluation management capacity, but also reflects the greater complexity associated with multiple-managers in joint evaluations.

Figure 5: Percentage overall ratings for different management arrangements
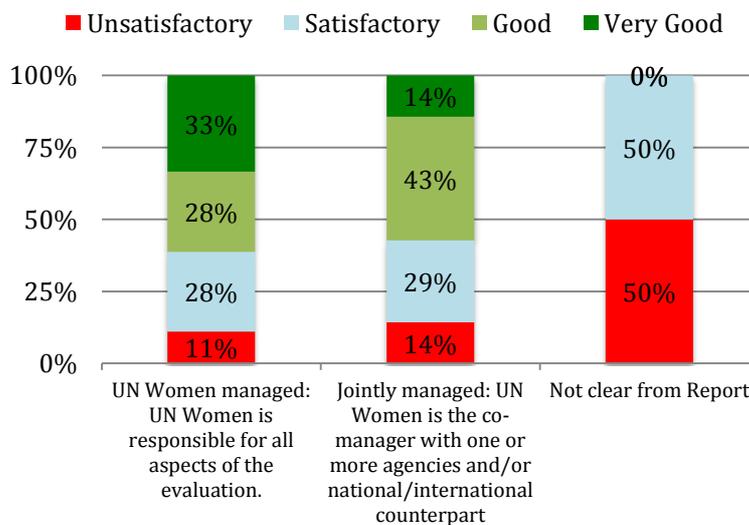
6.3.3 Type of intervention evaluated

Half (52%) of the reports related to programme evaluations (including six joint programmes[2]), 29% to projects and 11% to policies. There was only one evaluation relating to a pilot and only one country programme evaluation. Whilst programme evaluations tended to be associated with assessing outcomes and policy evaluations with assessing impacts, project evaluations considered a range of different results levels – reflecting the broad use of this modality by UN Women.

More than half of all project evaluations rated Good (25%) or Very Good (38%), and 86% of programme evaluations were rated at least Satisfactory (with 50% Good or above). Programme and project evaluations were most associated with outcome-level evaluations, which were consistently the strongest (see 6.3.4 below). The three corporate evaluations were classified as policy evaluations. The sample of country programme and pilot evaluations is too small to consider in terms of meaningful findings.

6.3.4 Results/level of changes

The 'level of changes' considered by reports were classified according to what the report actually considered. Reflecting the correlation with programme evaluations, exactly half (50%) of reports considered effectiveness in terms of outcomes. Closely associated with project evaluations, 31% of reports considered outputs. The remaining 19% of evaluation reports – mostly policy and joint programme evaluations – attempted to interrogate impacts (contribution to longer-term more complex effects), usually considering both UN Women's work and preceding work by UNIFEM.

Outcome evaluations were the most consistent, with 100% rating as Satisfactory or above (31% Very Good). By contrast, output evaluations tended to be considerably weaker in overall performance (25% Unsatisfactory), also reflect the fact that they tended to be associated with sub-national evaluations. Impact level evaluations (mostly policy-level) recorded the greatest relative proportion of Very Good reports (60%), although outcome level reports accounted for more Very Good reports (4).

> **Impact:** Final results of a programme or policy on the intended beneficiaries and, where possible, on comparison groups. 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

---

[2] The GERAAS methodology does not classify Joint Gender Programmes separately, although it is recommended that this be adjusted for future years if it is an issue of interest to the organization.

Figure 6: Percentage overall ratings for different levels of change
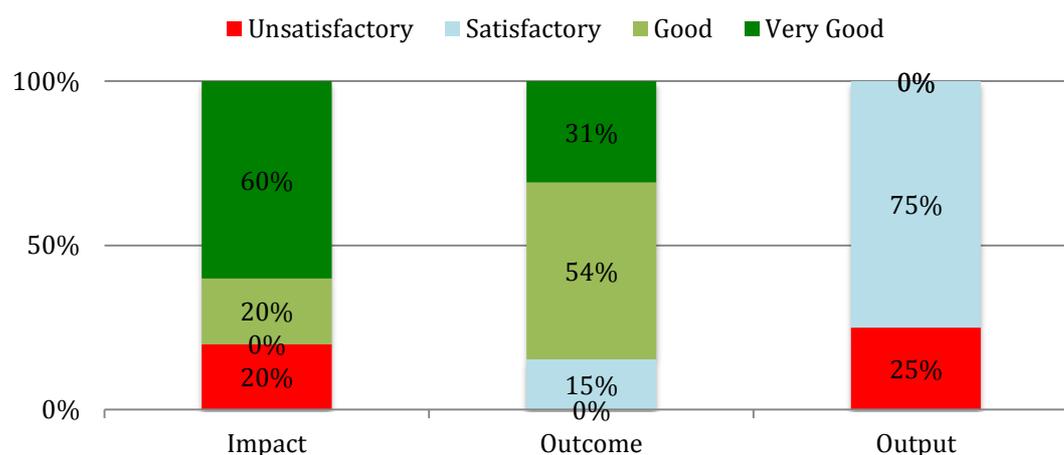
6.3.5 Stage

The majority of reports (69%) were final/summative evaluations, with the remaining 31% being mid-term/formative. Overall, formative reports consistently performed more strongly than summative reports, with 87% rating as Satisfactory or above (83% for summative), 63% rating as Good or above (56% for summative), and 38% rating as Very Good (17% for summative). There is no clear evidence in the data to explain the slightly better performance by formative evaluations.

6.3.6 Strategic Plan Correspondence

Evaluations considered all of the Strategic Plan Impact Areas, with the greatest consideration given to Impact Area 3 – VAW (13 reports) – and Impact Area 1 – participation and leadership (11 reports). All impact areas were considered in at least 6 reports.

Figure 7: Distribution of reports considering Strategic Plan Impact Areas
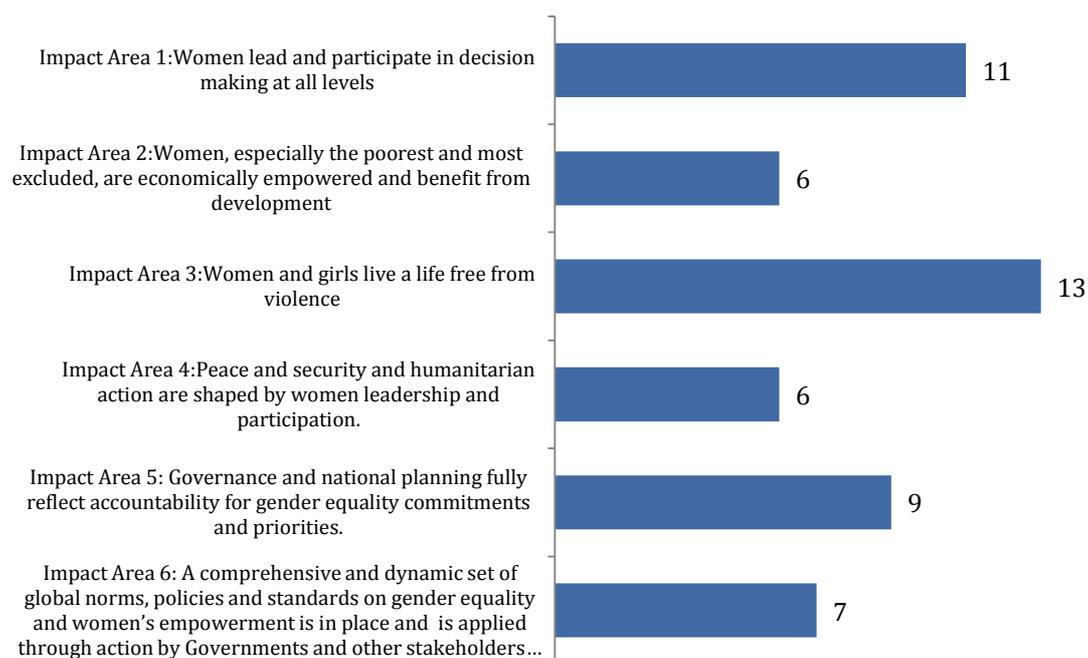
The overall performance in each strategic area presents a complex picture, with the highest percentage of reports being considered Satisfactory or above in relation to peace and security (83%) and the lowest in global norms (57%). However, considering only the reports that were rated as Good or above (which is one target for the future) reveals a different picture, with evaluations of economic empowerment rating most highly and violence least highly. One potential explanation to explain this could be the increased degree of complexity (including specialist knowledge and ethics) associated with EVAW and influencing global norms.

