UN Women
GERAAS 2014

Global Evaluation Reports Assessment and Analysis System

META EVALUATION REPORT

Version 1.2

30 March 2015

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

Prepared by ImpactReady LLP
Joseph Barnes, Sara Vaca

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 evaluation reports from 2014 that were submitted to UN Women's Global Evaluation Reports Assessment and Analysis System (GERAAS). GERAAS is an 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.

21 evaluation reports from across the UN Women regions HQ were included in the analysis, representing at least 51 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 39 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 caution should be exercised when using the findings or recommendations 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 four UN SWAP Evaluation Performance Indicators[1].

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

---

[1] This is an updated system from the 13 indicators used in GERAAS 2013. As a result, the UN SWAP scores for 2013 and 2014 are broadly but not directly comparable.

of analysis to fully understand utilization (such as follow-up interviews) is currently considered prohibitive.

Reviews were undertaken by a team of independent consultants with prior experience of evaluation report quality assurance for UN agencies and donors. A process was established to ensure consistency of ratings, both within GERAAS 2014 and to allow comparison with GERAAS 2013.

2.2 Findings

The review process found that 100% of evaluation reports could be considered as satisfactory or above according to UN Women standards: an significant improvement of 15 points. Whilst reports still have opportunities for improvements, 43% were found to be very good – exceeding UN Women Standards in multiple parameters.

Figure i: Overall ratings
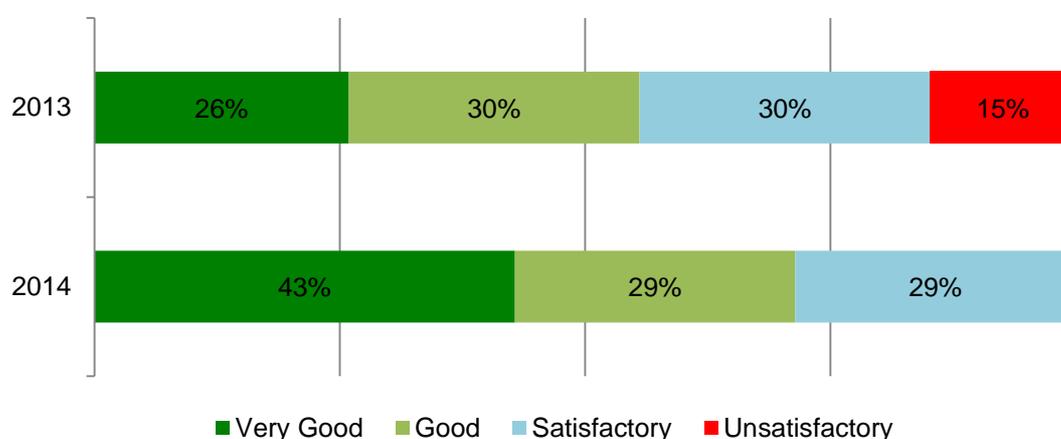
| | Very Good | Good | Satisfactory | Unsatisfactory |

The most evaluation reports were submitted from the Eastern and Southern Africa region (6) and the Asia Pacific region (5), the least (1) from Europe and Central Asia region and West and Central Africa region. The highest concentration of *very good* evaluations were in Asia Pacific (3), Corporate/HQ (3) and Latin America and the Caribbean (2). Eastern and Southern Africa, with 6 evaluations submitted, also included one *very good* report. Evaluation reports in the other regions were all rated either *good* or *satisfactory*.

The majority (38%) of evaluations were national level evaluations, or regional evaluations (29%). Global evaluations averaged highest, followed by region, then national and – lastly – sub-national evaluations.

Compared to the previous year, there were fewer joint evaluations. UN Women directly managed 90% of the evaluation reports, an increase of 23 points from GERAAS 2013. Many Joint Evaluations may have been scheduled for 2015 (to align with the MDGs).

Almost half (48%) of the reports related to programme evaluations, 29% to projects and 10% each to policies and pilots. There was only one country programme

evaluation[2]. This is very similar to GERAAS 2013. The majority of evaluations (67%) were final/summative in nature, with the purpose of assessing programme/project results.

Reports rated strongest in relation to recommendations (90% good or above) and findings and report structure (both 81% good or above). The next best performing parameter was describing the object of the evaluation (76% good or above). The weakest parameter was gender and human rights, with 52% of reports meeting or exceeding UN SWAP standards. The methods sections were also an area requiring attention, with 67% of reports rated as good or above.

Figure ii: Parameters rated good or very good
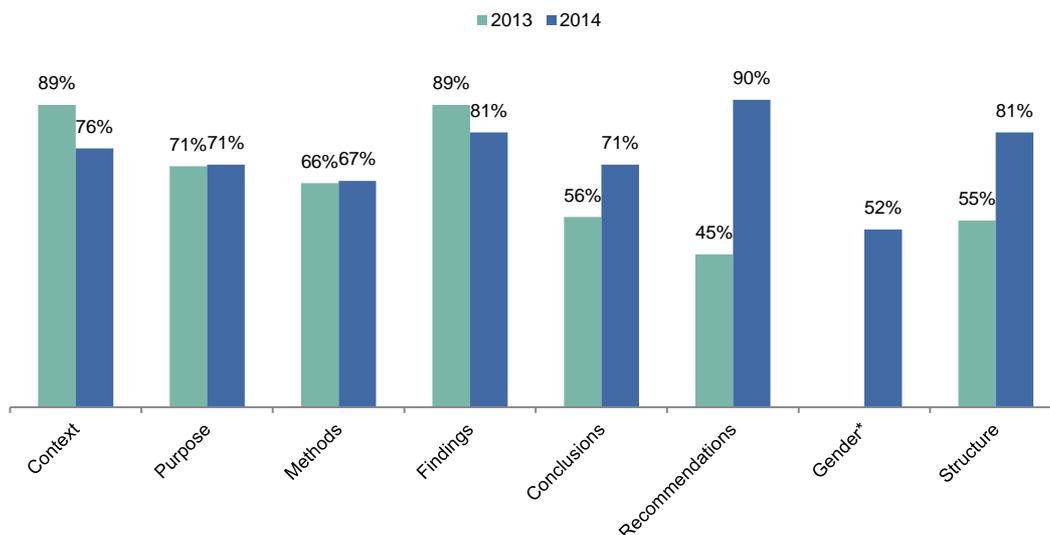
In relation to gender, reports were strongest in regard to gender responsive evaluation frameworks (including gender responsive indicators) and methods. The greatest improvement is available to be made in relation to gender analysis and the scoping of evaluations to consider power issues. It was also observed that few reports included gender or human rights as specific objectives, with most purpose and objectives statements focusing on institutional or organisational issues

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

☐ Several reports include details on **evaluation processes that extended participation** beyond the data collection stage. For example, holding a participatory workshop on conclusions and recommendations;

☐ Some reports were noticeable in terms of the use of **innovative or good-practice methods**, including mixed methods, theory-based designs and most significant change;

☐ A number of evaluation reports included a **reconstruction of the theory of change** where no programme logic was available;

---

[2] UN Women Independent Evaluation Office is currently developing Country Portfolio Evaluation guidance in response to recommendations made by the UNEG review of the evaluation function.

☐ One report included a table in the annexes explaining the **achievement of objectives** with traffic lights;

☐ **Gender analysis** is a key indicator under UN SWAP. Some reports managed to take this further than others, including gender-based breakdowns of data.

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: There is an overall improvement since 2013, throughout the decentralised evaluation function, with fewer but more consistent evaluation reports. Reports are increasing structured according to UNEG guidance, with strong findings and recommendations sections that rate well because they align strongly to the evaluation criteria/framework. Conclusions sections are also aligned to evaluation criteria; but are most often more syntheses of the findings, rather than taking the analysis deeper. Other areas that emerged as requiring focus include: placing more material about evaluation objects in the background section and leaving more space for analysis in the findings; supporting evaluation within the context of peace and security; and strengthening methods and gender responsiveness within the context of UN Women's macro, policy, normative and coordination work (see Conclusion 3).

Conclusion 2: The homogenuity of evaluation designs largely remains, with a few examples emerging of mixed methods approaches. A case exists for standardising different types of evaluation. GERAAS 2013 proposed that evaluation designs were largely similar across the portfolio (qualitative designs that included limited quantitative data). Preponderance for this type of evaluation still exists, although a number of serious attempts to execute mixed methods designs were included in GERAAS 2014. There is a case for strengthening use of mixed methods at all stages of evaluations, for clearly separating impact evaluations from outcome-level evaluations, and for exploring the use of developmental evaluation. Focusing on a few distinct evaluation 'types' would allow for greater standardisation of guidance. For example, model ethics statements and implementation arrangements.

Conclusion 3: The highest priority for future quality assurance is ensuring that gender responsive designs, analysis and reporting techniques are accessible and practiced throughout the decentralised evaluation function. In order to improve the rating of reports, it is necessary to further strengthen gender integration into the scoping, specification of methods, indicators, analysis and evaluation process. Despite increased availability of general guidance, there are a number of issues where more standardised guidance and tools from IEO could help to fill a gap: how to scope an evaluation; a 'default' stakeholder mapping and sampling tool; and a standalone criterion for Gender and Human Rights.

Conclusion 4: Participatory approaches are a common feature of UN Women evaluations at the data collection stage. Considerable scope exists for

enhancing participation at all other stages; and for emphasising empowerment alongside utilisation. All of the evaluations cited participatory approaches in the methods section. In general, participation was primarily concerned with the data collection stage: with affected groups variously consulted as part of the social enquiry process. The evidence of evaluations including participation at the design and analysis stage was far more limited. Very little information is available in reports in regard to feeding back the outcomes of evaluations to rights holders. Whilst it is unrealistic to conclude that participation can be simply or easily increased at all stages of future UN Women evaluations, there is a strong case for adaptation of GERAAS 2015 to provide more evidence and attention on this issue.

2.4 Recommendations

As with the conclusions, the reviewers have developed these recommendations based on the findings and conclusions, rather than a participatory process. They can be expected, therefore, to be subject to further validation.

Recommendation 1: For aspects of evaluations that can be standardised, IEO should consider developing simple checklists of minimum or default standards for all UN Women managed evaluations.

Urgency	Impact	Ease	Priority
★★★☆☆	★★★★☆	★★★☆☆	★★★★☆

Checklists – sometimes referred to as 'cheatsheets' – are frequently used in high-stress situations to ensure that professionals (such as pilots, medical staff or disaster responders) do not miss important details. It is recommended that checklists are piloted for issues that can most easily be standardised across evaluation reports. Given the ratings in GERAAS 2014, potential priority areas for checklists are: scoping the evaluation; evaluation indicators that include rubrics of success; mapping stakeholders and sampling; implementing UNEG guidance on ethics; exploring unexpected findings; and referencing evidence in report writing.

Recommendation 2: Continue existing efforts to strengthen Terms of Reference and recruit evaluators with expertise in gender responsive evaluation.

