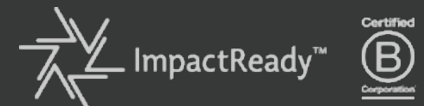


*Global Evaluation Report  
Assessment and Analysis System*

# UN WOMEN META EVALUATION 2015

*Joseph Barnes  
Messay Teferi (Manager)*

This document presents the major findings of the 2015 meta evaluation of UN Women evaluation reports submitted to the GERAAS process. It is accompanied by an online version of the report at <http://bit.ly/GERAAS15>



3

Background

5

Findings

16

Conclusions

23

Recommendations

# CONTENTS

# BACKGROUND

---

# PURPOSE

The Purpose of this meta-analysis is to capture the quality of evaluation reports according to UN Evaluation Group (UNEG) standards.

## **The Global Evaluation Report Assessment and Analysis System (GERAAS) has four main objectives:**

**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;

**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;

**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

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

*An explanation of the full GERAAS method is available to download*

[http://bit.ly/GERAAS\\_Method](http://bit.ly/GERAAS_Method)

This report assesses final evaluation reports uploaded in the UN Women Global Accountability and Tracking of Evaluation System (GATE) by January 2016.

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

## Part 2

---

Overall

---

2015 Evaluations

---

HQ

---

Arab States

---

Asia Pacific

---

Europe and Central  
Asia

---

Eastern and Southern  
Africa

---

Latin America and the  
Caribbean

---

West and Central  
Africa

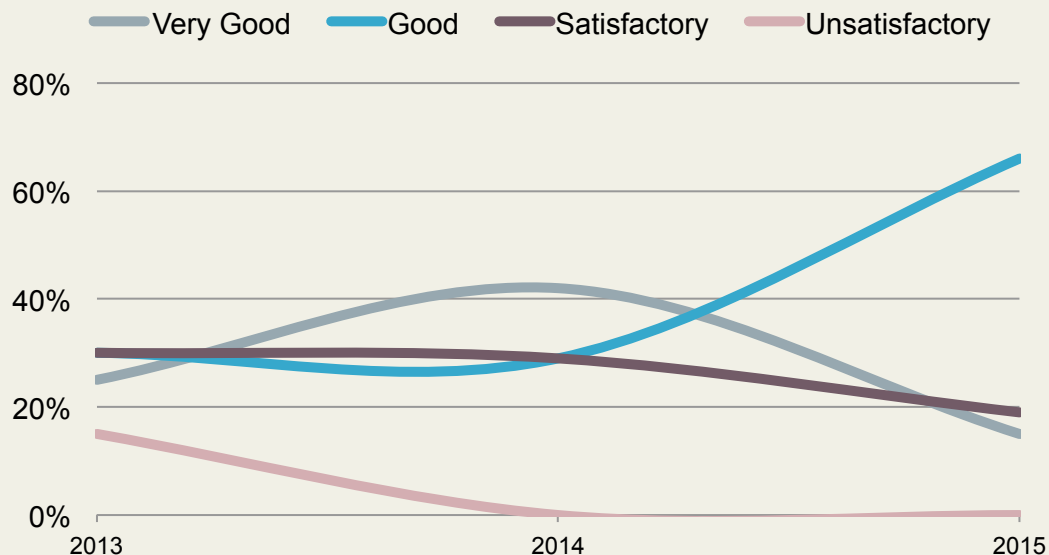
---

Strengths and  
Weaknesses

---

# FINDINGS

# OVERALL



## UN SWAP

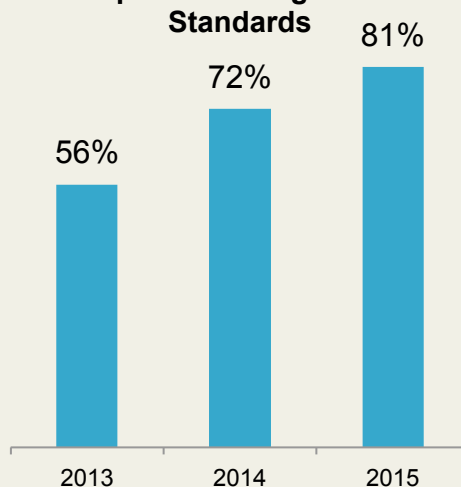
Average score 2015: 6.6

“Approaching Requirements”

Average score 2014: 7.5

In line with UNICEF, UNDP and UNESCO (independently rated or peer reviewed)

## Reports Meeting UNEG Standards



## Trend

The average quality of evaluations has risen year-on-year, with the proportion of reports rated Good or above increasing from 72% in 2014 to 81% for 2015. Once again, no report was rated unsatisfactory.