Figure 8: Percentage of reports rated as Good or above for strategic goals

3. Free from Violence	6. Global Norms	5. Governance Accountability	1. Women Lead	4. Peace and Humanitarian	2. Economic Empowerment
31%	43%	44%	45%	50%	67%
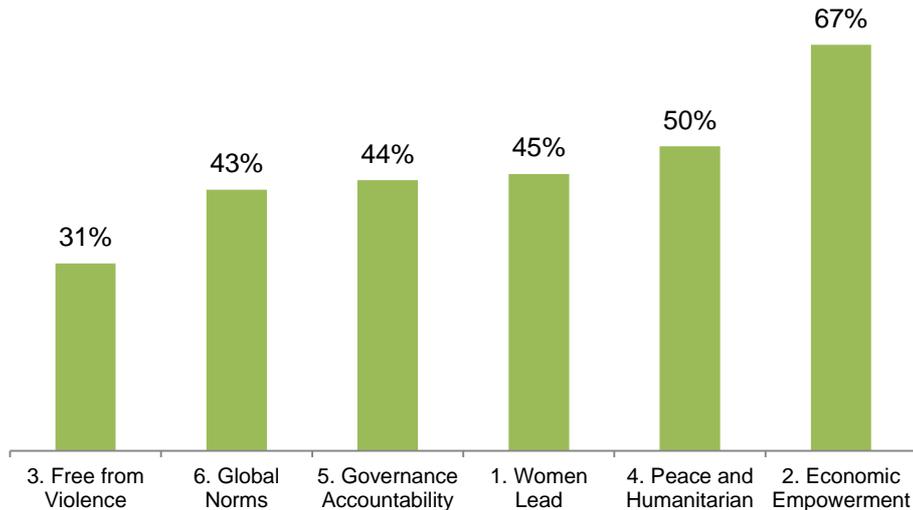
6.4 Trends by Quality Assessment Parameters

6.4.1 Overall trends Parameters 1-8

The strongest parameters of performance in terms of the overall standard of reports were Parameter 1 (Object and Context) and Parameter 7 (Gender and Rights). Both of these rated 100% Satisfactory or above. However, the most frequent examples of excellence were found in the Findings section, with 33% of reports rating as Very Good for Parameter 4.

The Parameters with the greatest challenge were conclusions (Parameter 5) and report structure (Parameter 8), with 22% of reports rating as Unsatisfactory. Recommendations (Parameter 6) is also an area of challenge, with fewer reports (45%) rating as Good in this section than any other. Many reports struggled to translate excellent findings into deeper insights for decision makers and relevant, actionable and targeted recommendations. These issues are explored further below.

Figure 9: Overall performance of parameters 1-8
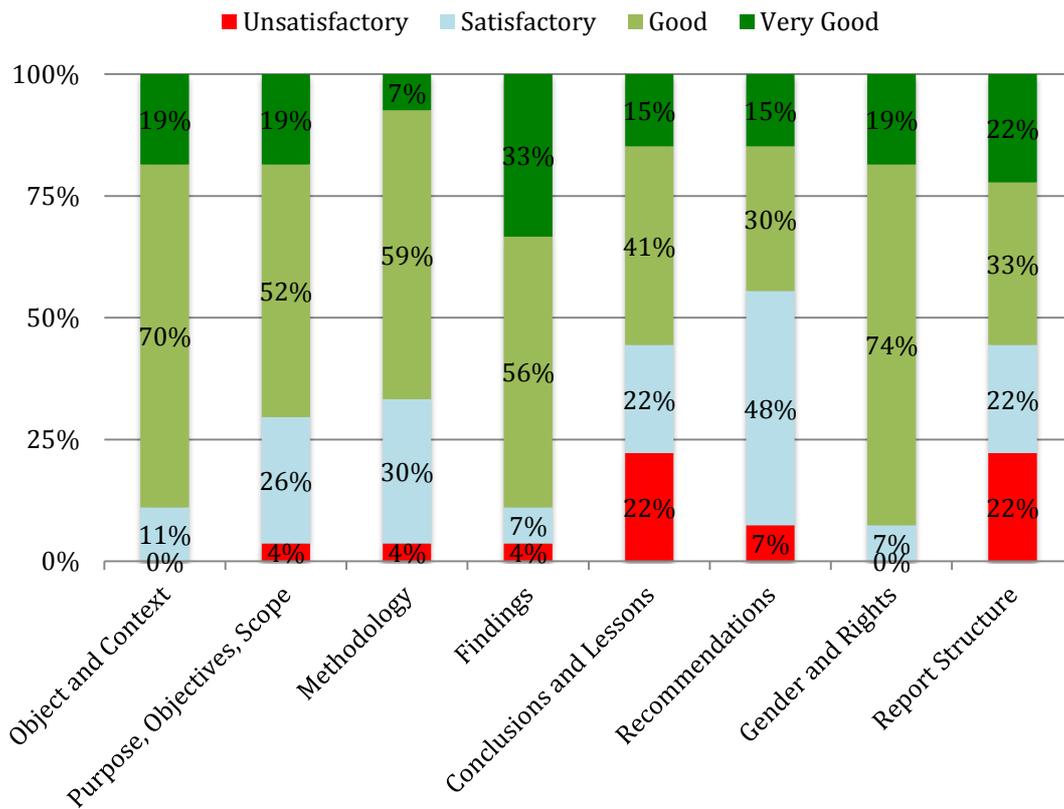
6.4.2 Parameter 1: Object and Context of the Evaluation

Whilst being a strong section overall, there are a couple of areas of refinement that could help improve the quality of many reports. Most noticeably, the identification of stakeholders was basic in many reports. There was an observed tendency to identify only institutional stakeholders (organisations, agencies and formalised groups), and many reports referred to 'women' as a homogenous group under the single stakeholder of 'beneficiaries'. The best reports include critical power analyses and socio-economic disaggregation of programme participants (and those who may have been excluded).

Many reports included the logframe or results framework for the evaluation object, which is considered good practice. Others reconstructed the theory of change, which is also recognised as good practice[3]. However, in many cases both of these are 'buried' in the reports rather than being introduced or explained as part of the background. Thus, they are either to be found in the annexes, or revealed piece-by-piece through reading the findings. This reduces the overall value of including these analytical frameworks, as the reader misses the 'big picture'.

The strongest performance was in relation to describing the implementation and nature of the evaluation object. The best reports differentiated themselves by

---

[3] Theories of Change is considered to be a gender-responsive approach to meeting the UNEG requirement of *"The subject to be evaluated should be described in terms of what it aims to achieve, how the designers thought that it would address the problem they had identified".* See, for example, guidance material available at mymande.org.

providing insightful histories (linked to the context analysis) and clear information on budgets, resources, and accompanying processes.

6.4.3 Parameter 2: Purpose, Objectives and Scope

Purpose and objectives of evaluations were frequently strongly described in reports (44% were considered Very Good). This tended to be grounded in good quality Terms of Reference. Describing the scope of the evaluation was a more challenging area, with most reports limiting their analysis to the years and place of evaluation, rather than describing what issues were to be included, excluded and why.

Nearly all reports used the OECD-DAC standard evaluation criteria. The reports rated as Good included explanations of how these were to be defined in the context of the evaluation. Around 4 out of every 5 evaluations included a clear evaluation framework in the annexes, linking the evaluation questions to the evaluation criteria. All evaluations considered gender and human rights issues, although this tended to be as a result of the evaluation objects rather than with specific evaluation questions.

It was observed that there was almost no discussion of how evaluations were to be communicated, or other utilization considerations. One corporate evaluation did describe an initial process centred on UN Women's planning for utilization, but the outcomes of this process were not described. It is noted that this process is normally to be undertaken separately by the commissioning office, but without reference to communication strategies in the evaluation report it is not fully clear on the extent to which the report (or the evaluator) is responding to the utilization needs of UN Women.