Urgency	Impact	Ease	Priority
★★★★★	★★★★☆	★★★★★	★★★★★

GERAAS 2013 identified Terms of Reference and gender-sensitive evaluation teams as key contributing factors to report quality. GERASS 2014 findings support the continuation of efforts to strengthen these aspects. Six specific actions identified by the meta-evaluation as contributing to further enhancing Terms of Reference are: strengthen focus on the design of evaluations; separate evaluations of impact from evaluations of the other OECD-DAC criteria; include requirement for evaluations to explicitly reference CEDAW concluding observations; ask the question of whether the evaluand was/is the *most relevant option* for UN Women; include a preliminary stakeholder map and gender/human rights role analysis; and specify a standalone criteria on *gender and human rights*.

Recommendation 3: Provide practical guidance and examples of gender analysis of policy, macro, coordination and normative work.

Urgency	Impact	Ease	Priority
★★★★★	★★★★☆	★★☆☆☆	★★★★☆

The meta-evaluation has identified a number of challenges in implementing the UN SWAP standards within the context of UN Women. It is recommended, therefore, that any forthcoming guidance – such as the Country Portfolio Evaluation guidance – should include very practical tools (with examples) that make gender analysis accessible within the decentralised contexts of UN Women.

Recommendation 4: Update GERAAS 2015 to highlight opportunities for participation at all stages of evaluation.

Urgency	Impact	Ease	Overall
★★★★☆	★★☆☆☆	★★★☆☆	★★★☆☆

Based on the qualitative observation that participation in UN Women evaluations is primarily concentrated around the data collection stage (and thus requires strengthening), it is recommended that an additional analysis of participation be included within GERAAS 2015. The proposed approach would involve the inclusion of a 'participation scan' to map the levels of participation of different stakeholders at different stages of each evaluation. Combined with a similar 'methods scan', this would allow the meta-evaluation to give a richer picture of participation within the organisation, and provide a stronger basis for recommendations to improve this.

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-evaluation and meta-analysis of 2014 evaluation reports submitted to GERAAS, including the UN-SWAP scores[3]. A total of 21 reports were included.

---

[3] ImpactReady LLP – see annexes for profiles of reviewers

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 2014 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 currently considered outside the scope of this analysis. It should be recognised, therefore, that this report only provides a partial view in answer to the question *"what is the quality of evaluation in UN Women?"*

5. Methodology

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

This report assesses final evaluation reports uploaded in the UN Women Global Accountability and Tracking of Evaluation System (GATE) by January 2015. IEO undertook an initial screening of all reports uploaded in the GATE system to a) decide whether the report can be classified as an evaluation as per UNEG definition and b) to ensure that the evaluation is managed or jointly managed by UN Women. The independent assessor undertook a secondary screening.

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

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

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

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:

Clear and full description of the 'object' of the evaluation	The report describes the object of the evaluation including the results chain, meaning the 'theory of change' that underlies the programme being evaluated. This theory of change includes what the programme was meant to achieve and the pathway (chain of results) through which it was expected to achieve this. The context of key social, political, economic, demographic, and institutional factors that have a direct bearing on the object is described. For example, the partner government's strategies and priorities, international, regional or country development goals, strategies and frameworks, the concerned agency's corporate goals and priorities, as appropriate.
The evaluation's purpose, objectives and scope are fully explained	The purpose of the evaluation is clearly defined, including why the evaluation was needed at that point in time, who needed the information, what information is needed, how the information will be used. The report provides a clear explanation of the evaluation objectives and scope including main evaluation questions and describes and justifies what the evaluation did and did not cover. The report describes and provides an explanation of the chosen evaluation criteria, and/or other criteria used by the evaluators.
Appropriate and sound methodology	The report presents transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve evaluation purposes. The report presents a sufficiently detailed description of methodology in which methodological choices are made explicit and justified and in which limitations of methodology applied are included. The report gives the elements to assess the appropriateness of the methodology. Methods as such are not 'good' or 'bad', they are only so in relation to what one tries to get to know as part of an evaluation. Thus this standard assesses the suitability of the methods selected for the specifics of the evaluation concerned, assessing if the methodology is suitable to the subject matter and the information collected are sufficient to meet the evaluation objectives.
Findings, conclusions, recommendations and lessons learned are based on evidence and sound analysis	*Findings* respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report. They are based on evidence derived from data collection and analysis methods described in the methodology section of the report. *Conclusions* present reasonable judgments based on findings and substantiated by evidence, providing insights pertinent to the object and purpose of the evaluation. *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.
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.

5.2 Meta-evaluation

The review process consists of five main parts. There were two 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) responding to comments/challenges by the Independent Evaluation Office.

One report had been produced by ImpactReady LLP – the authors of this meta evaluation. In order to avoid a conflict of interest, an independent evaluator, EvalQuality, was commissioned to undertake that review through direct dialogue with the Independent Evaluation Office.

PART 1: BASIC INFORMATION AND CLASSIFICATION OF THE EVALUATION
The Evaluation Assessment Matrix is a simple tool designed to capture or provide a snap shot of the key aspects of the evaluation and the evaluation report. This comprises basic information such as title, region/country, type, costs, geographic and thematic coverage, stage/timing and management of the evaluation.

PART 2: ASSESSMENT OF THE QUALITY AGAINST EACH OF THE EIGHT 'PARAMETERS'
The final review template is composed of *8 Parameters* (Object and Context of Evaluation; Evaluation Purpose, Objectives and Scope; Evaluation Methodology; Findings; Conclusions and Lessons Learned; Recommendations; Gender and Human Rights Considerations; and the report structure). The Eight Parameters are further defined by 39 Guiding Points.

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

Figure 1: Meta-evaluation process

Filtering by EO	
IEO checks that all submitted reports are evaluations	0 reports removed from the review as considered outside the scope of 2014

Filtering by Independent Reviewer	
Reviewer checks that all submitted reports are evaluations	No reports removed

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

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

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

Meta-evaluation draft report	
Quantitative analysis of report ratings using Excel	**Qualitative analysis of review comments matched to quantitative results**

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

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

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

Very good
Good
Satisfactory
Unsatisfactory

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

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

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

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

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

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

To ensure consistency in the rating of each parameter and the overall report, the reviewer undertook the review based on what is written in the evaluation report. In line with GERAAS specifications, methodology, findings, conclusions and recommendations were given more prominence to inform the overall rating. These are considered by UN Women to be the bedrock of a good quality report.

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

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

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

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

5.3 Changes Made From Previous Years
The main change from GERAAS 2013 is the replacement of the previous guiding points for *Parameter 7: Gender and Human Rights* with the four UN SWAP indicators specified in the revised guidance provided by UNEG. Furthermore, the rating for *Parameter 7* has been changed from a qualitative assessment to the quantified scoring required for UN SWAP reporting. This completes the integration of the UN SWAP and GERAAS assessment processes.

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 4 UN SWAP Evaluation Performance Indicators:

Gender mainstreaming	Score
Fully integrated	3.0
Satisfactorily integrated	2.0
Partially integrated	1.0
Not at all integrated	0.0

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

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

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

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 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 focusing only on 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-evaluation 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.

6. Findings

6.1 Overall Ratings and Feedback

The review process found that 100% of evaluation reports could be considered as satisfactory or above according to UN Women standards[4]. This is a significant increase of 15 points compared to GERAAS 2013, and is a major achievement. Whilst reports still have opportunities for improvements, 43% were found to be Very Good – exceeding UN Women Standards in multiple parameters. Once again, this represents a major improvement.

Figure 2: Overall ratings for all reviewed reports, 2013-2014
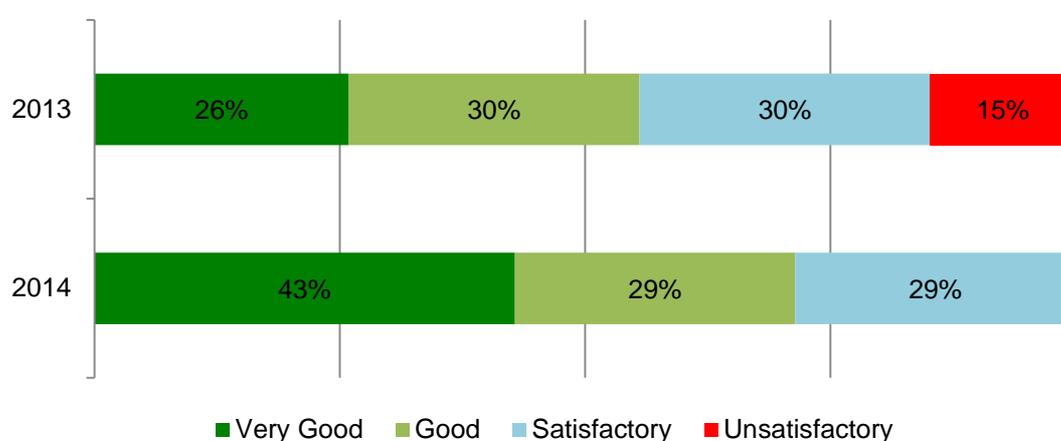
Figure 3: Numbers of reports with overall ratings for GERAAS 2014

The review process revealed a number of patterns in the evaluation reports that were considered:

1. As with GERAAS 2013, most 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 means that the overall portfolio continues to rely on one type of evidence and its associated limitations;

2. Only two of the evaluations included designs that were deemed sufficient to assess impact (as defined by UNEG). Despite this, impact was included as a

---

[4] This implies that decision makers could be reasonably confident in their findings, insights and recommendations.

criterion in many Terms of Reference. In response to this, several reports adapted the Terms of Reference to discuss 'progress towards' impact. There is, however, scope to consider whether evaluations of impact and the other OECD-DAC criteria should be commissioned separately;

3. When assessing the UNEG parameter of 'object and context' the reviewers found that a great deal of this information (political, social, gender and economic context; stakeholders, roles and contributions; and programme budget and implementation status) was placed in the Findings section rather than the Introduction. It seems that many 'facts' about evaluands are being classified as findings because they are hard for the evaluation team to identify (because of lack of records, etc). However, they are – in essence – just a synthesis of the current status of evaluands rather than new insights. The implication of this is that space for analysis and additional insights within the findings section is reduced.

6.2 Overall Regional Trends

The most evaluation reports were submitted from the Eastern and Southern Africa region (6) and the Asia Pacific region (5), the least (1) from Europe and Central Asia region and West and Central Africa region.