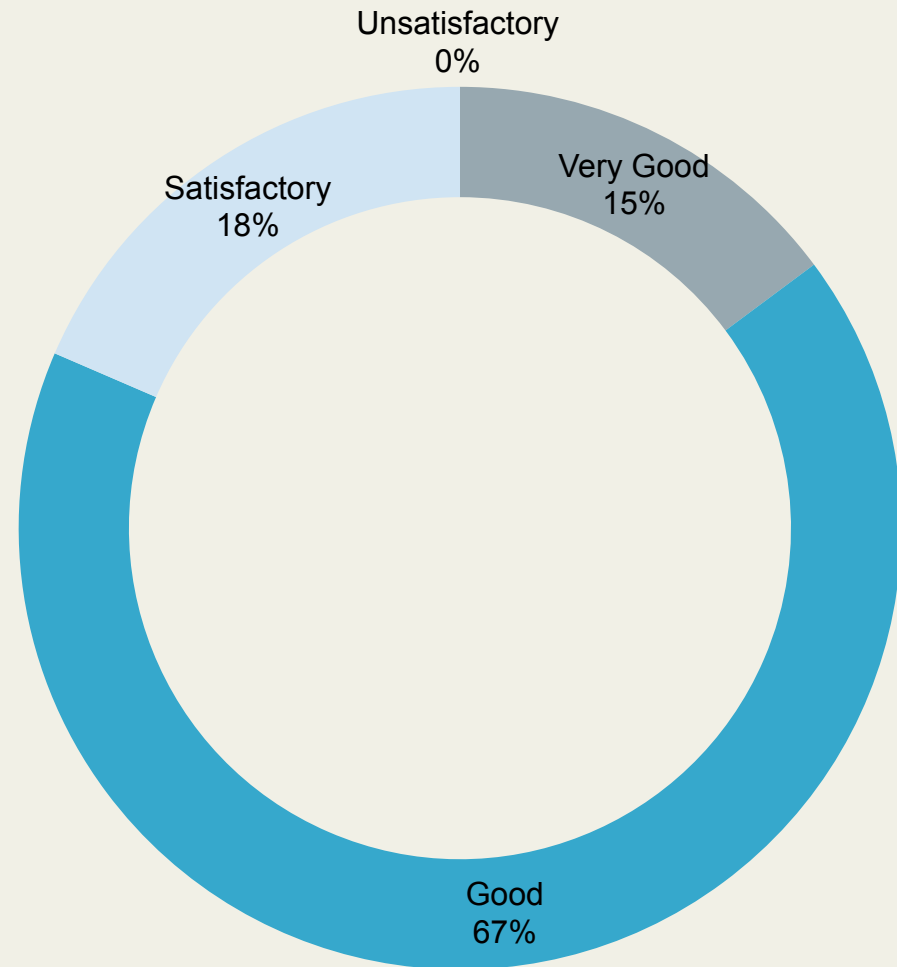
Whilst fewer reports were rated as very good examples of best practice, a large body of reports are now fully aligned with UNEG standards for evaluation reports.

# 2015 EVALUATIONS

Evaluation reports are now well written, well structured, and contain all of the essential elements required by UNEG standards.

As with previous years, nearly all evaluations were based on similar designs that rely primarily on triangulating qualitative data and document analysis.

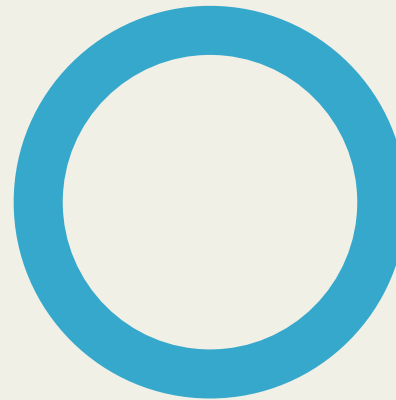
Just over half of reports (56%) were fully compliant with UNEG standards for descriptions of the methodology, and this – along with integration of gender responsiveness – are the key areas where improvements can be made.