6.4.4 Parameter 3: Methodology

The review found that reports tended to focus most strongly on describing data collection tools and data sources. With regard to technical aspects, description of the overall design and the detailed explanation of sampling tended to be the areas of greatest challenge.

It was particularly observed that reports tended to described constraints (time, resources, data) instead of methodological limitations (attribution, bias, representation). Even where limitations area described, most reports attempt to reassure the reader that these were successfully mitigated, rather than to acknowledge potential issues and explain the practical implications for decision makers.

The weakest area in the whole portfolio of evaluations is ethics. Reports tend to either be excellent in this regard (26% of reports rated Very Good), or particularly weak. Reports were still rated as Satisfactory for ethics even if there was no specific description of the ethical considerations but they fulfilled three conditions: a) the overall approach was participatory, human rights sensitive and reflected ethnographic norms in reporting, b) the data collection tools reflected an ethical approach, and c) there were no obvious vulnerable groups who might need special considerations (such as survivors of SGBV). Despite this – and the fact that most TORs provide a link to the UNEG standards on ethics – 22% of reports rated

Unsatisfactory. This suggests that ethics needs special attention in regard to future evaluation reports.

6.4.5 Parameter 4: Findings

The Findings is a strong section across nearly all evaluations. Reports presented structured analysis rather than raw data, and most often presented information in a logical way (either by evaluation criteria, or programme components for complex evaluations). Generally, evidence was well marshalled, although it was observed that effectiveness, relevance, and sustainability tended to be considered in greater depth than efficiency or impact.

A number of reports innovated well in the findings section, making good use of specific sections on unexpected findings, graphics, case studies, and ratings schemes to allow comparison across criteria. There was a noticeable tendency, associated with the dominant evaluation design used, to triangulate between different qualitative sources, rather than between different types of data. In many cases, this was because evaluations were reliant on project M&E systems for quantitative data, which was often not available or of poor quality.

In some cases, the findings section became rather bloated, encompassing information that would have been better placed in the context or conclusions sections. Nevertheless, this is a case of report editing and refinement rather than a fundamental issue.

6.4.6 Parameter 5: Conclusions and Lessons Learned

The conclusions section is the weakest parameter in terms of the number of reports rated as Unsatisfactory (22%). In two cases this was because reports missed out conclusions entirely, but in most cases it was found that conclusions were either primarily a summary of findings, or extremely brief. A number of reports chose to use bullet points, which tended to result in simple statements that were not elaborated in terms of greater insight or substantiating evidence.

A significant area of challenge related to substantiating conclusions without introducing new evidence. Reports often did not attempt to identify solutions or consider the full range of strengths and weaknesses of an evaluation object. Very few reports elaborated on the process of how conclusions or lessons learned were developed.

Not all reports presented lessons learned, as is appropriate. Where these were included, 44% of reports correctly interpreted the UN Women standards, elaborating generalised insights that are not common knowledge. Other reports listed points that were either more like conclusions or even recommendations, however, 89% of reports were still considered as meeting the minimum standards for lessons learned.

6.4.7 Parameter 6: Recommendations

All reports included a set of recommendations. In all cases, these were found to be relevant to the purpose of an evaluation, and in most cases they were well supported by the evidence.

The principle challenge faced by recommendation sections was the lack of any description about how they were developed, refined and validated – including whether or not relevant stakeholders were involved in this process. Methods sections of reports tended to stop the description of the process at the data gathering and analysis stage, whilst recommendations section tended to launch into the recommendations without any preamble. Having said this, a couple of reports did make clear provision for the participation of stakeholders in validation, and these serve as interesting examples.

Overall, it was observed that recommendations lack sufficient detail to be understood in terms of their feasibility, prioritisation, or target audience. Many recommendations are broadly targeted – such as to 'UN Women' or 'donors', but have not been developed with a specific unit or position in mind. The best reports differentiated themselves by having fewer recommendations, which were clearly prioritised and described in sufficient detail to be immediately actionable.

6.4.8 Parameter 7: Gender and Human Rights

All reports were written by people familiar with human rights and gender concepts, engaging with these issues through the documents. The nature of the evaluation objects also led to extensive discussion of gender issues and marginalised groups within reports, although not all reports followed this through in terms of disaggregated analysis of quantitative data or evaluation participants. A number of reports include discussions of rights-holders and duty bearers in the stakeholder analysis, but this is not consistent across all evaluations.

As discussed further under section 6.6 UN SWAP, most reports tended to derive their alignment with gender and human rights principles by applying standard OECD-DAC criteria and participatory methods to a gender-focused evaluand. Thus, findings, conclusions and recommendations were gender responsive to the extent that they recommended improvements to an already gender-focused intervention. Few reports went substantially beyond this. In attempting to identify those that did, it was noted that some of the most gender-responsive reports also tended to be those with weaknesses in the structure of the conclusions and recommendations section (such as summarising many findings rather than offering deeper insights, or making too many recommendations).

It was observed that there is an inherent tension between the utilization-focused and gender equality/rights based designs adopted by most evaluations. Thus, whilst issues of rights and social justice were considered, the majority of evaluations did so from within the predefined framework provided by UN Women in the TOR (as the primary intended users). A couple of reports described an inception process that included engaging stakeholders in refining the evaluation questions, however, none included a process for meeting the information-rights of UN Women's stakeholding groups – particularly marginalised women.

6.4.8 Parameter 8: The Report Structure

Parameter 8 found the easiest space for improvement within the overall analysis – with report editing contributing substantially to the overall coherence, usability and

confidence of an evaluation. Whilst most reports were broadly logical in their layout, a number became confused at the end with unusual arrangements of findings, conclusions, recommendations and lessons so that the flow of logic was broken. Nevertheless, nearly all reports had an accessible and easy to read writing style, and made use of clear language.

The areas of greatest challenge were the executive summaries and the annexes. Whilst only one report failed to include an executive summary, around one thirds of reports missed important aspects out of the executive summary (such as conclusions or methods) that are required for it to stand alone as a document. Most reports included annexes, although 38% missed out important elements or were poorly organised. Reports that differentiated themselves as very good (37%) combined well-edited annexes with cross-referencing in the main report to enhance the utility of the overall evaluation.

6.5 Examples of Good Practices in Evaluation Reports

The review identified a wide range of good evaluation practices across all regions and levels of evaluation.

6.5.1 Evaluation Process

Whilst the most extensive use of **evaluation reference groups** was in the corporate evaluations, a number of decentralised evaluation reports also identified this strategy for quality assurance. Evaluation reference groups were used to guide the inception-stage design, comment on emerging findings, and reflect on lessons learned. It was noted in some reports that the main challenge was in ensuring sufficient time to meet back with reference groups during the evaluation process.

Whilst it was common for national-level evaluations to include a **wide range of stakeholders** – from policy makers to rural women – as part of the process, a number of regional (or multi-country) and global evaluations also achieved this. The main challenge was in ensuring that the voice of these participants was heard in the report alongside the voices of power-holders.

A few reports mentioned the application of – and process for obtaining – **free and informed consent** as an ethical safeguard during data collection. Others went beyond consultation of marginalised groups during the enquiry stage, and elaborated clear processes by which marginalised stakeholders were identified and engaged in **participatory inception and validation** phases.

6.5.2 Evaluation Design

Due to the absence of theories of change, or limited results frameworks, in many evaluands, several reports **reconstructed the theories of change** used by a programme, project or policy – and then tested these theories as part of the evaluation process. One of the strongest examples of this also drew on 'generic theories of change' (e.g. change is created through mobilising sufficient numbers of people, or change is created through altering the calculation by elites).

In addition to restating the evaluation criteria and questions from the Terms of Reference, a number of reports **unpacked evaluation criteria** within the context of the evaluand. These added analytical frameworks that were relevant to the evaluation object in addition to generic OECD DAC definitions of the criteria. One report from Zimbabwe innovated by replacing the evaluation questions with **simple hypotheses** that could be more easily confirmed or refuted by the team (for example, *"More men are aware of women's sexual and reproductive rights and are able to respect them"*).