Figure 4: Number of overall reports and ratings per region
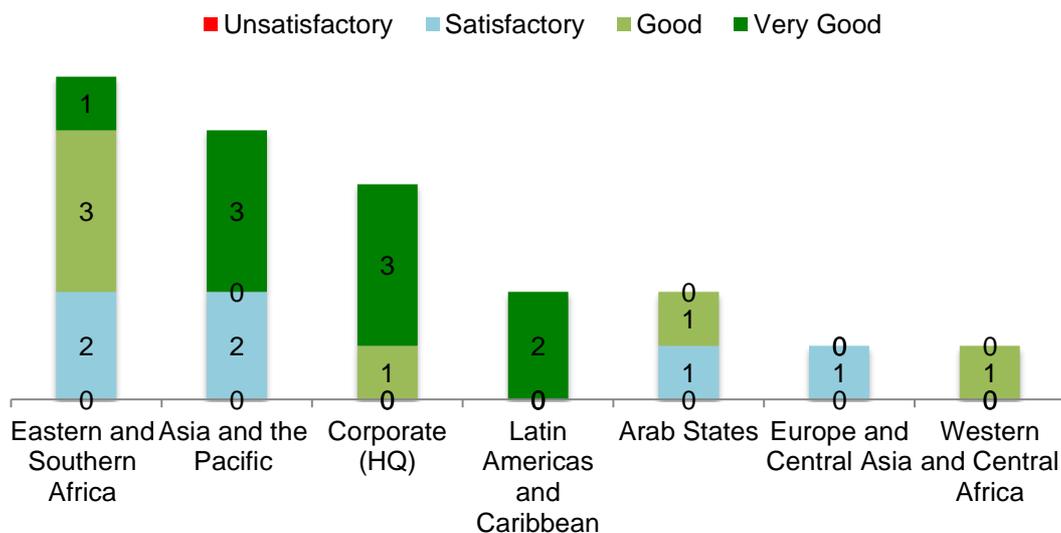
The highest concentration of *very good* evaluations were in Asia Pacific (3), Corporate/HQ (3) and Latin America and the Caribbean (2). Eastern and Southern Africa, with six evaluations submitted, also included one *very good* report. Evaluation reports in the other regions were all rated either *good* or *satisfactory*.

6.3 Trends by Type and Scope of Evaluation

6.3.1 Geography

The majority (38%) of evaluations were national level evaluations, or regional evaluations (29%). The average (mean) quality of reports – according to UNEG standards – was correlated with the geographic scope of the evaluation: global evaluations averaged highest, followed by region, then national and – lastly – sub-national evaluations.

Pending access to detailed evaluation budget data, it is not possible to triangulate whether the level of resources invested in an evaluation follows the same pattern. However, this meta-evaluation posits that a logical thought experiment allows us to hypothesise that the level or resourcing is connected to the level of decentralisation/evaluation-scope, and so is likely to be a strong factor in explaining this pattern of performance.

6.3.2 Management of Evaluation

Compared to the previous year, there were fewer joint evaluations. UN Women directly managed 90% of the evaluation reports, an increase of 23 points from GERAAS 2013.

The only two jointly managed evaluations were in Eastern and Southern Africa, both at the national level. One of these rated satisfactory and one rated good. All Very Good reports were managed directly by UN Women. In all regions (4) that included very good reports, these were spread fairly evenly across all levels (from sub-national to global), suggesting that the geographic level of an evaluation is not a strong determining factor of the quality of its report. There was no clear preference for any geographic scope of evaluations between the different regions.

6.3.3 Type of intervention evaluated

Almost half (48%) of the reports related to programme evaluations, 29% to projects and 10% each to policies and pilots. There was only one country programme evaluation[5]. This is very similar to GERAAS 2013.

Programme evaluations (10) were largely focused at outcome level results (9), and most of these were summative final evaluations (7). Project evaluations (6) were more focused on assessing outputs (4), with a balance between formative and summative purposes. Only one policy evaluation considered impacts.

Half of project evaluations rated *good* (17%) or *very good* (33%), and 90% of programme evaluations were rated at least *satisfactory* (with 70% *good* or above). All of these ratings represent a year-on-year improvement.

6.3.4 Stage

The majority of evaluations were final/summative in nature, with the purpose of assessing programme/project results. These reports focused on accountability to donors, with learning being framed in terms of similar future programmes. By comparison, the 7 formative/mid-term evaluations tended to focus on learning in order to improve an on-going programme or project (with accountability as a secondary consideration).

---

[5] UN Women Independent Evaluation Office is currently developing Country Portfolio Evaluation guidance in response to recommendations made by the UNEG review of the evaluation function.

6.3.5 Results/level of changes

Two thirds of evaluations (67%) primarily assessed outcomes. The most common evaluations were national-level outcome evaluations (6) and regional outcome evaluations (5). 75% of sub-national evaluations (3/4) were at the output level. There was only one impact-level evaluation: the Corporate Thematic Evaluation of WEE.

The most common type of evaluation in GERAAS 2014 were final evaluations of programme outcomes, constituting 33 percent of reports.

There was a strong correlation between outcome evaluations and high quality reports, with 86 percent rated as good (36 percent) or very good (50 percent). By comparison, only 33 percent of output level evaluations were rated as good or above.

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

6.3.6 Strategic Plan Correspondence

Women's participation was a major focus of GERAAS 2014. In total, 18 reports considered economic empowerment (10) or political participation (8) interventions. Governance was also a major focus, with 8 reports considering Strategic Priority 5. The least number of reports focused on women in peace and security (4) and translating global norms (5). Seven evaluations focused on ending violence against women.

Figure 5: Distribution of reports considering Strategic Plan Impact Areas
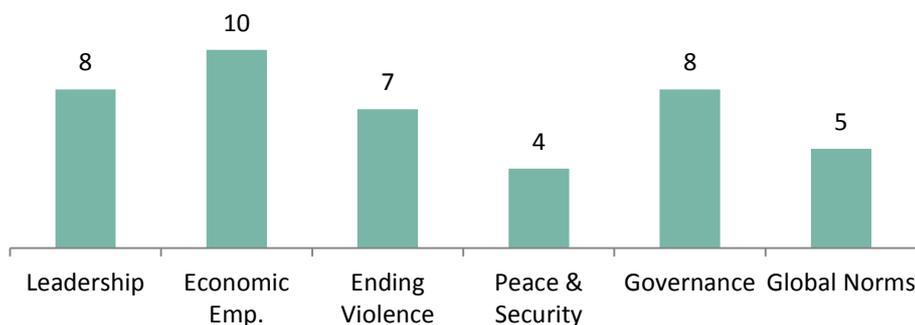
Proportionally, evaluations relating to governance (Strategic Priority 5) and ending violence against women (SP 3) were the strongest, with 88 percent and 86 percent (respectively) rating as good or above. By comparison, evaluations of peace and security (SP4) and translating global norms (SP6) achieved ratings of 50 percent and 60 percent (respectively) as good or above. Evaluations of political (SP1) and economic (SP2) empowerment rated midway between these levels.

Figure 6: Percentage of reports rated as Good or above for strategic goals
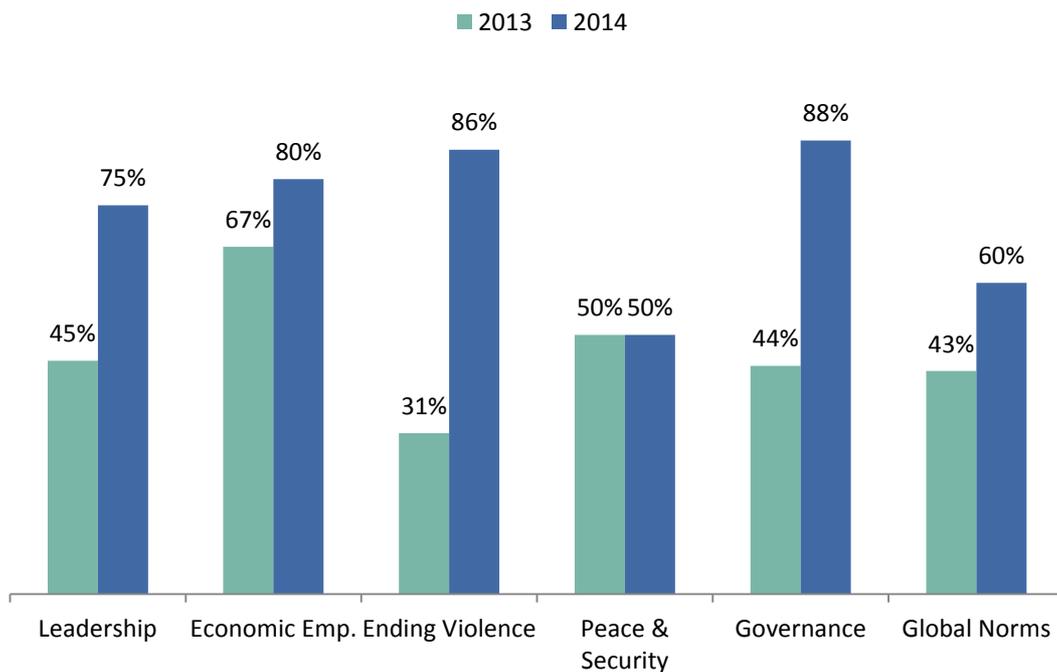
6.3.7 Evaluation Teams

On average, each evaluation was undertaken by a team of three evaluators (although there is a distorting effect created by a few very large evaluations). Based on the overall numbers, the average UN Women evaluation team would consist of 2 international evaluators and 1 national evaluator. The most frequent evaluators are female internationals (39%), and least frequent are male nationals (16%).

There is greater use of international evaluators (64%) than national evaluators (36%) – although the decentralised structure of UN Women (including multi-country offices) partly accounts for this weighting. More women evaluators (59%) are used than men (41%).

6.4 Trends by Quality Assessment Parameters

6.4.1 Overall trends Parameters 1-8

Reports rated strongest in relation to recommendations (90% good or above) and findings and report structure (both 81% good or above). The next best performing parameter was describing the object of the evaluation (76% good or above).

The weakest parameter was gender and human rights, with 53 percent of reports meeting or exceeding UN SWAP standards. The methods sections were also an area requiring attention, with 66 percent of reports rated as good or above.