---

# HQ

## 2 Reports



Good  
100%

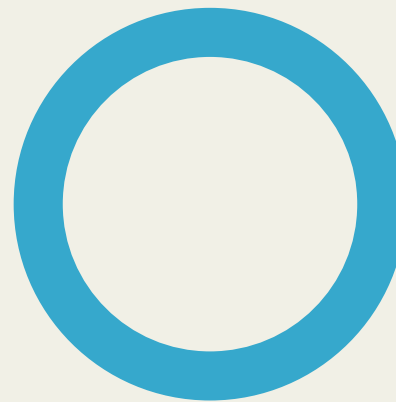
Only two evaluation reports were produced by HQ units, both by the same firm. The structure, background and findings sections of the reports were of good quality. The main gap is the low level of gender integration according to SWAP standards, especially with regard to gender analysis throughout all sections of the report.



---

# ARAB STATES

## 6 Reports



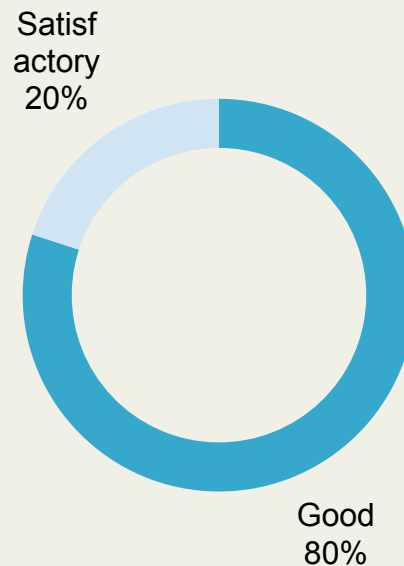
Good  
100%

Arab States region produced the most reports of any region, and all were rated Good. Strengths included the purpose/objectives/scope, the findings, and the overall structure of reports. The main area for improvement is with regard to integration of gender responsive approaches, for which only one report met the required standard.

---

# ASIA PACIFIC

## 5 Reports

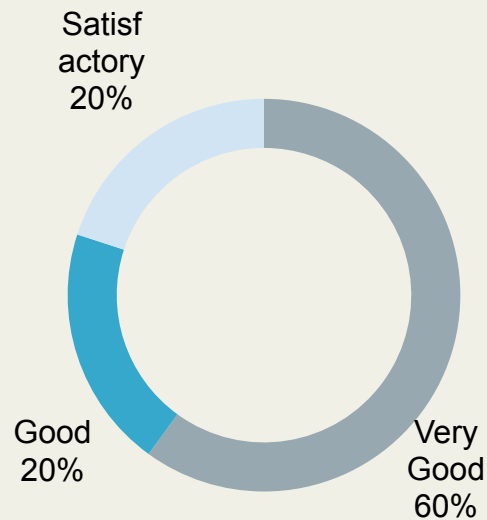


Asia Pacific produced the second highest number of evaluations, and all but one were rated Good (the other was rated Satisfactory). The reports generally had good opening sections. The main areas to gain further quality were descriptions of methodologies, development of recommendations, and integration of gender responsive approaches.

# EUROPE AND CENTRAL ASIA

# 5

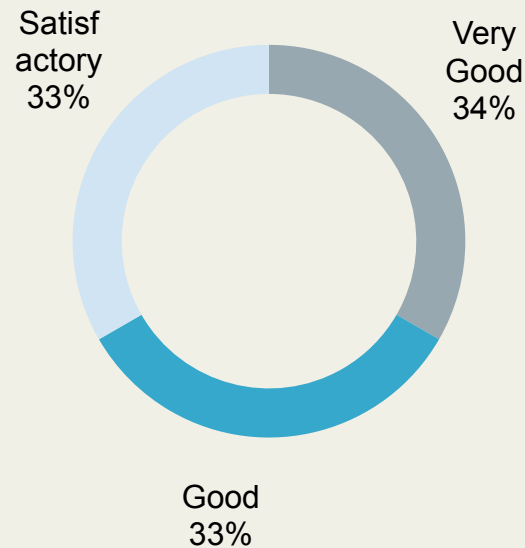
## Reports



Europe and Central Asia produced the most Very Good reports. Strengths of the reports included purpose/objectives/scope and the recommendations section. Also, the region accounted for more reports to meet UN SWAP standards for gender than any other region. The main area to strengthen is the description of methodologies.

# EASTERN & SOUTHERN AFRICA

## 3 Reports

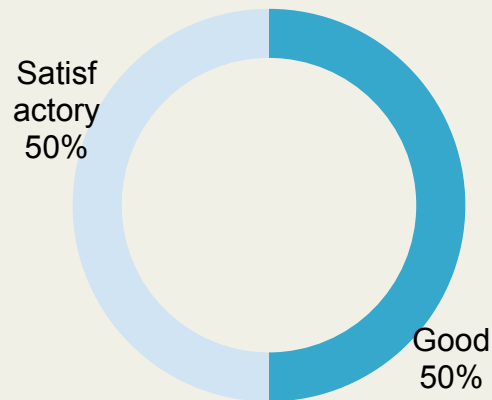


Eastern and Southern Africa produced less reports than in previous years, but they were of a high standard, including the highest rated report for UN SWAP (from Uganda). Reports had strong purpose/objectives/scope, recommendations, and structure. The main area to strengthen is the completeness of conclusions.