Other evaluations innovated the design around limitations in resources or available data. One report describes how elements of **contribution analysis** were adapted to the context and resources available even though full use of this approach – whilst relevant – was not feasible. Another report triangulated qualitative data from case studies with a wider **quantified survey** in order to apply mixed methods despite insufficient monitoring data being available.

6.5.3 Evaluation Analysis and Reporting
At least two reports **mapped the history of a programme**, including how it has evolved overtime and where major delays (such as financial disbursements) have had implications on results. These were presented graphically, clearly showing the reader the implications that funding delays had on reducing the time available for implementation.

Recognising the different interpretations that people apply to common gender and development concepts, one corporate evaluation developed a **context-specific glossary** to define key concepts used in the evaluation. Several reports found ways to enhance learning by providing quick **summaries of findings** in the text (e.g. using text boxes), in addition to detailed descriptions.

A number of evaluations applied a **Strengths Weaknesses Opportunities Threats (SWOT)** framework to assess a programme as part of developing conclusions[4]. Other interesting analytical approaches included use of text-coding tool *WorditOut* (an alternative is Wordle.net) to count and **analyse open text for key words**; and developing a **ratings rubric** (mostly a scale-of-five) to compare relative performance across the evaluation criteria.

It was also noted that a number of reports included a specific section **on unexpected results** from the programme and tried to explain these. One report also made **practical suggestions for future indicators** that were viewed to be missing from the programme M&E Framework.

6.5.4 Integration of Gender Equality
Not all reports provided a detailed **gender disaggregation of evaluation participants**, and thus it is noted as a good practice by those that did. One report from the State of Palestine used the gender-disaggregated data from a survey to

---

[4] It should be noted that, although SWOT is the most well known, there are many strategic analysis frameworks available and SWOT may not always be the optimal approach from an Organisation Development perspective.

disaggregate perceptions of different stakeholders and one report from Kenya systematically considered specific **implications for different stakeholding groups**.

It was noted that a number of reports were exceptional in their theoretical approach to gender. Of particular note from Grenada was a report that used of **feminist theory to identify strategic challenges** for the programme in terms of its operating environment and the patriarchical power systems within the institutions that the programme is seeking to influence.

6.6 UN SWAP

In addition to the GERAAS ratings, the review process considered the UN SWAP scorecard. This is a reporting tool organised around 13 scoring criteria which are articulated around 3 headings that capture the overall elements related to mainstreaming gender equality throughout the evaluation process.

Each criterion was rated as Fully Met (score 2), Partially Met (score 1), Missing (score 0) or not applicable. Scores have been aggregated and assessed using the protocol specified in UN SWAP guidance.

It was observed that evaluation in UN Women is well suited to performing strongly in relation to UN SWAP criteria, as all the evaluated objects pertain to gender issues (thus making gender an inherent consideration). Overall, the SWAP score translated to a rating of **Meeting Requirements**. There was, however, variation across the criteria. The review observed– based on evaluation reports and GATE – that UN Women performed most strongly in relation to the Methodology criteria – exceeding SWAP requirements.

As noted in the discussion under 6.4.8 (above), however, the strong performance in criteria such as 'evaluation questions' and 'analysis' is often a consequence of evaluating a gender-focused object, rather than as a result of gender-specific questions. This review notes the stated intention of UN SWAP to contribute to progressive improvement in gender-responsive evaluations. It should be expected, therefore, that despite this strong SWAP score, improving the gender-responsiveness of evaluation remains a priority for UN Women through, for example, more frequent use of gender specific questions, feminist analysis, human rights based analysis of roles, co-design of evaluation frameworks, and participatory validation.

Criteria relating to Evaluation Preparation were found to meet requirements overall, with Report Use criteria approaching the required standards. The major challenge faced by UN Women appears to be articulating clear plans for evaluation dissemination, and to ensure that such a strategy is gender responsive. As this particular aspect may arguably be outside the scope of evaluation reports, other areas requiring stronger focus are the inclusion of stockholding groups in a gender-responsive validation process, and ensuring that all reports have a management response that includes addressing gender issues (or highlighting where these have been missed).

Figure 10: Performance across UN SWAP criteria of reviewed reports and management responses
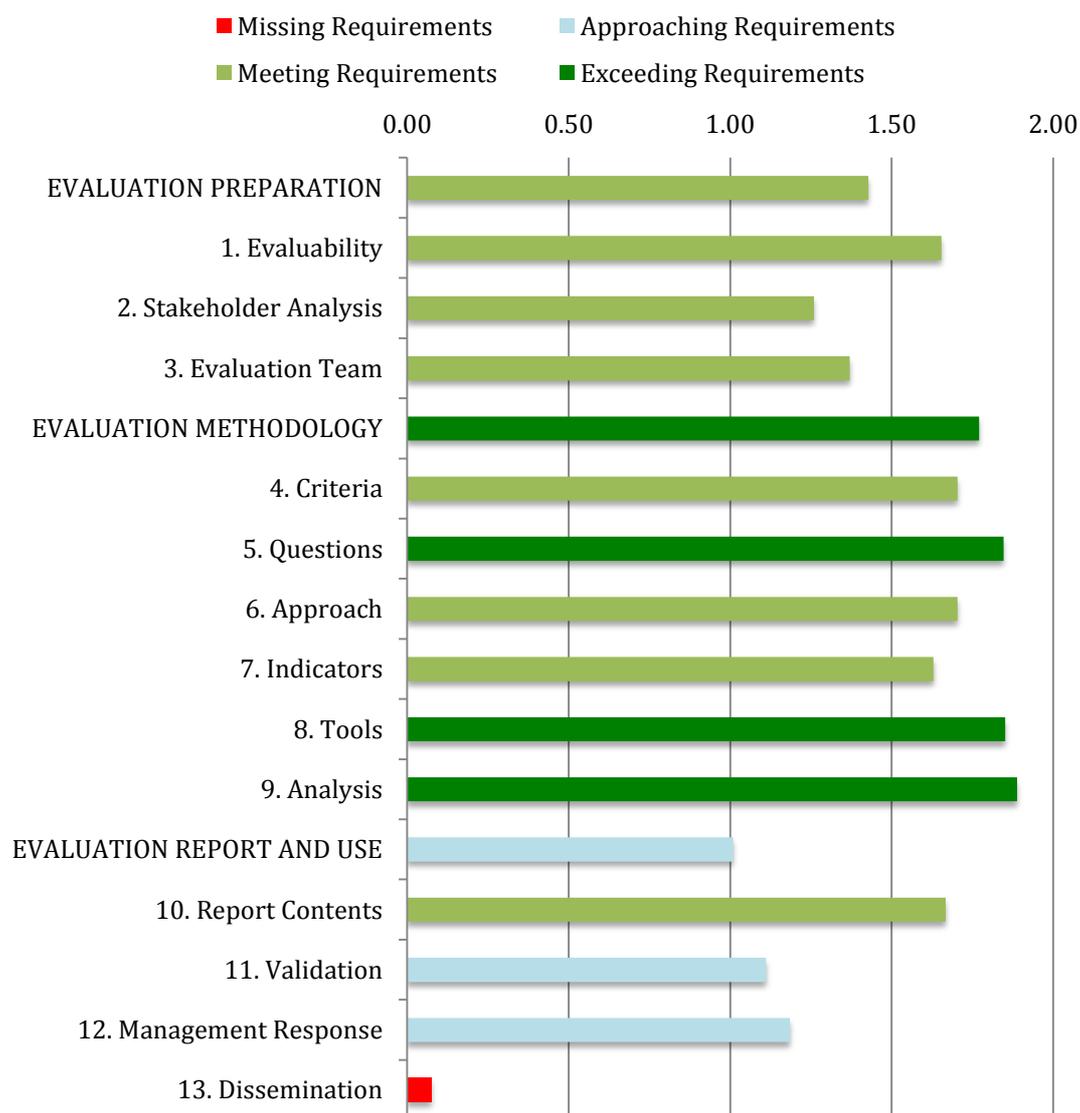
Table 2: Detailed UN SWAP scores