Figure 7: Overall performance of parameters 1-8
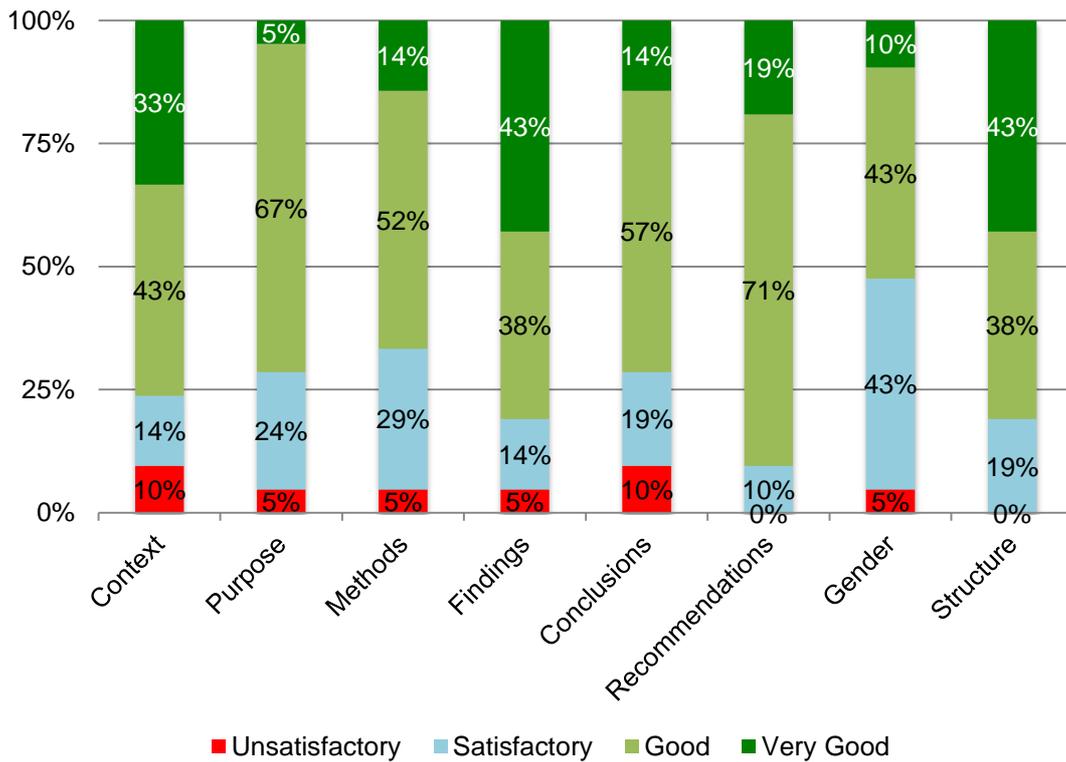
Figure 8: Year-on-year comparison in percent of reports rated good or very good
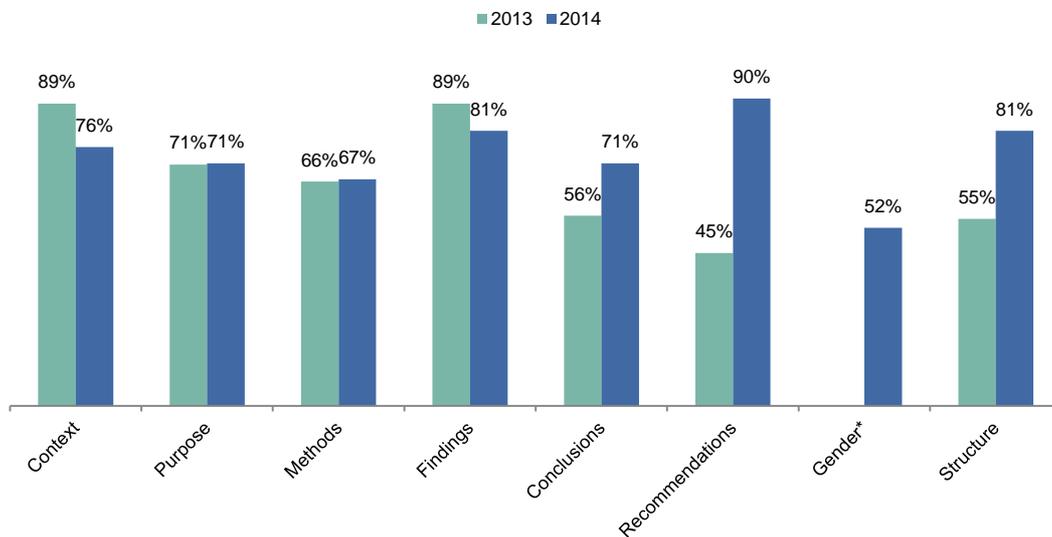
*The indicators for rating Gender changed for GERAAS 2014 and are not directly comparable to GERAAS 2013.

6.4.2 Parameter 1: Object and Context of the Evaluation

Reports rated strongly in regard to describing the implementation status of policies, programmes and projects (86% rated good or above). Whilst most reports included a clear explanation of the logic model of the evaluation object, 1 in 4 reports require substantive improvement in this regard. In the few cases where projects or programmes do not have a logical framework to present it is necessary for

evaluations to reconstruct the theories of change. A number of reports presented both.

Generally, evaluation reports were less robust in terms of presenting a full description of the context, object and stakeholders. For each of these areas, 30 percent of reports required improvement. It tended that reports would focus on one or two of these aspects of the context, at the expense of others (rather than some reports being weak in all regards).

Characteristics of good reports included: making specific reference to CEDAW and other relevant normative frameworks; presenting clear illustrations of the logic model or reconstructing the theories of change; and mapping the stakeholding groups using a systematic approach.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:
- ☐ The use of a table to identify stakeholding groups, their role in the evaluation, and the size of the sample frame;
- ☐ Expanding the range of techniques for identifying stakeholders (see: www.betterevaluation.org/resources/guides/mapping_stakeholders/guidelines and www.betterevaluation.org/plan/approach/critical_systems_heuristics)
- ☐ Analysing the validity of the theories of change used by interventions;
- ☐ Making explicit the assumptions underlying programme; and
- ☐ Widening the analysis of contextual factors from a gender perspective (see: www.uneval.org/document/detail/1616 pages 49-50).

6.4.3 Parameter 2: Purpose, Objectives and Scope
Reports were found to be particularly strong in terms of identifying the objectives (90% good or above) and evaluation criteria (81% good or above). It was noted that the best reports presented a clear definition of how each of the OECD-DAC criteria were being interpreted within the context of the evaluation.

Reports were not always so specific with regard to defining the purpose (76% good or above), or in describing human rights and gender considerations (71% good or above). There was frequently some confusion between purpose and objectives – with the rationale for evaluating a programme/project at any point in time most often being implied (rather than clearly stated).

The aspect of greatest concern, however, is evaluation scope. Only 38 percent of reports were rated as good in regard to scope. Many of the report implied the scope of the evaluation in the purpose and objectives, but few included a substantive discussion of boundaries. Unlike other aspects of evaluation reports, the meta evaluation found that there is very little practical guidance to point to in regard to defining the scope: indicating a possible reason for its weak rating.

Characteristics of good reports included: extended discussions on the scope and the boundaries of the evaluation (including justification of excluded issues); detailed evaluation frameworks; and elaborating the definition and interpretation of evaluation criteria within a specific context.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:

☐ Including a specific evaluation criterion on gender and human rights in the evaluation framework;

☐ Using diagrams to illustrate how the main elements of the evaluation 'fit together';

☐ Making specific reference to one or more of the purposes identified in Point 5 of the UN Women Evaluation Policy;

☐ Developing rubrics or standards of 'what success looks like' for evaluation indicators as part of the framework; and

☐ For large multi-country programmes to consider instigating Developmental Evaluation from the design stage, rather than (or in addition to) ex-post evaluation (see: www.betterevaluation.org/plan/approach/developmental_evaluation).

6.4.4 Parameter 3: Methodology

Methods was one of the weakest overall parameters, with 1 in 3 reports requiring substantive improvements in order to meet UNEG standards. Nevertheless, some aspects were rated strongly across most reports, with 81 percent of evaluations rated as good or above with regards to data collection and stakeholder consultation. This reflects the extensive use of participatory approaches in UN Women evaluations, and the use of annexes to present data collection tools.

The areas requiring greater focus are ethics (38% rated good or above) and the definition of the sampling frame (48% rated good or above). Although many reports include one sentence on the use of UN Women/UNEG ethics standards, few explained how these were interpreted and realised in the context of the particular evaluation. This left an impression that ethics is known to be important, but is not yet being considered deeply or meaningfully in evaluation designs. In part this may be due to participatory approaches being construed as inherently ethical, whereas no such automatic link exists.

With regard to the sampling frame, the most common sampling approach was convenience (although this was rarely specified), or a broad version of purposive (this may reflect the predominance of qualitative designs among the evaluations). Very few reports discussed the representativeness of these samples or established a clear sample frame in order to do so – an issue linked to the challenge of scoping that was previously highlighted.

Characteristics of good reports included: transparent descriptions of the methods used, making reference to external sources (for example, Bamburger and Segone, 2011, where relevant); specifying precise techniques of analysis (such as content or comparative analysis); strong participatory elements; highlighting methodological limitations and their implications for decision makers; and genuinely applying mixed methods.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:

- ☐ Clearly stating and justifying the overall design of the evaluation (with reference to an establishing framework), rather than just listing a collection of methods and tools;
- ☐ Detailing participation of marginalised/target groups at stages other than just data collection;
- ☐ Broadening the use of different approaches (see: www.bit.ly/GERAAS5, or www.betterevaluation.org/approaches);
- ☐ Expanding mixed methods to all stages of the evaluation (see: www.betterevaluation.org/blog/mixed_methods_part1); and
- ☐ Providing detailed elaboration on how UNEG ethics standards were implemented in the evaluation.

6.4.5 Parameter 4: Findings

Findings was the strongest section of evaluation reports overall. In particular, these sections rated well in terms of addressing the criteria, responding to evaluation questions, and identifying underlying reasons for the observations (all over 90% good or above). In order to strengthen the findings sections further, the meta evaluation identified a need to improve the systematic marshalling of evidence in order to support statements of findings – in particular identifying where contradictory or limited evidence exists.

One of the challenges of the predominance of qualitative designs, it seems, is that evaluations generally struggle to establish a clear method for weighting or prioritising different sources and types of evidence. As a result, many reports described the consensus view among stakeholders, without including discussion on the frequency, representativeness or validity of these perceptions. Whilst this challenge is not clearly reflected in the ratings of the findings parameter, it is mirrored in the scoring of gender analysis under parameter 7, gender (see below).

Characteristics of reports rated good include: inclusion of frequency analysis data with qualitative evidence; inclusion of a specific discussion or section on unexpected findings and their implications; including gender and human rights analysis alongside institutional analysis; reference to CEDAW concluding observations; and strong adherence to the evaluation framework.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:
- ☐ Expanding the use of lists summarising 'most important factors' used by one of the evaluations at the end of each set of findings;
- ☐ Making more disciplined references to the evidence that supports and contradicts each evaluative statement throughout the report;
- ☐ Disaggregating expected and unexpected findings;
- ☐ Critically analysing 'consensus views' among the evaluation stakeholders;
- ☐ Only reporting the results of analysis in the findings section (many reports include background information such as expenditure or coverage in the findings section rather than the context section – even though this information is a 'discovery' to the evaluators, it is not an evaluative finding per se); and
- ☐ Expanding the discussion of relevance to consider whether the evaluand is the *most relevant option* available to UN Women, rather than simply whether it has some relevance.

6.4.6 Parameter 5: Conclusions and Lessons Learned

Conclusions sections were most often found to convey the basic premise of the UNEG standards – providing an overall assessment of each of the evaluation criteria. However, many were notable for being very short (around one page), leaving little room for discussion of supporting evidence, deeper insights, or possible solutions (these indicators rating 57%, 71% and 67% good or above, respectively). This observation highlights the potential for improvement by increasing the level of time and report space committed to elaborating conclusions.

Nevertheless, conclusions were largely found to be balanced, presenting both the strengths and weaknesses of the evaluation object (86% rated good or above). In addition, 16 reports chose to include lessons learned, and 75 percent of these rated as good or above: suggesting that the evaluation function is generating generalizable insights that can be used across the organisation.

Characteristics of good reports included: a balanced and constructive analysis of the underlying factors in evaluation findings; responsiveness to the evaluation framework and questions; implications for future programming; and lessons learned that were supported with evidence and generalised to a wider context.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:
- ☐ Ensuring traceability between findings and conclusions with explicit cross-referencing or a summary table;
- ☐ For reports that are struggling to develop conclusions, structure the section according to the evaluation framework – with one conclusion for each major question or criterion;
- ☐ To ensure conclusions go beyond a synthesis of findings, and add additional insights into causes and implications of the evaluation observations (for example, operational and organisational challenges).