---

# LATIN AMERICA & CARIBBEAN

## 2 Reports

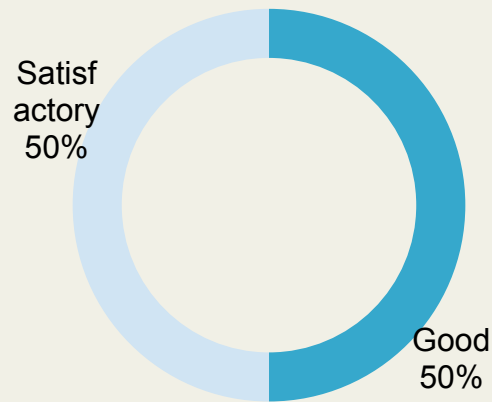


Latin America and the Caribbean region only produced two evaluations for the year, but both were rated Good. The reports were consistent across all of the main parameters, with the main potential for strengthening in relation to the description of the methodology and the integration of gender responsive approaches.

# WEST & CENTRAL AFRICA

# 4

## Reports



The four reports from West and Central Africa were set-up strongly in regard to good quality background and purpose/objectives/scope sections. The main gaps – and areas with highest potential for improvement – are with regard to the description of methodologies, the development of robust findings, and the integration of gender responsive approaches.

---

# STRENGTHS & WEAKNESSES

*For a full analysis with graphs of performance according to report sections, management arrangements, thematic areas, etc, see <http://bit.ly/GERAAS15>*

---

The majority of reports were focused at the country-level and were designed to assess outcome-level results. All thematic area were well covered by the body of evaluations, with an even spread of quality. The main item of concern is the low coverage by evaluations of the contribution of UN Women to global norms and standards (Goal 6).

## **Strength**

Trend analysis suggests that overall performance across most criteria is variable. However, the areas that Regional Evaluation Specialists are most able to influence (the purpose, objectives and scope set in the ToR, the structure, and the recommendations) are both the strongest sections and reflect continuous improvement.

## **Weakness**

The issue of greatest concern is the performance in relation to UN SWAP standards, for which the average score of the overall body of reports is not yet meeting UNEG standards. There were, however, individual reports that were rated extremely highly in regard to UN SWAP (including one from ESARO that exceeded requirements).

# CONCLUSIONS



---

# CONCLUSION

1

*Performance differences between regions are becoming less pronounced, potentially reflecting a consolidation of evaluation management capacity in UN Women.*

Continuing the trend from the previous year, there were no reports that were completely unsatisfactory. There are various potential explanations for this observation that could include: the full complement of Regional Evaluation Specialists helping to establish a 'performance floor' that is approaching full compliance with UNEG standards (the evaluation component of the Regional Architecture was completed in 2014, covering the two years in which no reports have been unsatisfactory); access to better evaluators through an expanding gender-responsive evaluation community (e.g. EvalGender+) of which UN Women is a key convenor (although this is regionally uneven and is not reflected in SWAP performance); or greater attention to evaluation planning in terms of feasibility to conduct evaluations, strategic decision-making to cancel an evaluations when it's clear that the time and resources to conduct a quality process are not present (reflected in the lower delivery rate identified in the audit process, but higher overall quality).

Whilst the precise explanations are not able to be determined from the evidence generated through the GERAAS process, the implication of the 'performance floor' is that the evaluation function is delivering evidence that is increasingly reliable and can be used for the intended purposes of decision-making, accountability and learning. This should be caveated with the observation that particular evaluation criteria are assessed as being higher quality than others. For example, the treatment of efficiency and impact is notably weaker, whilst relevance and effectiveness are strongest (with increasing levels of evidence and analysis pertaining to sustainability).

---

# CONCLUSION

1

There are several potential explanations for the reduction in the proportion of 'very good' evaluations. In particular, there were no corporate evaluations included in the 2015 review (each evaluation accounts for 4-5% of the overall rating, and all previous corporate evaluations have been rated as very good. Whilst the percentage of very good reports is a significant reduction from previous years, it should be recalled that the central purpose of GERAAS is to support all evaluations to meet UNEG standards. Thus, whilst it is positive when more evaluations are very good, achieving a high proportion of very good evaluations is not a central tenet of success for GERAAS, and neither is the mean average rating of reports. The main objective is that all reports should be rated 'green' (Good or Very Good).