Dimension	Score	Classification
Preparation	**1.43**	**Meeting Requirements**
Evaluability of the GEEW aspects of the intervention is assessed	1.65	Meeting
A gender- responsive Stakeholder Analysis is undertaken.	1.26	Meeting
Evaluation Team is recruited that has the capacity to conduct gender-responsive evaluation	1.37	Meeting
Methodology	**1.77**	**Exceeding Requirements**
GEEW is integrated in Evaluation Criteria.	1.70	Meeting
Evaluation Questions are included that specifically address how GEEW has been integrated	1.85	Exceeding
A gender- responsive Evaluation Approach and Methodology is selected.	1.70	Meeting

Evaluation indicators are included to ensure GEEW- related data is collected.	1.63	Meeting
Data Collection Methods and Tools are developed to collect GEEW-related data	1.85	Exceeding
Data Analysis techniques were deployed that include gender analysis	1.89	Exceeding
Report and Use	**1.01**	**Approaching Requirements**
The Evaluation Report includes findings, conclusions, recommendation s and lessons	1.67	Meeting
The Validation Process of evaluation findings, conclusions and recommendation s	1.11	Approaching
The Management Response to the evaluation	1.19	Approaching
The Evaluation Dissemination Strategy is gender responsive	0.07	Missing
OVERALL	**1.46**	**Meeting Requirements**

7. Conclusions

These conclusions have been developed by the reviewer based on the evidence presented in the findings, and have drawn on UN Women/UNEG 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: Evaluation Reports in UN Women perform well against UN Women, UNEG and UN SWAP standards despite the organisation having limited access to specialist evaluation resources. This is hypothesised as being an outcome of having regional evaluation expertise, good quality TORs, recruiting evaluators with credible expertise in gender responsive techniques, and focusing on evaluation objects that are inherently concerned with rights and gender equality.

Whilst there remains clear room for improvement in multiple dimensions of evaluation reports, the overall performance is very positive, especially given the (relatively) small number of evaluation specialists in UN Women.

It would appear that, given the nature of the issues UN Women is addressing and the strategic nature of its action, significant attention is being given in evaluation TORs to methodological and process rigour (based on the brief reviews of TORs where they were included in report annexes). This is reflected in evaluation reports that, by-and-large, include all of the main elements required under UNEG standards. Indeed, more reports are rated as Very Good – and these are considered to be excellent examples for others to follow – than are rated Unsatisfactory.

Based on the distribution and number of evaluation reports, it is hypothesised that a significant contribution to this performance is having only a few reports per office, allowing managers to focus on quality over quantity. The implication of this is that by remaining focused and selective in its evaluations, UN Women can achieve strong results in terms of quality and continue to make the best use of evaluation technical assistance provided by regional experts.

Conclusion 2: The portfolio of evaluations is largely reliant on a single design and set of methods. Whilst these may be implemented with varying quality, the variance in GERAAS ratings comes primarily from the extent to which evaluation reports address issues of ethics, the participation of affected populations in the evaluation, and structured approaches to developing conclusions and recommendations.

It is noticeable that, in terms of design and methods, all of the evaluations submitted to GERAAS take a very similar approach (mixed methods, primary qualitative data triangulated with secondary quantitative data). The implications of this in terms of the range and limitations of evidence available to the organisation is beyond the scope of this analysis.

However, it is relevant to note that the main differentiating factors in report ratings relate primarily to how findings are translated through a structured process into high quality conclusions and recommendations. The extent to which evaluations identify, respond to, and analyse ethical considerations is also a differentiating factor in terms of overall quality, with the Very Good reports giving particular attention to this dimension.

The implications for evaluation managers are: i) at the inception report stage of evaluations, evaluation teams that give considerable attention to ethics may well be indicative of higher overall quality in other aspects; and ii) increased attention should be given to the process by which conclusions and recommendations are developed and validated.

Conclusion 3: Evaluations are addressing gender issues within the scope of the social enquiry process, but few are reporting validation and utilization processes that fully reflect gender-responsive evaluation techniques and human rights based approaches.

The overall strong performance in gender and human rights is explainable, to a degree, by the nature of the evaluation objects as demanding discussion of gender issues. In the majority of cases, however, this gender responsiveness is limited to the boundaries of the 'traditional' enquiry process – the evaluations starts with a predetermined set of questions set by UN Women, extracts data from stakeholders (using more-or-less participatory data) and then reports back the conclusions and recommendations based on a report-editing process with UN Women.

Whilst there are some notable exceptions to this trend, it remains unusual for reports to specifically and clearly articulate participatory processes in relation to the evaluation design and the development and validation of conclusions and recommendations. The implication of this, is that commissioning and managing evaluations with the goal of meeting the UN standards is unlikely to push the boundaries of participatory practice to the degree that might be possible given UN Women's knowledge of gender equality and empowerment.

Conclusion 4: There is a good spread of evaluation experience across UN Women regions, with many examples of excellence – and the potential to learn from them – found everywhere.

The review found that it was easy to identify examples of good practice in almost all evaluation reports, including those that did not meet the overall standards. This suggests that there is capability and experience within UN Women, and that the priority must therefore be a transfer of knowledge between different regions and levels of the organisation.

It is particularly noted that project evaluations would benefit strongly from the learning that has been developed in achieving good quality programme evaluations, and, similarly regional evaluations benefitting from the experience of corporate

evaluations. The implication of this is that the challenge is not one of achieving good quality in the first place, but of translating this experience into consistently delivering the same quality in evaluations with lower levels of resourcing.

Conclusion 5: Ensuring that all reports are logically structured is a priority for maximising existing strengths, especially strong Findings sections and frequent use of evaluation frameworks.

Report structure and editing is the 'low-hanging-fruit' identified by this review. Having a clear and logical structure – that is aligned to the UN Women standards – is correlated with reports that meet both the requirements for each parameter and achieve overall coherence. It is easy to include a draft report structure in any TOR as a means to improving this aspect.

More challenging will be to ensure that the evaluation frameworks and theories of change elaborated in the early stages of reports are referred back to in the development of conclusions and recommendations. The implication of not addressing this aspect of report writing is that the evidence developed in highly-rated findings section will not be used to its full potential in informing learning and accountability within UN Women.

8. Recommendations

As with the conclusions, the reviewer has 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.

Recommendation 1: Evaluation managers in UN Women can set their sights on performing well beyond the requirements of UN standards for gender and human rights; and systematically apply gender analysis to evaluation design, interpretation of findings, and development of conclusions.

The current performance in terms of evaluation report quality is generally robust, with a good number of exceptionally strong reports. In some regards, UN Women is advantaged in terms of the nature of its programmes and the questions that are being asked of them.

Building on this capability, UN Women could be content to deliver more highly rated evaluation reports in subsequent years. However, this can be achieved without making a significant impact on the extent to which evaluation processes – and information products (reports) – empower stakeholding groups (including rights holders).

UN Women can afford to, and need to, aim higher than the minimum standards, and learn from the best reports in its portfolio to push the boundaries of participation at all stages of the evaluation process. Evaluation commissioners in UN Women should thus continue to be guided by the UNEG standards, but also encouraged by the Executive Board and the Independent Evaluation Office to commit the resources and time to deepening the use of participatory, gender-responsive, and empowerment evaluation.

Recommendation 2: Evaluation commissioners will benefit from maintaining and strengthening the existing quality of Terms of Reference and recruitment of evaluators with expertise in gender responsive evaluation.

The current combination of TORs (that clearly identify evaluation criteria, questions, designs and methods) and recruitment of evaluation teams with credible GEWE experience appears to be delivering results in terms of report quality. This may be hypothesised as being because evaluators that are engaged with gender and human rights are more likely to be committed to professional development in others areas of evaluation competency.

Whilst there is no substantive to support such a claim, both practices are considered to be good in their own right, and it is, therefore, recommended to evaluation commissioners that they should remain focused on continuous improvement in the development of TORs and recruitment of evaluation teams. This should also be

supported as a key consideration in systems strengthening initiatives undertaken by the Evaluation Office.

Recommendation 3. The UN Women Evaluation Office will help fill a gap by providing simple model guidance on methodological limitations, ethical considerations and utilization approaches in mixed methods participatory evaluations.