6.4.7 Parameter 6: Recommendations

Recommendations represent a strength in evaluation reports, with almost all reports identifying relevant, targeted and actionable recommendations (all more than 86% rated good or above). 4 out of 5 reports rated strongly in terms of marshalling evidence from the findings and conclusions to support the recommendations.

One area to improve the quality of recommendations is to enhance prioritisation. One report estimated the urgency and impact of different recommendations in order to do this, which could be a useful example to learn from. The lowest rated indicator for recommendations was the description of the process by which they were developed. Only 1 in 5 reports rated good or above in this regard. Given UN Women's commitment to participatory and empowerment approaches, future reports should more clearly describe how recommendations were developed – and who was involved in the process.

Characteristics of good reports included: the provision of practical examples to illustrate recommendations; a specific attempt to prioritise either through clustering

recommendations around major themes or with a numbering system; and made explicit references to the groups intended to implement recommended actions.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:
- ☐ Starting the recommendations section with a paragraph explaining the process by which recommendations were developed (and validated), including who was involved in the process;
- ☐ Developing recommendations through a participatory process, such as the Summit Workshop used in the Collaborative Outcomes Reporting Technique (see: www.betterevaluation.org/plan/approach/cort);
- ☐ Rating each recommendation in terms of 'urgency' and 'potential impact', or similar criteria designed to aid with prioritisation;
- ☐ Adding detail to generic recommendations relating to improving the efficiency of organisational practices (such as fund management).

6.4.8 Parameter 7: Gender and Human Rights

The assessment of gender in GERAAS 2014 represents a significant change from the indicators used in GERAAS 2013, and holds the entity to a higher standard. For example, whereas the previous UNEG/UN SWAP indicators took into account the recruitment of gender specialists in evaluation teams (an aspect where UN Women score highly), under the four new indicators evaluation teams are not considered.

Whilst the overall UN SWAP score equated to 'meeting requirements', only 52 percent of reports met this standard or higher. Overall, gender and human rights – assessed using the 4 UN SWAP indicators – scored as the weakest of the GERAAS 2014 parameters. Nearly half of reports are not yet *fully* compliant with UN SWAP. This is because UN SWAP now considers evaluation methods and analysis independently of the evaluation object (just because an evaluation is about a gender-related subject does not make it automatically gender-responsive).

Figure 9: GERAAS scores for UN SWAP indicators (max score 3)
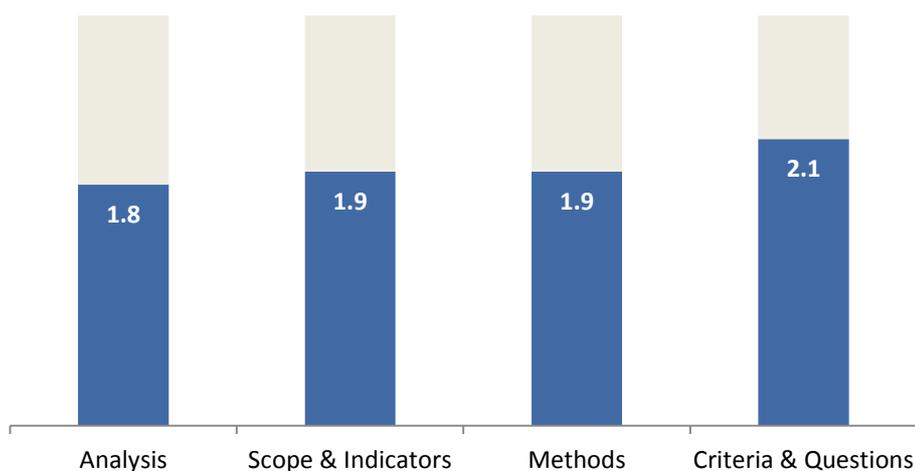
Some areas of reports, however, are fairly strong. For example, reports scored well with regard to integrating gender into evaluation criteria and questions. Whilst few reports included a standalone gender or human rights criterion, 81% met the UN SWAP standards by mainstreaming gender in the evaluation questions.

Whilst only two reports rated very good in terms of gender analysis, the most number of reports (13) rated good in this regard – meaning that 71% of reports rated good or above. The challenge for gender analysis seemed to be threefold:
1. Reports generally did not specify an approach to gender analysis in the methods section;
2. Findings sections were organised according to evaluation criteria – with the gender analysis implied by a mainstreamed approach becoming 'lost' in the process; and
3. The nature of programmes and projects – many intervening at the policy level or in terms of capacity development – and lack of disaggregated baseline data meant that few reports were in a position to include quantitative analysis (disaggregated by gender).

The scoring of the indicators for both scope/indicators, and methods highlights this final point: with 1 in 3 reports requiring substantive improvements. In the case of methods, although there was the wide application of participatory tools, 34% of reports did not explain how such methods were being implemented so as to identify and explore gender.

There was similar observation with regard to indicators – many of which focused solely on the experience of women participants or institutions, without being designed to probe relationships and dynamics between socially constructed groups. The challenge moving forward is thus guiding evaluations to apply gender responsive techniques at all stages (not just during data collection), and supporting gender analysis of work at the macro and institutional levels.

All reports with details of the evaluators included team members familiar with human rights and gender concepts. As noted in GERAAS 2013, the nature of the evaluation objects also led to extensive discussion of gender issues and marginalised groups within reports, although few reports followed this through in terms of disaggregated analysis of quantitative data or evaluation participants.

Policy and Pilot evaluations (two each) average the highest quality, followed by programme evaluations – which are distributed exactly evenly across the UN SWAP scoring spectrum. Project evaluations average slightly below satisfactory overall.

Reports were strongest in regard to gender responsive evaluation frameworks (including gender responsive indicators) and methods. The greatest improvement is available to be made in relation to gender analysis and the scoping of evaluations to consider power issues. It was also observed that few reports included gender or human rights as specific objectives, with most purpose and objectives statements focusing on institutional or organisational issues.

Evaluations submitted by HQ units and Latin America and the Caribbean scored highest in regard to UN SWAP. Both African regions and Europe and Central Asia averaged as approaching UN SWAP requirements, suggesting that there is room for greater focus on gender responsive methods and analysis. Reports from Asia Pacific and Arab States represented a fairly even spread of quality in regard to gender and human rights.

Characteristics of good reports included: consistent discussion of gender issues at all stages of the evaluation (context, objectives, scope, framework, design, process, data, analysis, conclusions and recommendations); presence of gener experts in the evaluation team; and relevant references to CEDAW and/or other normative frameworks.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:
- ☐ Exploring roles and power of different stakeholding groups using tools such as Critical Systems Heuristics (see: www.mymande.org/elearning/course-details/1 Lesson 8);
- ☐ Making greater use of quantitative gender analysis (see: http://gaap.ifpri.info/files/2010/12/GAAP_Toolkit_Feb_14.pdf);
- ☐ To make specific reference the UNEG guidance on Integrating Human Rights and Gender in Evaluation during the inception phase, including a table in the main report that explains how the guidance was applied; and
- ☐ Considering the more frequent use of an evaluation criterion dedicated to gender equality, human rights and/or social inclusion.

6.4.8 Parameter 8: The Report Structure

Reports were almost entirely found to be highly aligned to UNEG guidance in terms of the structure and provision of basic information – suggesting that all evaluation teams are now being provided with the requisite guidance (this is supported by the citing of UNEG standards in all terms of reference). By comparison, the writing of executive summaries would benefit from additional support. Although 67 percent of reports were rated good as above, and nearly all reports included an executive summary, 1 out of 3 executive summaries contained insufficient information to stand alone. Enhancing these briefs – including stating the limits of the evaluation method – is a clear opportunity to enhance usability.

Some constructive recommendations made by the reviews to enhance evaluation reports in this aspect included:
- ☐ Sharing a model UNEG-standard evaluation report with evaluation teams at the beginning of the evaluation – and example of a proforma report with space for the basic elements has been made available (see: http://bit.ly/GERAAS1); and
- ☐ Ensuring that executive summaries include a clear statement on the methodological limitations of the evaluation and there implications for decision makers in terms of the evidence presented.

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

Participation was an important feature of data collection approaches. For example, one evaluation tailored the language of the process to the capacity of the available interlocutor. [6] Several reports include details on **evaluation processes that extended participation** beyond the data collection stage. For example, holding a participatory workshop on conclusions and recommendations[7]; recommendations based on mapping out options for users to choose from rather than the preferences of the evaluators[8]; and the use of multi-country teams (including UN Women staff members) to ensure knowledge is retained inside the organisation[9].

6.5.2 Evaluation Design

Several reports included clear tables that clearly presented the **evaluation criteria and questions**[10], including how these were contextualised[11], or presenting a scoring matrix and rubric to assess each criterion.[12] One evaluation included a **specific criterion for Inclusion**[13], and another for Gender and Human Rights.

Although the methods parameter did not rate as strongly as some others, some reports were noticeable in terms of the use of **innovative or good-practice methods**. Mixed methods and theory-based designs are both identified in UN Women and EvalPartners guidance as being relevant approaches to gender responsive evaluation. A number of reports discussed and developed these approaches in detail. [14]

---

[6] *Mid-term Evaluation of the Women Entrepreneurs Project in South Africa.* South Africa. [Eastern and Southern Africa]

[7] *Formative Evaluation of the Pacific Regional Ending Violence Against Women (EVAW) Facility Fund.* Fiji, Kiribati, Nauru, Papua New Guinea, Solomon Islands, Samoa, Tonga, and Vanuatu. [Asia and the Pacific]

[8] *UN Women's Contribution to Women's Economic Empowerment 2011-2014.* HQ, South Africa, Zimbabwe, Moldova. Nepal, Bolivia, and Jordan. [Corporate (HQ)]

[9] *Mid-Term Evaluation Report Strengthening Responses to Create Wealth and Reduce Poverty for Women in Cross Border Trade (WICBT) in Africa.* Rwanda, Burundi, DRC, Swaziland, Lesotho, Zimbabwe, Mozambique, Botswana, Liberia, Mali, Cameroon, and Nigeria. [Corporate (HQ)]

[10] *End of Project Evaluation Empowering Youth to Advocate for Women's Human Rights through Volunteerism.* Jordan. [Arab States]

[11] *Ending Violence Against Women 2010-2013 Joint UN Women-UNFPA Programme Ethiopia End of Programme Evaluation.* Ethiopia. [Eastern and Southern Africa]

[12] *Mid-term Evaluation of the Women Entrepreneurs Project in South Africa.* South Africa. [Eastern and Southern Africa]

[13] *Formative Evaluation of the Pacific Regional Ending Violence Against Women (EVAW) Facility Fund.* Fiji, Kiribati, Nauru, Papua New Guinea, Solomon Islands, Samoa, Tonga, and Vanuatu. [Asia and the Pacific]