Observations from 2015 reveal that an increasing number of evaluation reports meet or exceed UNEG standards, and that therefor the trend is a positive one despite the decrease in Very Good evaluations. This trend is also consistent across regions. The disaggregated analysis of the UNEG parameters suggests that it is built upon strong evaluation purposes, objectives and scope. These elements are all heavily controlled by the Terms of Reference – implying that the guidance provided by Regional Evaluation Specialists is successfully preventing poor-quality evaluations from being undertaken.

---

# CONCLUSION

## 2

*Strengthening and diversifying evaluation designs and methodologies is a key item for the evaluation agenda in UN Women.*

The majority of evaluations commissioned by UN Women apply a similar design: an unspecified method of triangulating qualitative primary data (interviews, focus groups) with quantitative secondary data (monitoring indicators, finance). To some extent this is a reflection of the work being done at country and regional levels or in global programmes – often trying to contribute to national systems or convene stakeholders and evidence around a particular issue. It is also due to the fact that the majority of evaluations are being designed at the mid-point or end of interventions, rather than having interventions designed to be evaluable (analysis from the Asia Pacific region in 2014 identified that a ‘Very Good’ evaluation was designed and planned 4 years prior to implementation to ensure that the required data was available through the RBM system).

As reported in the previous meta-evaluation, the implication of this situation is that nearly all of the evaluative evidence available to UN Women has similar strengths and weaknesses. There is, therefore, a strong case to be made for diversifying the designs used across the UN Women portfolio. This has already been attempted with the introduction of Country Portfolio Evaluation guidance: the three CPEs included in this meta evaluation all focused on questions of the strategic positioning of UN Women at the country level for the first time, and were rated as Good quality.

One approach would be to commission Impact Evaluations as an alternative design (Impact is the OECD DAC criterion that is least-well addressed in the existing body of evaluations), this has significant resource and strategic implications when resources and time are already highly constraining factors in the design of outcome evaluations. Impact Evaluations are more expensive, more time consuming, require specific expertise, and do not provide answers to other questions such as UN Women’s organisational performance. They are suited to answering questions about specific interventions that UN Women may wish to pilot or scale – but there are most likely insufficient numbers of these to diversify the evaluation portfolio.

---

# CONCLUSION

## 2

This suggests that alternative approaches to programme evaluation are required that would further extend the quality of evaluation reports and capture the dimension of impact in the context of UN Women's interventions. Given the resource constraints and complexity faced by UN Women evaluations, this will be a challenge and will require stepping outside of quantitative quasi-experimental methods for achieving validity. In view of the current SWAP indicators, elaboration of alternative options for programme evaluation would certainly provide an opportunity to explore and strengthen the greater use of gender responsive designs.

Identifying relevant approaches and developing guidance is an important step in diversifying the evaluation base. As previously noted, this has already begun through UN Women's publishing of the Evaluation Handbook, CPE Guidance, and wider support to EvalPartners learning materials. However, observations of Regional Evaluation Specialists suggests that model approaches and guidance can also be applied in a 'cut-and-paste' manner at country-level, without adopting evaluation frameworks or methods to the particular context. The key challenge for the Independent Evaluation Office is, therefore, to develop a tool that makes diverse approaches accessible and adaptable to programme staff-members who are not evaluation specialists. One tool that was explored in the context of the CPE guidance was a decision-tree to guide the identification of new and appropriate evaluation approaches. There is a case for revisiting this approach in a broader context as part of the UN Women evaluation agenda.

---

# CONCLUSION

## 3

*The highest priority for strengthening the evaluation system is in regard to gender responsive evaluation.*

Unlike in previous years, reports reviewed for GERAAS 2015 included only decentralised evaluations: 25 managed by country/regional offices and 2 global programme evaluations managed by headquarters. There were no corporate evaluations (i.e. those managed directly by the Independent Evaluation Office) included in the sample. Comparison with the previous year reveals a similar trend in terms of the four UN-SWAP indicators: with stronger performance in terms of gender responsive criteria, questions, scoping and indicators, and weaker performance in regard to methods and gender analysis. Overall, however, the reports rated 6.6 (at the top end of “Approaching Requirements”), representing an average 0.9 point drop over the previous year. This was spread evenly across the four criteria, suggesting that the difference was in the overall body of reports rather than with regard to specific aspects of gender responsive evaluation.