The frequent use of a similar design in all of the decentralised evaluations is an advantage in terms of being able to more easily develop relevant guidance. Some of the weakest aspects of the overall portfolio of evaluations related to the discussion of limitations and ethical safeguards. Many of these will be common to evaluations across the organisation because of the homogeneity in the methods used.

It is therefore recommended to the Independent Evaluation Office to produce the following capacity-strengthening products written for non-specialist evaluation managers:

5. **For the ToR stage:** a brief on the wider range of evaluation designs that may be appropriate or adapted to UN Women's context (e.g. Collaborative Outcomes Reporting, Multiple Lines and Levels of Evidence, Systems Theory);
6. **For the inception stage:** a quick reference guide identifying the typical epistemological/ontological limitations and ethical issues faced by theory-based mixed method designs that rely on semi-structured interviews and focus group discussions (including the implications in interpreting data that should be clearly communicated to decision makers, and effective mitigating strategies);
7. **For the implementation stage:** model *free and informed consent* protocols and communications materials that can be used for focus groups and semi-structured or structured interviews;
8. **For the reporting stage:** a brief on ideas and a clear set of expectations for the dissemination, follow-up and utilisation of evaluation outputs (including reports).

Recommendation 4: By focusing on developing a learning community among evaluation managers in UN Women, the Independent Evaluation Office can maximise the proven capacity of non-specialist staff managers to deliver high quality evaluation reports.

The prevalence of good practice and innovation in evaluation throughout the UN Women universe of offices suggests that there is plenty of opportunity to share and enhance good evaluation practice among committed and effective evaluation managers, even where they are not evaluation specialists. It is, therefore, recommended that the Independent Evaluation Office explore opportunities to nurture the learning community for evaluation within UN Women to share experience and inspiration across regions.

Whilst there is limited research available on effective evaluation learning communities, there are some common elements of professional learning communities in other sectors that may offer some practical guidance. For example,

research into professional learning communities (PLCs) for educators[5] reveals eight important characteristics of effective PLCs:

Openness, networks and partnerships	Learning communities work best when they are open to participation from organisations and individuals outside their boundaries – suggesting the importance of linking with other initiatives such as MyM&E, Community of Evaluators, etc.
Collaboration focused on learning	Teaching members how to collaborate and supporting this with regular contact to jointly identify issues and cases to work through together.
Collective responsibility	All members of the community feel and exercise responsibility for the shared vision (e.g. quality evaluation). Peer pressure supports continuous development rather than relying on 'top-down' momentum.
Shared values and vision	Developing a strongly held focus on the core purpose of the learning community and developing a process for initiating new members into this vision.
Individual and collective professional learning	Whilst recognising the importance of individual learning, appreciating that most professional learning is collective – and creating spaces, events and opportunities to enable this sharing and dialogue with peers.
Reflective professional enquiry	Establishing habits and skills so that community members continuously analyse their experiences, performance data, and insights towards continuous improvement.
Inclusive membership	Not restricting an evaluation learning community to only evaluators/managers – although recognising that these may form the core group. Operating from the principle that support staff, and other technical or external experts are an important element of improving evaluation.
Mutual trust, respect and support.	Continuously nurturing a culture of peer-to-peer support, honesty, openness, and trust.

Such a community can celebrate innovation and experimentation as much as it does achieving UNEG standards. As a contribution to this knowledge community, it is also recommended that the Independent Evaluation Office continue to refine GERAAS, including enhancing the consistency of budget information available on GATE to allow for analysis of quality according to the resource-scale of evaluations and evaluands.

9. Lessons Learned

Whilst the data for this meta-analysis has been derived solely from the review of reports, it would appear that one lesson has emerged in relation to overall evaluation quality.

---

[5] http://www.educationscotland.gov.uk/Images/Creating%20and%20Sustaining%20PLCs_tcm4-631034.pdf

The overall achievement of 85% of reports satisfactorily meeting UNEG standards and strong performance in UN SWAP would appear to suggest that **designing evaluations with a focus on gender and human rights can also have wider beneficial effects** on other aspects of evaluation quality, even in situations where specialist resources are (relatively) constrained. This is *not* suggesting that gender-responsive evaluations are automatically better evaluations overall – merely that in real world evaluation the two factors (strong gender-responsiveness, strong report quality) are correlated.

A comparison between GERAAS overall ratings and UN SWAP scores reveals that Very Good reports also scored as either Meeting (50%) or Exceeding (50%) SWAP requirements. Similarly, reports rated Good also scored as either Meeting (66%) or Exceeding (33%) SWAP requirements, but with a slightly lower average. By comparison, Satisfactory reports mostly Met (43%) or were Approaching (43%) SWAP requirements (1 report Exceeded). Finally, Unsatisfactory reports were mostly Approaching (60%) SWAP requirements (1 report Met and 1 report Missed).

The hypothesis for explaining this correlation, made in the conclusions section, is that the recruitment of evaluation teams based on credible GEWE experience may be leading to self-selection of evaluators with a greater range of skills and commitment to professional development. If this is indeed the case, then UN Women's continued commitment to empowerment and human rights based approaches is likely to be advantageous to pursuing wider aspects of technical excellence in evaluation.

10. Annexes

10.1 Terms of Reference

Meta-Evaluation of 2013 UN Women Evaluation Reports
Terms of Reference (ToR) for Evaluation Consultant

| 1. | Introduction |

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 country, regional and global levels.

The year 2013 marked the coming into effect of a new UN-Women Evaluation Policy and approval of a new Strategic Plan (2014-2017). A landmark System-wide Action Plan (UN-SWAP) on gender equality and women's empowerment was also adopted in 2012 that requires annual reporting against a performance indicator on gender-response evaluation. These developments further underpinned the prominence of evaluation for enhanced organizational accountability, learning and evidence generation for decision-making. In light of this, UN Women has increased the capacity of the Evaluation Office, made considerable investment in evaluation-related activities and carried out several corporate and decentralized evaluations in recent years.

The Evaluation Policy, inter-alia, calls for supporting the implementation of the Strategic Plan (SP) through generation of a critical mass of high quality credible evaluations that provide useful evidence to accelerate achievement of gender equality and the empowerment of women. In congruence with this, the UN Women independent Evaluation Office has put in place a Global *Evaluation Report Assessment and Analysis System (GERAAS)* aligned with the UNEG Norms and Standards to ensure application of sound approaches and methods in order to continuously improve the quality and credibility of evaluation methods and reports.

Towards this end, two meta-evaluations have been conducted since 2011. These assessments not only helped to assess the quality and credibility of the evaluation reports but also provided an opportunity to generate lessons through synthesizing recurrent findings and recommendations useful for organizational effectiveness, decision-making and corporate knowledge management and learning.

The EO intends to continue its practice and is planning to undertake a meta-evaluation of all evaluations conducted and completed in 2013. The Meta-Evaluation will be carried out by an independent consultant.

| 2. | Purpose |

Specifically, the meta-evaluation will serve four purposes:

- *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;
- *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 gender responsive evaluations;
- *Improve UN Women's performance and organizational effectiveness***:** provide senior management with better understandings and insights into key UN women performance areas requiring attention through a synthetic approach; and
- *Promote learning and knowledge management:* help promote organizational learning and knowledge management through capturing experiences and lessons learnt from credible evaluations.

3.   Scope

The consultancy will cover all UN Women managed evaluation reports completed in 2013, including joint evaluations. The estimated total number of evaluation reports to be reviewed will be a minimum of 25 and a maximum of 30[6].

Evaluation is related to, but distinct from, other reviews and assessments. Reports meeting the below three criteria will be subject for the meta-analysis and synthesis.

1. Full-fledge evaluations (including Annexes), as opposed to reviews, evaluability assessments, or other kind of assessments according to criteria gathered in Annex I.
2. Corporate and decentralized evaluations of UN Women programmes uploaded in the Global Accountability and Tracking of Evaluation System by mid-January of the following year.
3. Joint programme evaluations in which UN-Women fully participates and had a decision making process.

4.   Methodology

The assessment of the quality of the evaluation reports will be done by an external independent consultant/s or consultancy firm and peer reviewed by the Evaluation Office to ensure rigor and validity of the reviews.