[14] *Final Evaluation of the UN Women Sudan Gender Equality and Women's Empowerment (GEWE) Programme: "Consolidating Efforts that Contribute to Gender Equality and Women's Empowerment in Sudan".* Sudan. [Eastern and Southern Africa]; *Final Project Evaluation "The National Implementation Support Plan In Policies For Women".* Brazil. [Latin Americas and Caribbean]; *'Empowering Women in Rural and Informal Settings through Capacity Development' Final Evaluation Report.* India. [Asia and the Pacific]

Participatory approaches were frequently cited in principle, but a number of reports added substantive detail on specific methods and tools. These included Most Significant Change and participatory videos [15]. The Corporate Thematic Evaluation of Women's Economic Empowerment extended the range of methods and tools, using **systems approaches** such as fuzzy sets Qualitative Comparative Analysis, forcefield analysis, organizational capacity assessment, social network mapping, and Critical Systems Heuristics.[16]

From a technical perspective, several reports reflected strong evaluation practice. For evaluands that did not have an established project or programme logic, a number of evaluation reports included a **reconstruction of the theory of change**. [17] One evaluation contained an intriguing discussion on programme risks[18]; another was very rich in quantitative data analysis [19]; and a third identified and reported on negative effects. [20] In one of the few quasi-experimental evaluations, the same evaluation team carried out baseline, mid-term and endline evaluations to ensure consistency.[21]

6.5.3 Evaluation Analysis and Reporting

Some evaluation reports provided **accessible summaries on the achievement of results**. One report included a table in the annexes explaining the achievement of objectives with traffic lights[22], another included a summary table in the findings section.[23] At least three reports stood out in terms of **identifying major contributing factors**, challenges and potential solutions. [24] In one discussion of strengths and

---

[15] *Le Projet d'Appui a la Legislation Sensible au Genre et Promotion du Leadership Feminin a l'Est de la RD Congo; le Projet de Promotion et Relevement Socio-economique des Femmes Vulnerables en Milieu Rural du Sud-Kivu; le Projet d'Appui au Relevement Economique des Femmes et Filles Victimes des Conflits Armes dans la Province du Nord Kivu.* DRC. [Western and Central Africa]; *Evaluación temática regional sobre participación política de las mujeres en América Latina y el Caribe 2011-2014.* [Latin Americas and Caribbean]; *UN Women's Contribution to Women's Economic Empowerment 2011-2014.* HQ, South Africa, Zimbabwe, Moldova. Nepal, Bolivia, and Jordan. [Corporate (HQ)].
[16] *UN Women's Contribution to Women's Economic Empowerment 2011-2014.* HQ, South Africa, Zimbabwe, Moldova. Nepal, Bolivia, and Jordan. [Corporate (HQ)]
[17] *Ending Violence Against Women 2010-2013 Joint UN Women-UNFPA Programme Ethiopia End of Programme Evaluation.* Ethiopia. [Eastern and Southern Africa]; *Final Evaluation of the UN Women Sudan Gender Equality and Women's Empowerment (GEWE) Programme: "Consolidating Efforts that Contribute to Gender Equality and Women's Empowerment in Sudan".* Sudan. [Eastern and Southern Africa]
[18] *Mid-Term Review of the CEDAW South East Asia Programme II.* Cambodia, Indonesia, Lao PDR, Viet Nam, Philippines, Thailand, Timor-Leste, and Myanmar. [Asia and the Pacific]
[19] *UN Women's Anti Human Trafficking Program.* India. [Asia and the Pacific]
[20] *Evaluation Thématique de l'Action du Bureau Multi-pays d'ONU Femmes Maghreb dans le domaine de la Lutte Contre les Violences à l'égard des Femmes et des Filles.* Algeria, Morocco and Tunisia. [Arab States]
[21] *UN Women's Anti Human Trafficking Program.* India. [Asia and the Pacific]
[22] *Final Evaluation of the Government of Kenya (GoK)/United Nations (UN) Joint Programme on Gender Equality and Women's Empowerment (JP GEEW).* Kenya. [Eastern and Southern Africa]
[23] *Contribution de l'ONU Femmes dans l'Atteinte des Resultats de Developpement de la Côte d'Ivoire 1er Janvier 2012- 31 Decembre 2013.* Côte d'Ivoire. [Western and Central Africa]
[24] *Final Evaluation of the Government of Kenya (GoK)/United Nations (UN) Joint Programme on Gender Equality and Women's Empowerment (JP GEEW).* Kenya. [Eastern and Southern Africa]; *Contribution de l'ONU Femmes dans l'Atteinte des Resultats de Developpement de la Côte d'Ivoire 1er Janvier 2012- 31 Decembre 2013.* Côte d'Ivoire. [Western and Central Africa]; *Mid-Term Review of the CEDAW South East Asia Programme II.* Cambodia, Indonesia, Lao PDR, Viet Nam, Philippines, Thailand, Timor-Leste, and Myanmar. [Asia and the Pacific]

weaknesses, "change resistance" was identified as an issue for achieving intended behaviour change effects.[25]

Although the UNEG standards do not specifically call for reports to use **visualisation** or other approaches to utilization, a few reports did attempt to use enhanced techniques. One report included maps illustrating zones of intervention and sites visited [26], another included a figure illustrating pathways of data through the evaluation design, and infographics at the beginning of each section summarizing key figures.[27]

6.5.4 Integration of Gender Equality

A number of reports included the voice of stakeholders through the use of selected quotes and stories of change. These helped to **illustrate findings with concrete, real life, example**s that helped to explain the analysis in human terms[28]. Other reports listed specific achievements to illustrate the findings [29], or included descriptions of example case studies.[30]

Gender analysis is a key indicator under UN SWAP. Some reports managed to take this further than others, including: gender-based breakdowns of data[31]; discussion of relevance based on women's needs and situation in addition institutional frameworks[32]; and mapping funded projects to different normative frameworks to assess their relevance.[33]

---

[25] *Evaluation Thématique de l'Action du Bureau Multi-pays d'ONU Femmes Maghreb dans le domaine de la Lutte Contre les Violences à l'égard des Femmes et des Filles.* Algeria, Morocco and Tunisia. [Arab States]

[26] *Contribution de l'ONU Femmes dans l'Atteinte des Resultats de Developpement de la Côte d'Ivoire 1er Janvier 2012- 31 Decembre 2013.* Côte d'Ivoire. [Western and Central Africa]

[27] *UN Women's Contribution to Women's Economic Empowerment 2011-2014.* HQ, South Africa, Zimbabwe, Moldova. Nepal, Bolivia, and Jordan. [Corporate (HQ)]

[28] See *Final Evaluation of the Government of Kenya (GoK)/United Nations (UN) Joint Programme on Gender Equality and Women's Empowerment (JP GEEW).* Kenya. [Eastern and Southern Africa]; *Le Projet d'Appui a la Legislation Sensible au Genre et Promotion du Leadership Feminin a l'Est de la RD Congo; le Projet de Promotion et Relevement Socio-economique des Femmes Vulnerables en Milieu Rural du Sud-Kivu; le Projet d'Appui au Relevement Economique des Femmes et Filles Victimes des Conflits Armes dans la Province du Nord Kivu.* DRC. [Western and Central Africa]; *Contribution de l'ONU Femmes dans l'Atteinte des Resultats de Developpement de la Côte d'Ivoire 1er Janvier 2012- 31 Decembre 2013.* Côte d'Ivoire. [Western and Central Africa]; *End-of-Programme Evaluation of EC-UN Women's "Supporting Gender Equality in the Context of HIV/AIDS".* Cambodia, Jamaica, Papua New Guinea, Kenya, Rwanda. [Corporate (HQ)].

[29] *Report of Findings of the Final Evaluation ASEAN Regional Mechanisms Project to Protect the Human Rights of Women and Girls in Southeast Asia 2010 – 2014.* Cambodia, Laos, Indonesia, Philippines, Thailand, and Vietnam. [Asia and the Pacific]; *Evaluación temática regional sobre participación política de las mujeres en América Latina y el Caribe 2011-2014.* [Latin Americas and Caribbean]

[30] *UN Women's Anti Human Trafficking Program.* India. [Asia and the Pacific]

[31] *End of Project Evaluation Empowering Youth to Advocate for Women's Human Rights through Volunteerism.* Jordan. [Arab States]

[32] *Contribution de l'ONU Femmes dans l'Atteinte des Resultats de Developpement de la Côte d'Ivoire 1er Janvier 2012- 31 Decembre 2013.* Côte d'Ivoire. [Western and Central Africa]

[33] *End of Programme Evaluation of the Project 'Action to Promote the Legal Empowerment of Women in the Context of HIV and AIDS' 03/2010 – 11/2013.* Kenya, Uganda, Tanzania, Cameroon, Ghana, Malawi, Nigeria, Rwanda, and Zimbabwe. [Corporate (HQ)]

7. Conclusions

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

Conclusion 1: There is an overall improvement since 2013, throughout the decentralised evaluation function, with fewer but more consistent evaluation reports.

The data reveals an overall improvement in the quality of evaluation reports in nearly all regards: a higher proportion of reports rated very good; a higher proportion of reports meeting UNEG standards; a lower proportion of reports rated unsatisfactory; improved ratings across nearly all strategic priority areas; and more complete evaluation reports. The meta-evaluation unknowingly rated a report that was included in GERAAS 2013 (this was later removed from the analysis), which was given the same rating as the previous year – supporting the view that overall improvement in quality is real.

There are a number of factors that may have contributed to these results:
- ☐ Fewer evaluations were submitted to GERAAS 2014, suggesting that more time and resources were available to focus on each report;
- ☐ The Independent Evaluation Office has been engaged in completing the roll-out of the regional architecture for the evaluation function (Regional Evaluation Specialists) – making greater evaluative expertise available at the decentralised levels;
- ☐ There is a strong focus on programme evaluations, which are associated with measuring outcomes (and likely to be better resourced than project/output level evaluations); and
- ☐ UN Women participated in fewer joint evaluations, having direct management control over more reports[34].

Reports are increasing structured according to UNEG guidance, with strong findings and recommendations sections that rate well because they align strongly to the evaluation criteria/framework. Conclusions sections are also aligned to evaluation criteria; but are most often more syntheses of the findings, rather than taking the analysis deeper. There is a good case, therefore, for making available guidance on how to write (and assure) stronger conclusions.

Other areas that emerged as requiring focus include: placing more material about evaluation objects (logic models, theories of change, implementation status, spend, coverage) in the background section and leaving more space for analysis in the findings; supporting evaluation within the context of peace and security; and

---

[34] One explanation for this low number of joint evaluations in 2014 may be that many development programmes have been aligned to the Millennium Development Goals, which end in 2015. We would, therefore, expect to see a large number of joint evaluations of programmes completing in 2015.

strengthening methods and gender responsiveness within the context of UN Women's macro, policy, normative and coordination work (see Conclusion 3).

Conclusion 2: The homogenuity of evaluation designs largely remains, with a few examples emerging of mixed methods approaches. A case exists for standardising different types of evaluation.

GERAAS 2013 proposed that evaluation designs (and associated limitations) were largely similar across the portfolio (qualitative designs that included limited quantitative data). Preponderance for this type of evaluation still exists, although a number of serious attempts to execute mixed methods designs were included in GERAAS 2014.