Within this overall rating, a number of trends emerge. In particular, significant improvement in the performance of Eastern and Southern Africa and Asia and the Pacific regions, with one evaluation of the Joint Gender Programme in Uganda rating 11 points out of a maximum 12. GERAAS 2015 is the first to include examples of Country Portfolio Evaluations – decentralised evaluations undertaken to specific IEO guidance – and these represented stronger overall performance, supporting the case for the expansion of this approach. One area for further inquiry as a result of the 2015 UN-SWAP analysis is the scope for strengthening of gender in the francophone body of evaluations, particularly in relation to gender responsive methods and analysis (reports in English, Spanish and French were all rated by the same multi-lingual reviewers). The review notes that the launch of the new UN Women professionalization initiative, including guidance on managing gender responsive evaluation took place during the period covered by GERAAS 2015, and is thus unlikely to have been available early enough to affect these results.

---

# CONCLUSION

## 3

Arguably, UN Women needs to be the global leader on gender responsive evaluation. Current data and trends for the UN SWAP indicators on evaluation suggest that this is not the case: even accounting for the known variability with which these indicators are applied across different entities. However, the challenges faced by evaluation commissioners should also not be underestimated. Most UN Women decentralised evaluations are undertaken by a single evaluator or pair of evaluators over a short timeframe and with a limited budget relative to other multilateral entities. The evaluations focus on policy-level, normative, coordination and agenda-setting work. This is a challenging context in which to engage rights holders or easily assess disaggregated impacts on different social groups.

Despite these challenges, there are important implications for UN Women of being able to demonstrate leadership in applying gender responsive evaluation beyond the corporate level. To address this, the outstanding examples of integrating gender into evaluation processes and reports – such as the Joint Programme Evaluation in Uganda – need to become the norm. During 2015, the Independent Evaluation Office launched guidance on managing gender responsive evaluations to address precisely this issue. The evidence from the meta evaluation ratings suggests that implementing this guidance will require the same level of organisational follow-up as has been demonstrated with regard to evaluation objectives and recommendations.

## Part 4

---

Recommendation 1

---

Recommendation 2

---

Recommendation 3

---

Recommendation 4

---

Recommendation 5

---

Recommendation 6

---

Recommendation 7

---

# RECOMMENDATIONS

---

# RECOMMENDATION 1

*Oversight of UN Women evaluation reports in 2016 should prioritise, first and foremost, meeting and exceeding UNEG Design Standard 3.7 ('... Methodology should explicitly address issues of gender and under-represented groups') and UNEG Report Standard 4.8 ('The evaluation report should indicate the extent to which gender issues and considerations were incorporated where applicable') in order to satisfy UN SWAP standards.*

---

The UN SWAP indicators address gender responsiveness in evaluation reports in two main dimensions:

1. Gender responsiveness in the evaluation method and process; and
2. Assessment of gender in the object of the evaluation.

Based on the type of evaluations included in the UN Women portfolio, reports will meet or exceed UN SWAP standards where they include a number of practical features. It is recommended that this list is used by IEO for quality assurance and shared with evaluators to improve the performance of UNW evaluations in SWAP:

1. A specific reference in the Objectives of the evaluation to assessing how gender was mainstreamed in the design of the object of the evaluation;
2. One or more evaluation questions specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved;
3. A standalone criterion on gender and/or human rights in the evaluation framework;
4. Mainstreaming of gender into one or more indicators under other evaluation criteria – by being gender-disaggregated, gender-specific (relevant to a specific social group), or gender-focused (concerning relations between social groups);
5. Inclusion of evaluation sub-questions and/or criteria that address participation and social inclusion in UN Women interventions;
6. A mix of quantitative and qualitative indicators in the evaluation framework;



---

# RECOMMENDATION 1

7. A background section that includes an intersectional analysis of the specific social role groups affected by the issue that is being addressed by the evaluation object. The best reports attempt to quantify the size of these groups and to differentiate the ways in which they are affected by a particular issue;
8. Presentation or reconstruction of the theories of change used by the intervention and subjecting these ToCs to feminist critical analysis;
9. Description of an evaluation design that includes substantial utilisation-focused and participatory elements – including the participation of a range of duty-bearers and rights-holders in scoping the evaluation and making meaning from evaluation data (i.e. not just being a source of data);
10. A statement in the main report or the annexes that explains how data collection protocols ensured that women and men were included in ways that avoid gender biases or the reinforcement of gender discrimination and unequal power relations;
11. Data analysis in all findings that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data; and
12. At least one finding, one conclusion, and one recommendation that explicitly address the extent to which the intervention contributes to transforming the structural relationships between the social role groups identified in the background section of the report.