The assignment will have *three major components*. The first component is <u>*an assessment of the quality of UN Women evaluation reports*</u>. The second component is <u>*meta-analysis of evaluation report*</u> while the third component is <u>*meta-synthesis of the main findings, conclusions and recommendations including lessons learned presented in different evaluation reports*</u>.

---

[6] The exact number will be provided once all reports are posted in the Global Accountability and Evaluation Tracking System – GATE. Majority of reports are in English, but some are also available in Spanish, Portuguese and French.

A. _Assessment of the quality of individual evaluation reports._ This comprises assessment of the quality of each evaluation report and an executive feedback to be sent to the offices concerned. Each individual evaluation report is assessed using the UN Women-UNEG evaluation report standards. Based on the UNEG Quality Checklist for Evaluation Reports and UN SWAP Evaluation Scorecard, this tool 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) that are further defined by guiding points.

- Parameters and overall reports are rated against a 4-point scale: very good, good, satisfactory, and unsatisfactory.
- Constructive feedback is provided for each report to the commissioning office on how to improve future evaluation practice.

B. _Meta-analysis of evaluation report quality:_ The assessment will focus on the observed strengths, weaknesses, best practices and challenges for evaluation practice and based on

C. _Synthesis of the reports' findings, recommendations and lessons learned._ The meta-synthesis analyzes and synthesizes the findings, conclusions, recommendations and lessons learned presented in the evaluation reports. This helps to provide a basis for better understanding of UN Women operations and key areas requiring management attention.

5.   Expected Deliverables

The main expected deliverables of the exercise will be:

- Refinement of the UN Women Global Evaluation Report Assessment and Analysis Methodology in conformity with the UN SWAP and UNEG Methodology
- Recording of key report details that allow the report to be tracked
- A draft meta-evaluation report that includes:
    - Qualitative and scored feedback on the eight UN Women Adapted UNEG standards for each individual report
    - A meta-analysis
    - Meta synthesis
- Executive Feedback for each evaluation report and final products for the above three based on agreed outline with UN Women Evaluation Office.

6.   Management of the Evaluation

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

7.   Timeframe and Payment

A total payment of up-to $29,000 will be made on the completion and approval of individual review of 25 evaluation reports, meta-analysis and meta-synthesis by the respective deadlines. An additional $500 will be paid for any evaluation report to be added on top of the original 25 evaluation reports. In the same line, the total payment will be re-adjusted/reduced if the total number of reports turned out to be less than 25. The work is expected to begin in January 6, 2014 and to be finalized in February 28, 2014.

Timeframe for key deliverables and payment modalities would be as follows:

	Deliverables	Submission Deadline	Payment
1.	Refinement of the UN Women Global Evaluation Report Assessment and Analysis Methodology in conformity with UNEG and SWAP Methodology	January 15, 2014	-
2.	Recording of key report details that allow the report to be tracked	January 15, 2014	-
3	Draft individual evaluation report quality assessment and scored feedback on the nine UN Women Adapted UNEG standards for each individual report	February 20, 2014	12,500
4.	Draft meta-analysis and meta-synthesis reports		
5.	Final individual evaluation report quality assessment, meta-analysis and meta-synthesis reports, executive feedback for each evaluation report and power point presentation outlining the key findings	February 28, 2014	12,500
			25,000*

- the payment may be adjusted accordingly up to a maximum of $29,000 based upon the number of reports approved and received

8.   Required Competencies and Skills

As part of the recruitment process, potential consultants will be asked to do a qualitative and quantitative assessment of one evaluation report based on UN Women Evaluation Quality Assessment Tool. The exercise will serve as a basis to determine the best candidate for this particular consultancy.

In general, the consultant should have

- At least a Master's degree related to any of the social sciences, preferably gender or evaluation
- At least 10 years of working experience in evaluation and previous experience in designing and conducting meta-evaluations

- Experience and background in gender equality/gender analysis and gender responsive evaluations
- Demonstration of strong analytical and communications skills
- Ability to work with the UN Women Evaluation Office to produce high quality report delivered in a timely basis
- Experience working with the UN; knowledge of UN Women context a strong asset
- Familiarity with United Nations Evaluation Group (UNEG) Norms and Standards
- Fluent in English and working language of Spanish, French and Portuguese an asset

9.   UN Women Evaluation Office Contact

Messay   Tassew: Evaluation   Specialist   –   messay.tassew@unwomen.org

10.2 List and Ratings of Evaluation Reports Reviewed

Country	Title Of The Evaluation Report	Overall rating
Corporate (HQ)		
Global	The Contribution Of UN Women To Prevent Violence Against Women And Expand Access To Services	Very Good
Global	The Contribution Of UN Women To Increasing Women's Leadership And Participation In Peace And Security And In Humanitarian Response	Very Good
Global	Joint Evaluation Of Joint Programmes On Gender Equality In The United Nations System	Very Good
Eastern and Southern Africa		
Ethiopia	Ethiopia Joint Flagship Programme On Gender Equality And Women's Empowerment (JP GEWE)	Good
Mozambique	AVALIAÇÃO Das Intervenções Da ONU Mulheres Na Área De Prevenção E Eliminação Da Violência Contra A Mulher, Com Enfoque Na Mobilização Social No Âmbito Da Campanha Unidos Pelo Fim Da Violência Contra A Mulher E Rapariga	Satisfactory
Rwanda	Final Evaluation Of Rwandan Government And ONE UN ISANGA One Stop Centre, Final Report	Satisfactory
Uganda	Mid Term Evaluation Of UN JOINT PROGRAMME ON GENDER EQUALITY (UNJPGE) --UGANDA	Good
Zimbabwe	End Of Project Evaluation Of The Gender Support Programme (GSP)	Good
Burundi	Evaluation Finale Du Programme Pays 2010-2013 D'onu FEMMES Au Burundi	Unsatisfactory
Kenya	Evaluation Of Gender And Governance Programme, Kenya	Satisfactory
DRC, Rwanda, Burundi	Evaluation Of UN Women Partnership With The International Conference On The Great Lakes Region (ICGLR)	Very Good
Western and Central Africa		
Mali	Programme De La Promotion De La Justice Pour Les Femmes Dans Le Gouvernance Democratique Au Mali	Unsatisfactory
Liberia	Evaluation Of The UN / Government Of Liberia Joint Programme On Gender Equality And Women's Economic Empowerment (JP-GEWEE)	Good
Arab States		
oPt	The Mehwar Centre Final Project Evaluation	Satisfactory
Asia and the Pacific		
Nepal	Final Evaluation Of Project On Making Politics Work With Women (MP3W) In Nepal	Good
India	Mid-Term Evaluation Of UN Women's Anti-Human Trafficking Programme	Unsatisfactory
Nepal	Final Evaluation Of Sustaining The Gains Of Foreign Labour Migration Through The Protection Of Migrant Workers' Rights Programme	Good
Afghanistan	External Evaluation Of The EVAW Commission Project	Good
Europe and Central Asia		
Georgia	Final External Formative Evaluation Of The Project "Women For Equality, Peace And Development In Georgia"	Very Good
Tajikistan, Pakistan, Georgia	Women Connecting Across Conflict Final Evaluation	Good
Albania, BiH, FYR Macedonia	Final Evaluation Of The UN Women Project "Promoting Gender Responsive Budgeting In South East Europe"	Very Good

Serbia, Montenegro	Final Project Evaluation – UN Women: Advancing Women's Economic And Social Rights In Serbia And Montenegro	Very Good
Bosnia and Herzegovina , Serbia, fYR Macedonia, Kosovo	UN Women Project: Advancing The Implementation Of UNSCR 1325 On Women Peace And Security In The Western Balkans	Satisfactory
Latin Americas and Caribbean		
Argentina, Brasil, Uruguay y Paraguay	Evaluación Final Programa Mujeres Jóvenes ONU Mujeres	Satisfactory
Mexico	De Lo Local A Lo Global: Políticas Públicas Para El Empoderamiento De Las Mujeres Indígenas Evaluación Del Acuerdo De Colaboración ONU-CDI Mujeres, En El Marco Del Programa De Acciones Para La Igualdad De Género Con Pueblos Indígenas	Unsatisfactory
Grenada	Evaluation of Grenada's Man-To-Man Batterer Intervention Programme	Satisfactory
Guatemala	Final Evaluation of The Safe Cities Programme	Satisfactory