Given the emergence of demand in the Entity for mixed methods designs, there is a case for making stronger mixed methods at all stages of evaluations more accessible to decentralised evaluations (for example, through the forthcoming Country Portfolio Evaluation guidance).

Greater support for high quality mixed methods designs is, however, unlikely to satisfactorily address the issue of 'impact' being included as a criteria in terms of reference, even when the resources/time/design is insufficient to measure long terms changes in people's lives or realisation of their rights. This suggests that there is a case for clearly separating impact evaluations from outcome-level evaluations (including effectiveness, efficiency, sustainability and relevance, but excluding impact) within the UN Women portfolio: and encouraging evaluation managers to focus each ToR on either one or the other type.

Establishing clarity between what is intended to be an impact-level evaluation or an outcome-level evaluation would – potentially – allow for greater standardisation of guidance. For example, model ethics statements and implementation arrangements could be provided for these different types of evaluations: beginning to address one of the weaknesses identified in both GERAAS 2013 and GERAAS 2014. A similar approach might be taken to addressing the inconsistency observed in sampling and sample frames.

Alongside ex-post mixed methods outcome and impact evaluations, the meta-evaluation identified a third type of evaluation with potential application to UN Women: Developmental Evaluation. This was specifically suggested in the feedback for one of the global programmes: assuming that evaluation can begin to be integrated at the design stage of future programmes (rather than waiting until the mid-term or endline). Given the emphasis on learning as a purpose in many of these evaluations – and the challenges associated with difference-in-difference designs at the (multi-)regional scale – developmental evaluation is both an appropriate method for UN Women and could help diversify the evaluation portfolio.

Conclusion 3: The highest priority for future quality assurance is ensuring that gender responsive designs, analysis and reporting techniques are accessible and practiced throughout the decentralised evaluation function.

The changes made to GERAAS 2014 to align the system of scoring gender with UNEG guidelines for UN SWAP has significant implications for the rating of this parameter. The criteria for being rated highly are both narrower and more difficult to achieve. The use of gender-specialist evaluation teams and the assessment of gender-related evaluation objects is no longer recognised in the scoring used by GERAAS. In order to improve the rating of reports, it is necessary to further strengthen gender integration into the scoping, specification of methods, indicators, analysis and evaluation process.

Increasingly, resources and guidance are being made available to support this, including the UN Women Evaluation Handbook, the UNEG Guidance on Integrating Human Rights and Gender in Evaluation, the EvalPartners eLearning course on equity evaluation, and BetterEvaluation.org.

Despite these general resources, the process of developing constructive feedback for evaluation managers revealed that there are a number of issues where more standardised guidance and tools from IEO could help to fill a gap:
1. There is relatively little practical guidance on how to scope an evaluation, particularly on how to map boundaries in a way that is gender responsive (outside of material on systems evaluation);
2. The application and presentation of stakeholder mapping is inconsistent – suggesting that there is a real case for providing all evaluation managers and teams a 'default' stakeholder mapping tool. For example, this could be a table with spaces for required information, such as socio-economic groups, gender roles, human rights roles, stake in the programme, main contributions, primary intended uses for the evaluation, and utilisation needs. To extend the gender responsiveness of evaluations, the process of drafting this table could be included in the preparation of the Terms of Reference; and
3. Given the importance of UN Women demonstrating leadership in gender, a standalone criterion for Gender and Human Rights could be added as standard to all future evaluations (alongside mainstreaming gender in the other OECD-DAC criteria). Indeed, if this proves to be a useful addition, then it would provide a test-case for including a standalone criterion in future versions of UNEG and OECD-DAC standards.

Conclusion 4: Participatory approaches are a common feature of UN Women evaluations at the data collection stage. Considerable scope exists for enhancing participation at all other stages; and for emphasising empowerment alongside utilisation.

All of the evaluations cited participatory approaches in the methods section. The quality of participation varied widely: for some evaluations, participation constituted group interviews with project participants; whilst for others participation included

involvement of rights holders in data collection and interpretation (e.g. through participatory video).

In general, however, participation was primarily concerned with the data collection stage: with affected groups variously consulted as part of the social enquiry process. For a few utilisation-focused evaluations, the reports included evidence of primary intended users being included in the validation of conclusions and/or recommendations. In all cases, these users were duty bearers with whom UN Women works directly, rather than rights holders.

The evidence of evaluations including participation at the design stage was far more limited. Although some evaluations, such as the Corporate Thematic Evaluation of WEE, included extensive consultations with intended users in the development of both the terms of reference and the inception report, no evaluation included rights holders at this stage.

Similarly, rights holders were largely excluded from involvement in data collection and analysis (with the exception of participatory video in two evaluations); although a number of mid term evaluations did include duty bearers and UN Women staff as participants in this part of the process. Finally, very little information is available in reports in regard to feeding back the outcomes of evaluations to rights holders.

It should be recalled that the majority of UN Women evaluations are undertaken under significant time and financial constraints; and participatory processes tend to require more of both of these resources (the evaluations that were strong in terms of participation were also larger in size). It is unrealistic to conclude, therefore, that participation can be simply or easily increased at all stages of future UN Women evaluations.

Nevertheless, the current findings suggest that there is a strong case for working to enhance participation in UN Women evaluations: both in terms of participation at different stages; and balancing participation from a utilisation perspective (duty bearers) with participation from an empowerment perspective (rights holders). Adaptation of GERAAS 2015 to provide more specific information on the levels of participation achieved in different evaluations is one obvious starting point for drawing attention to this issue.

8. Recommendations

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

Recommendation 1: For aspects of evaluations that can be standardised, IEO should consider developing simple checklists of minimum or default standards for all UN Women managed evaluations.

Urgency	Impact	Ease	Priority
★★★☆☆	★★★★☆	★★★☆☆	★★★★☆

Checklists – sometimes referred to as 'cheatsheets' – are frequently used in high-stress situations to ensure that professionals (such as pilots, medical staff or disaster responders) do not miss important details. A number of writers have attempted to produce generic checklists for evaluators, for example Patton (2002 and 2003)[35].

Given the relatively unusual context of UN Women (in terms of the triple mandate and work at the macro-level), combined with the presence of inconsistencies in some core aspects of evaluation reports, it is recommended that checklists are piloted for issues that can most easily be standardised across evaluation reports.

Given the ratings in GERAAS 2014, potential priority areas for checklists are:
1. Scoping the evaluation and describing the scope in the report;
2. Developing evaluation indicators that include minimum standards, benchmarks, or other rubrics of success;
3. Mapping stakeholders, establishing the sample framework, and reporting on sampling;
4. Implementing the UNEG guidance on ethics and reporting on this;
5. Exploring unexpected findings and important contributing factors; and
6. Marshalling, referencing and cross-referencing evidence in report writing.

Recommendation 2: Continue existing efforts to strengthen Terms of Reference and recruit evaluators with expertise in gender responsive evaluation.

Urgency	Impact	Ease	Priority
★★★★★	★★★★☆	★★★★★	★★★★★

GERAAS 2013 identified Terms of Reference and gender-sensitive evaluation teams as key contributing factors to report quality. Three findings support the continuation of efforts to strengthen these aspects:

---

[35] See: http://www.wmich.edu/evalctr/archive_checklists/ufe.pdf, and
http://www.wmich.edu/evalctr/archive_checklists/qec.pdf.

1. Overall ratings for GERAAS 2014 have shown improvement compared with GERAAS 2013;
2. With the exception of unrealistically requiring assessments of impact, Terms of Reference were noted as contributing factors to good report quality in nearly all cases; and
3. Gender responsive evaluation needs to be a major focus on efforts to further strengthen the evaluation function.

Specific actions that were identified by the meta-evaluation as contributing to further enhancing Terms of Reference are:

1. Strengthen the focus on the design of evaluations (difference-based, agreement-based, mixed-methods, participatory, appreciative) in addition to listing the suggested methods – so as to better ensure a coherent overall approach;
2. For the time being, separate evaluations of impact from evaluations of the other OECD-DAC criteria in order to improve quality in both regards;
3. Include a clear requirement for evaluations to explicitly reference CEDAW concluding observations and other relevant normative frameworks in their analysis;
4. Extend the examination of relevance to ask the question of whether the evaluand was/is the *most relevant option* for UN Women;
5. Include a preliminary stakeholder map and gender/human rights role analysis; and
6. Specify a standalone criteria on *gender and human rights* in addition to mainstreaming gender across the evaluation framework.

Recommendation 3: Provide practical guidance and examples of gender analysis of policy, macro, coordination and normative work.

Urgency	Impact	Ease	Priority
★★★★★	★★★★☆	★★☆☆☆	★★★★☆

The meta-evaluation has identified a number of challenges in implementing the UN SWAP standards within the context of UN Women. For example, a lot of guidance material on gender analysis relates to traditional development programming. This can be expected to have generated large disaggregated datasets on populations. In nearly all of the evaluations included in GERAAS 2014, this type of data was not available to the evaluation team. Given the 'upstream' level at which UN Women interventions work, there are also too many other contributing factors to reliably use third-party data sets in place of UN Women's own baselines.

Furthermore, the predominant design of evaluations (largely qualitative, with non-statistical sampling of key informants) is not well suited to producing sufficient quantities of clean data with which to undertake quantitative gender analysis. An alternative approach would be to use techniques inspired by systems theory to analyse relationships and dynamics between different gender groups using qualitative evidence. However, as discussed in the conclusions, few reports included

a stakeholder map that was sufficiently disaggregated to use as a basis for such an analysis.

It is recommended, therefore, that any forthcoming guidance – such as the Country Portfolio Evaluation guidance – should include very practical tools (with examples) that make gender analysis accessible within the decentralised contexts of UN Women.

Recommendation 4: Update GERAAS 2015 to highlight opportunities for participation at all stages of evaluation.

Urgency	Impact	Ease	Overall
★★★★☆	★★☆☆☆	★★★☆☆	★★★☆☆

The current design of GERAAS is capable of providing two sets of insights regarding participation in UN Women evaluation:
1. A qualitative rating (with description) linked to relevant UNEG standards as part of the meta-evaluation report; and
2. Any lessons learned from participation as part of the meta-analysis report.

Based on the qualitative observation that participation in UN Women evaluations is primarily concentrated around the data collection stage (and thus requires strengthening), it is proposed that the two sets of insights described above are an insufficient basis for future capacity development.

It is thus recommended that an additional analysis of participation be included within GERAAS 2015, within the given resource and evidence limits of the system. The proposed approach is an adaptation of the Evaluation Dashboard developed by EvalQual[36]. Specifically, this would involve the inclusion of a 'participation scan' to map the levels of participation of different stakeholders at different stages of each evaluation. Combined with a similar 'methods scan', this would allow the meta-evaluation to give a richer picture of participation within the organisation, and provide a stronger basis for recommendations to improve this. Both scans could be achieved through a very simple survey of evaluation managers, triangulated with the document review.