---

# RECOMMENDATION 2

*As a first step to addressing limited practice of gender responsive evaluation, IEO should develop a mandatory and time-bound pre-assessment of an evaluator's knowledge of gender responsive evaluation as part of the recruitment process, in accordance with UNEG ethics Norm 11 ('In light of the United Nations Universal Declaration of Human Rights, evaluators must be sensitive to and address issues of discrimination and gender inequality').*

---

To give real emphasis to gender responsiveness in evaluation design and implementation, it is recommended that UNW no longer relies on the CV/profile of evaluators alone but complements this with a mandatory online test on gender responsive and human rights based evaluation for all evaluators (individuals or team members of firms) engaged in an UN Women managed evaluation. Drawing on the experience of UNDSS basic and advanced security training in the field certifications, evaluators should be given the opportunity to retake the test until they achieve a pass, and certification should be time limited in validity so as to contribute to refreshing awareness of established and emerging gender-responsive techniques.

Whilst such a test will be insufficient to 'teach' gender-responsive evaluation, it will be sufficient to ensure that evaluators are explicitly aware of the standards expected by UN Women, and could help to prompt further individual learning on issues or approaches that the evaluator does not feel fully conversant in. The course could also be made publicly available for other UN entities and development organisations to make use of – thereby contributing to UN Women's systems strengthening mandate.

---

# RECOMMENDATION 3

*IEO should develop 2-3 practical briefs on alternative designs to programme evaluation and make these available to evaluation commissioners through the RES system.*

---

Given the homogeneity of UN Women's current evaluation portfolio, it is recommended that practical steps be taken to avoid the risk of the organisation's evaluative evidence base being subject to the same set of strengths and limitations. Building on the existing work to elaborate approaches to country portfolio evaluations, this might include providing several different models for programme evaluations that can be specified in future terms of reference. These can be broad enough to allow for evaluator-interpretation and refinement, but still provide a coherent intellectual framework.

For example, IEO could choose to elaborate a highly participatory or democratic programme evaluation design that combined techniques such as photo stories, participatory video, and collaborative outcomes reporting. It could juxtapose this with a more quantitative evaluation design than is currently the norm, drawing on techniques such as social return on investment. These briefs should include specific guidance on implementing the model designs within a gender responsive paradigm.

---

# RECOMMENDATION 4

*Develop and maintain a list of good examples and FAQ on methods for evaluating UN Women's normative, partnership and coordination work.*

Across the spectrum of UN Women's integrated mandated, there is a strong tendency for operational (programmatic) aspects of UN Women interventions to feature most frequently and prominently in evaluative analysis. It can be hypothesised that this is the case both because evaluators are more familiar with this modality, and because methods for evaluating programmes are more numerous and more advanced than the other elements of the UN Women mandate. Since evaluation of normative, coordination and partnership working should, most often, be integrated with evaluation of operational work it is not recommended to develop standalone-guidance on specific evaluation methods for these aspects.

Drawing from experiences captured in evaluations of UN Women's policy and programme teams, one promising option for providing useful (and flexible) guidance is to publish a maintain a list of frequently asked questions (FAQs) and practice case studies of existing work in the area of interest. With regard to normative, coordination, and partnership evaluation, IEO could synthesise lessons and concrete examples from the existing database of evaluations on GATE, as well as examples from other entities. The ultimate aim would be to produce and communicate a list of practical 'tips' for evaluation commissioners, managers, and team leaders on better evaluation of these essential aspects of UN Women's work.

---

# RECOMMENDATION 5

*Provide evaluation managers with clear guidance on expectations for ethics in evaluation reports through the RES system.*

---

To address the continued low-rating of ethics in evaluation reports (something that is not unique to UN Women) it is recommended to provide practical guidance to evaluation managers and evaluators on concrete steps to meeting UNEG guidance. This is considered appropriate as there is no evidence to suggest that UN Women evaluations are fundamentally unethical in terms of process, simply that reports are not fully capturing ethical standards as required by the standards. To address this recommendation, the following table can be reviewed and adapted by IEO before being distributed through the RES system and included as a tool in the report quality assurance process of future evaluations. The updated table should also be shared with the GERAAS review team to ensure that a common approach to assessing ethics is adopted and to avoid focusing solely on consent and protection of evaluation participants.