10.3 Report Review Format

UN WOMEN Global Evaluation Reports Assessment and Analysis System (GERAAS) – Including UN SWAP

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.				
Evaluability of the GEEW aspects of the intervention is assessed and steps/measures are taken to maximize the evaluability of GEEW aspects. *The evaluability assessment should determine whether an intervention is ready to be evaluated. It should also determine whether 1) its GEEW aspects can be evaluated or not (i.e. the evaluability of the GEEW aspects of the intervention are identified and determined to be either high, medium or low) and 2) identify and implement the measures needed to address/maximize the evaluability of the GEEW aspects. More specifically, the evaluation design documents (i.e.ToR/Inception report) assess the evaluability of the GEEW aspects of the intervention and identify evaluation related challenges and mitigating measures.*				
A gender- responsive Stakeholder Analysis is undertaken. *1) The stakeholder analysis identifies who the different groups in an intervention are, why, how and when they should be included in the evaluation process. 2) Direct and indirect stakeholders of the intervention are identified with an aim to ensure that representatives of women (women's groups), men, boys and girls from different groups (e.g. beneficiaries, implementers, rights- holders, and duty- bearers) are identified and their level of participation in the evaluation is determined.*				
PARAMETER 2: PURPOSE, OBJECTIVES AND SCOPE			**Guiding Question**	**Constructive**

GUIDING POINTS/EXPLANATORY NOTE	RATING	Remarks	Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation?	feedback for future reports Including how to address weaknesses and maintaining good practice
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.				
GEEW is integrated in Evaluation Criteria. *GEEW dimensions are integrated into all Evaluation Criteria and/or criteria derived directly from GEEW principles are used (e.g. equality, participation, social transformation, inclusiveness, empowerment, etc.).*			⬛	
Evaluation Questions are included that specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved.				
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				
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.				
A gender- responsive Evaluation Approach and Methodology is selected. *The method and approach are the overarching framework of an evaluation that will determine what methodologies to employ and how to use the tools in gender-responsive fashion, e.g., a mixed-method approach to make visible diverse perspectives and promotes participation of both women and men, boys and girls from different stakeholder groups; Triangulation of data is done to ensure that the voices of both women, men, boys and girls are heard and used; additional time or resources (time, staff, funds) to implement a gender-responsive approach is considered and planned for, etc.*			⬛	
Evaluation indicators are included to ensure GEEW- related data is collected. *1) Indicators for the evaluation of the intervention include the GEEW dimension and/or additional indicators are identified specifically addressing GEEW. 2) Inclusion of mixed indicators (including quantitative and qualitative indicators); availability of disaggregated data.*				
Data Collection Methods and Tools are developed to collect GEEW- related data, deployed in a gender- responsive manner, and follow ethical codes of conduct. *Data collection methods including, desk reviews, focus groups, interviews, surveys, etc. are identified and accompanying tools, e.g. questionnaires, observational tools, interview guides etc. developed integrating GEEW considerations (e.g. interview guides ensure that women and men are interviewed in ways that avoid gender biases or the reinforcement of gender discrimination and*				

GUIDING POINTS/EXPLANATORY NOTE	RATING	Remarks	Guiding Question	Constructive feedback for future reports
unequal power relations, etc.)			████████	
Data Analysis techniques were deployed that include gender analysis. *Special attention is paid to data and information that specifically refer to GEEW issues, and making the best possible use of these in the overall assessment of the intervention.*			████████	

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.				
The Evaluation Report includes findings, conclusions, recommendation s and lessons on the extent to which 1) the design and implementation of the intervention was gender responsive 2) GEEW results were achieved and 3) a gender- responsive methodology used.*Conducting a gender responsive process refers to ensuring that the main steps in conducting an evaluation in a gender-responsive fashion be included and described in the evaluation report. The 'methodological' section in the evaluation report should reflect the methods employed and describe the appropriateness for analyzing gender in the evaluation scope (i.e methodology, data collection and analysis tools integrating a gender lens are described in and annexed to the report). The evaluation reports should also provide lessons/challenges/recommendations for conducting gender-responsive evaluation based on the experience of that particular evaluation.*			████████	
The Validation Process of evaluation findings, conclusions and recommendation s includes the participation of women, men, boys and girls from the stakeholder groups identified and particularly include the voice of women's groups. *At the end of the data collection stage, all stakeholder groups are consulted to discuss findings and hear their views on conclusions and recommendations of the evaluation.*			████████	

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 presented in a priority order?**	Including how to address weaknesses and maintaining good practice
Recommendations				
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.				

GUIDING POINTS/EXPLANATORY NOTE	RATING	Remarks	Guiding Question	Constructive feedback for future reports
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
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?	Including how to address weaknesses and maintaining good practice
GENDER AND HUMAN RIGHTS				
7.1 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.				
7.2 The report uses gender sensitive and human rights-based language throughout, including data disaggregated by sex, age, disability, etc.				
7.3 The evaluation approach and data collection and analysis methods are gender equality and human rights responsive and appropriate for analyzing the gender equality and human rights issues identified in the scope.				
7.4 The report assesses if the design of the object was based on a sound gender analysis and human rights analysis and implementation for results was monitored through gender and human rights frameworks, as well as the actual results on gender equality and human rights.				
7.5 Reported findings, conclusions, recommendations and lessons provide adequate information on gender equality and human rights.				

PARAMETER 8: THE REPORT STRUCTURE			Guiding Question	Constructive feedback for future reports
GUIDING POINTS/EXPLANATORY NOTE	RATING	Remarks	Is the report well structured, logical, clear and complete?	Including how to address weaknesses and maintaining good practice
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				
Evaluation Team is recruited that has the capacity to conduct gender-responsive evaluation. *The Evaluation Team has: 1) GEEW expertise with clear assigned responsibilities and adequate seniority; 2) is gender balanced; 3) is culturally diverse and/or makes use of regional/national evaluators .*		■	■	

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				
The Management Response to the evaluation addresses the GEEW issues raised in the report and is developed in consultation with a diverse group of stakeholders who have an interest in and/or are affected by GEEW issues.		■	■	
The Evaluation Dissemination Strategy is gender responsive. *Dissemination of evaluation findings on GEEW issues to diverse group of stakeholders who have an interest in and/or are affected by GEEW issues (including women's groups, networks and individuals) in accessible formats*		■	■	

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.				
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.	ToRs			
	Other			

10.4 Performance Graph of all GERAAS Criteria

Legend: ■ Unsatisfactory ■ Satisfactory ■ Good ■ Very Good

Criteria (top to bottom):
- 1.1 Object Logic
- 1.2 Context
- 1.3 Scale and Complexity
- 1.4 Stakeholders
- 1.5 Implementation
- 2.1 Purpose
- 2.3 Objectives
- 2.3 Scope
- 2.4 Critera
- 2.5 Gender and Human Rights
- 3.1 Methods
- 3.2 Data Collection
- 3.3 Data Sources
- 3.4 Sampling
- 3.5 Stakeholders
- 3.6 Data Quality
- 3.6 Gender and Human Rights
- 3.7 Ethics
- 4.1 Findings
- 4.2 Analysis
- 4.3 Respond to Criteria
- 4.2 Based on Evidence
- 4.5 Gaps and Unexpected
- 4.6 Reasons and Implications
- 5.1 Insights
- 5.2 Reasonable Judgement
- 5.3 Substantiaed by evidence
- 5.4 Identify solutions
- 5.5 Strengths and Weaknesses
- 5.6 Lessons Learned
- 6.1 Supported by Evidence
- 6.2 Process of development
- 6.3 Relevant to purpose
- 6.4 Targeted
- 6.5 Prioritised
- 6.6 Actionable
- 7.1 Incorporation of Gender and HR
- 7.2 Language and Disaggregation
- 7.3 Approach and Analysis
- 7.4 Design of Object
- 7.5 Included in Findings
- 8.1 Logically Structured
- 8.2 Opening Pages
- 8.3 Executive Summary
- 8.4 Annexes