Figure 10: Example participation scan from the EvalQual dashboard (requires adaption for GERAAS 2015)

---
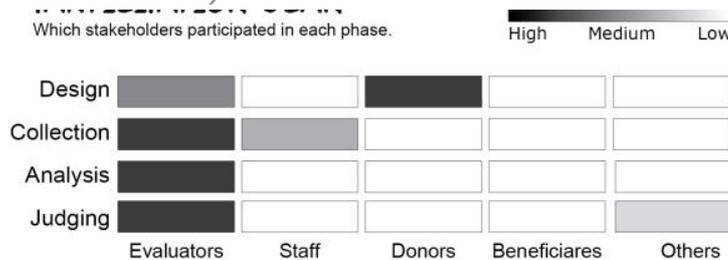
[36] EvalQual is Sara Vaca – a member of ImpactReady team involved in reviews for GERAAS 2014

9. Annexes

9.1 Terms of Reference

Background

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

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

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

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

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

The IEO is seeking to establish a long term agreement with a well-established firm to conduct a meta-evaluation and meta-analysis of final evaluation reports[37] on a yearly basis. The selected firm will review final evaluation reports (on average 30 final reports[38]), rate them

---

[37] Majority of evaluation reports are in English language but some are also available in Spanish, French and Portuguese.
[38] The number of reports varies from one year to another. The cost required will be adjusted accordingly.

against UNEG/UNWOMEN standards, write an executive feedback to be sent to the CO concerned, make analysis of trends, key weaknesses and strengths of UN-Women managed evaluation reports and produce a meta-analysis report by synthesizing the recurrent findings, recommendations, conclusions, and lessons learned of all evaluation reports completed in a given year.

Purpose & Methodology
In general, the GERAAS has four main objectives:

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

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

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

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

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

UN SWAP

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

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

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

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

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

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

Expected Deliverables
The main expected deliverables of the exercise will be
Meta-evaluation report - summarizing key trends (by region, type, scope, results, thematic areas, stage, management etc), weaknesses and strengths, as well as lessons learned and good practices emerging from the review of the evaluation reports.

Assessment of the quality of individual reports and scoring using the tool to be provided
Executive Feedback to commissioning offices using the tool to be provided
Scoring against the UN SWAP defined scoring criteria using the tool to be provided
Meta-Analysis report– analysis and synthesizes the findings, conclusions, recommendations and lessons learned presented in the evaluation reports rated satisfactory and above

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

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

Required Qualifications
Excellent and proved knowledge of evaluation methodologies and approaches
Proven experience with meta-evaluation and meta-analysis of evaluation reports, preferably with UN agencies
Experience and background in gender equality/gender analysis and gender responsible evaluations
Proven practical professional experience in designing and conducting major evaluations
Excellent analytical and writing skills in English required. Working language of Spanish, French or Portuguese as asset
Familiarity with UNEG evaluation standards is an asset
Knowledge and expertise of other or similar quality assurance systems will also be an asset
 Duration of contract
The is a yearly contract renewable up to four years upon satisfactory performance.

9.2 List and Ratings of Evaluation Reports Reviewed

Country(ies)	Title	Rated
Arab States		
Jordan	End of Project Evaluation Empowering Youth to Advocate for Women's Human Rights through Volunteerism	Good
Algeria, Morocco and Tunisia	Evaluation Thématique de l'Action du Bureau Multi-pays d'ONU Femmes Maghreb dans le domaine de la Lutte Contre les Violences à l'égard des Femmes et des Filles	Satisfactory
Asia and the Pacific		
Fiji, Kiribati, Nauru, Papua New Guinea, Solomon Islands, Samoa, Tonga, Vanuatu	Formative Evaluation of the Pacific Regional Ending Violence Against Women (EVAW) Facility Fund	Very Good
India	'Empowering Women in Rural and Informal Settings through Capacity Development' Final Evaluation Report	Very Good
India	UN Women's Anti Human Trafficking Program	Very Good
Cambodia, Indonesia, Lao PDR, Viet Nam, Philippines, Thailand, Timor-Leste, Myanmar	Mid-Term Review of the CEDAW South East Asia Programme II	Satisfactory
Cambodia, Laos, Indonesia, Philippines, Thailand, and Vietnam	Report of Findings of the Final Evaluation ASEAN Regional Mechanisms Project to Protect the Human Rights of Women and Girls in Southeast Asia 2010 – 2014	Satisfactory
Corporate (HQ)		
South Africa, Nepal, Zimbabwe, Bolivia, Moldova, Jordan	Corporate Thematic Evaluation of Un Women's Contribution to Women's Economic Empowerment 2011-2014	Very Good
Cambodia, Jamaica, Papua New Guinea, Kenya, Rwanda	End-of-Programme Evaluation of EC-UN Women's "Supporting Gender Equality in the Context of HIV/AIDS"	Very Good
Kenya, Uganda, Tanzania, Cameroon, Ghana, Malawi, Nigeria, Rwanda, and Zimbabwe.	End of Programme Evaluation of the Project 'Action To Promote The Legal Empowerment of Women in the Context of HIV and AIDS' 03/2010 – 11/2013	Very Good
Rwanda, Burundi, DRC, Swaziland, Lesotho, Zimbabwe, Mozambique, Botswana, Liberia, Mali, Cameroon, Nigeria.	Mid-Term Evaluation Report Strengthening Responses to Create Wealth and Reduce Poverty for Women in Cross Border Trade (WICBT) in Africa	Good
Eastern and Southern Africa		
Sudan	Final Evaluation of the UN Women Sudan Gender Equality and Women's Empowerment (GEWE) Programme: "Consolidating Efforts that Contribute to Gender Equality and Women's Empowerment in Sudan"	Very Good
Ethiopia	Ending Violence Against Women 2010-2013 Joint UN Women-UNFPA Programme Ethiopia End of Programme Evaluation	Good
Kenya	Final Evaluation of the Government of Kenya (GoK)/United Nations (UN) Joint Programme on Gender Equality and Women's Empowerment (JP GEEW)	Good
Namibia	Final Evaluation of the Pro Poor Governance, Gender Equality and Women's Empowerment from a HR perspective in Namibia	Satisfactory

South Africa	Mid-Term Evaluation of the Women Entrepreneurs Project in South Africa	Satisfactory
DRC	Le Projet d'Appui a la Legislation Sensible au Genre et Promotion du Leadership Feminin a l'Est de la Rd Congo; Le Projet de Promotion et Relevement Socio-Economique des Femmes Vulnerables en Milieu Rural du Sud-Kivu; Le Projet a'Appui au Relevement Economique des Femmes et Filles Victimes des Conflits Armes dans la Province du Nord Kivu	Good
Europe and Central Asia		
Georgia	Mid-Term Review of the UN Women's Project Women for Equality, Peace and Development (WEPD) II	Satisfactory
Latin Americas and Caribbean		
Brazil	Final Project Evaluation "The National Implementation Support Plan In Policies For Women "	Very Good
Latin America	Evaluación temática regional sobre participación política de las mujeres en América Latina y el Caribe 2011-2014	Very Good
Western and Central Africa		
Ivory Coast	Contribution de l'ONU Femmes dans l'Atteinte des Resultats de Developpement de la Côte d'Ivoire 1er Janvier 2012- 31 Decembre 2013	Good

9.3 Report Review Format

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

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

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

VII

GUIDING POINTS/EXPLANATORY NOTE	RATING	Remarks	Guiding Question	Constructive feedback for future reports
3.1 **Methodology:** The report presents transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve evaluation purposes and objectives.				
3.2 **Data Collection:** The report describes the data collection methods and analysis, the rationale for selecting them, and their limitations. Reference indicators and benchmarks are included where relevant.				
3.3 **Data Sources:** The report describes the data sources, the rationale for their selection, and their limitations. The report includes discussion of how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limits.				
3.4 **Sampling Frame:** The report describes the sampling frame – area and population to be represented, rationale for selection, mechanics of selection, numbers selected out of potential subjects, and limitations of the sample.				
3.5 **Stakeholders Consultation:** The evaluation report gives a complete description of stakeholder's consultation process in the evaluation, including the rationale for selecting the particular level and activities for consultation.				
3.6 **Data Quality:** The report presents evidence that adequate measures were taken to ensure data quality, including evidence supporting the reliability and validity of data collection tools (e.g. interview protocols, observation tools, etc.)				
3.6 **Gender and Human Rights considerations:** The methods employed are appropriate for analyzing gender and rights issues identified in the evaluation scope.				
3.7 **Ethics:** The evaluation report includes a discussion of the extent to which the evaluation design included ethical safeguards and mechanisms and measures that were implemented to ensure that the evaluation process conformed with relevant ethical standards including but not limited to informed consent of participants, privacy and confidentiality considerations.				

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

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

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

	RATING	Remarks		
developed with the involvement of relevant stakeholders.				
6.2 The report describes the process followed in developing the recommendations including consultation with stakeholders.				
6.3 Recommendations are relevant to the object and purposes of the evaluation.				
6.4 Recommendations clearly identify the target group for each recommendation.				
6.5 Recommendations are clearly stated with priorities for action made clear.				
6.6 Recommendations are actionable and reflect an understanding of the commissioning organization and potential constraints to follow up.				

PARAMETER 7: GENDER AND HUMAN RIGHTS			Guiding Question	Constructive feedback for future reports
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 GEEW is integrated in the evaluation **scope** of analysis and **indicators** are designed in a way that ensures GEEW-related data will be collected				
7.2 Evaluation **criteria and evaluation questions** specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved.				
7.3 A gender-responsive evaluation **methodology,** methods and tools, and data analysis techniques are selected.				
7.4 The evaluation findings, conclusions and recommendation reflect a **gender analysis**				

PARAMETER 8: THE REPORT STRUCTURE			Guiding Question	Constructive feedback for future reports Including how to address weaknesses and maintaining good practice
GUIDING POINTS/EXPLANATORY NOTE	RATING	Remarks	**Is the report well structured, logical, clear and complete?**	
THE REPORT STRUCTURE				
8.1 **Report is logically structured** with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations).				
8.2 **The title page and opening pages** provide key basic information A. Name of the evaluation object B. Timeframe of the evaluation and date of the report C. Locations (country, region, etc.) of the evaluation object D. Names and/or organizations of evaluators E. Name of the organization commissioning the evaluation 6. Table of contents which also lists Tables, Graphs, Figures and Annexes G. List of acronyms.				
8.3 The Executive Summary is a stand-alone section that includes A. Overview of the evaluation object B. Evaluation objectives and intended audience C. Evaluation methodology D. Most important findings and conclusions E. Main recommendations				
8.4 **Annexes** increase the credibility of the evaluation report. They may include, inter alia: A. TORs B. List of persons interviewed and sites visited. C. List of documents consulted D. More details on the methodology, such as data collection instruments, including details of their reliability and validity E. Evaluators biodata and/or justification of team composition F. Evaluation matrix G. results framework				

Additional Information	
Assess the extent to which the evaluation successfully addresses the ***Terms of Reference:*** If the report does not include a ToR then a recommendation should be given to ensure that all evaluations include the ToR in the future. Some evaluations may be flawed because the TORs are inappropriate, too little time etc. Or, they may succeed despite inadequate TORs. This should be highlighted.	
Identify aspects of ***good practice*** of the evaluation	

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