# RECOMMENDATION 5

Principle	How to implement in evaluation report	When important
Utility	<ol style="list-style-type: none"> <li>1. Identification of users and uses in the purpose</li> <li>2. Identification of stakeholder groups and their interests</li> <li>3. Evaluability analysis</li> <li>4. Specification of “utilisation-focused approaches” [optional]</li> </ol>	All evaluations
Necessity	Clear definition of purpose and objectives	All evaluations
Independence	Identification of possible inhibitors to independence in a statement on evaluability	All evaluations should include summary of evaluability analysis
Impartiality	<ol style="list-style-type: none"> <li>1. Include stakeholder mapping and collect data from multiple groups of stakeholders</li> <li>2. Specify a recognised approach to data analysis in the methods section</li> <li>3. Present findings and evidence that are transparently based on evaluative analysis</li> </ol>	All evaluations
Credibility	Clearly state methodological limitations and gaps in the data used by the evaluation	Methods section of all evaluations
Conflicts of interest	Clearly state that no conflicts of interest exist in a paragraph synthesising the evaluability analysis, or clearly state mitigating actions when they do	All evaluations

# RECOMMENDATION 5

Principle	How to implement in evaluation report	When important
Respect for diversity	Culturally sensitive evaluation data collection instruments and processes in annexes.	Especially participatory evaluations and surveys of rights holders
Right to self-determination	Paragraph describing recognised process of free, prior and informed consent	Evaluations working with rights holders and marginalised groups
Fair representation	Inclusive sampling of multiple stakeholder groups, including marginalised groups in methods section.	Programme evaluations, participatory evaluations
Protection of vulnerable groups	Specific statement citing codes to protect vulnerable groups included in the evaluation process, such as young people, sex workers, survivors of violence, migrants, etc	All evaluations, especially those with control groups
Redress	Specific statement in a paragraph in the methods section or the annexes stating that rights holders and other consulted groups were provided with options to register complaints	Especially participatory and impact evaluations
Confidentiality	Evaluation reports protect the identity of participants in the findings	All evaluations

---

# RECOMMENDATION 6

*For the majority of evaluations, the efficiency criterion should be more tightly focused on 'organisational efficiency' to generate useful learning.*

Whilst it is not captured by the ratings data, the reviewers observed that efficiency findings are frequently weaker than the other DAC criteria (with the exception of impact). In many cases this appears to be because either cost data for an intervention is not available, or there is a lack of evidence on what represents fair value for the type of interventions that UN Women engages in (convening, movement building, policy advocacy, coordination, etc) across different contexts. In most cases, this results in 'shallow' findings on efficiency, that are based on small amounts of weak data.

On the basis of this, it is recommended to learn from the experience of UN Women's corporate evaluations and to encourage all decentralised evaluations to refocus the efficiency criterion around organisational efficiency. Within organisational efficiency, there are far more established frameworks for assessing whether the organisation is fit-for-purpose and examples of these applied to UN Women successfully exist within the existing body of corporate evaluations.

Given the evolving nature of the regional and global architecture of UN Women – including the shift to the flagship programme initiatives – it is proposed that insights into organisational efficiency would be of more immediate benefit to the overall performance of the entity. Furthermore, the suggestion in the ratings that the IEO RES system is contributing to more consistent evaluation frameworks means that the mechanism to propagate a policy of concentrating of organisational efficiency already exists.



---

# RECOMMENDATION 7

*Apply OECD DAC 'impact' criterion more selectively, and replace this as a default standalone criterion on gender and human rights.*

Within the context of UNEG standards, impact refers to long term changes in peoples lives – which for UN Women equates to the progressive realisation of women’s human rights. Given UN Women’s integrated mandate and the type of interventions that result from this, the chain of contribution to impact is long in terms of both causation and time. As a consequence, the majority of evaluations are commissioned too soon to meaningfully identify or isolate impact – and state this in the final reports.

Considering this situation, and the unlikelihood that UN Women evaluations will significantly change in terms of either timing or methods, it is recommended that impact is replaced as a default evaluation criterion with a standalone criterion on gender equality and human rights. Impact can still be assessed in selective cases where it is appropriate, useful and feasible to do so. At the same time, it is proposed that a standalone criterion on GEHR would help to strengthen performance in SWAP, would help give greater emphasis to the SDG commitment that no person should be left behind, and would also capture an early indicators of impacts that are of interest to UN Women.

To implement this recommendation, it is recommended that IEO update the evaluation policy, evaluation management handbook, and other evaluation guidance to reflect the new prioritisation of criteria.