UN WOMEN

GLOBAL EVALUATION REPORT
ASSESSMENT AND ANALYSIS
SYSTEM

2017 META EVALUATION

March 2018

PREPARED FOR UN-WOMEN INDEPENDENT EVALUATION OFFICE
Messay Tassew

PREPARED BY IMPACTREADY LLP
Joseph Barnes, Maria Borisova

INCLUDES INDEPENDENT RATINGS BY
Katherine Garven, Fernando Garabito

Contents

1. ACRONYMS

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

2. EXECUTIVE SUMMARY

2.1 BACKGROUND

The Purpose of this meta-evaluation is to report on the quality of evaluation reports from 2017, and the trends in evaluation quality since 2013. Evaluation in UN Women contributes to learning on the best ways to promote gender equality and women's empowerment, enhancing UN Women's accountability, and informing decision-making. To address the organizational demands for ensuring good quality and credible evaluations particularly at decentralized level, the Independent Evaluation Office has designed a Global Evaluation Reports Assessment and Analysis System (GERAAS).

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

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

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

2.2 FINDINGS

While evaluation coverage is widespread, there are still some gaps in terms of programme-presence countries that have not been included in evaluations. The meta-evaluation assessed 39 reports from 2017; an increase from the 36 reports from 2016, 27 reports rated from 2015 and 21 from 2014. Over the course of the 2014-2017 Strategic Plan, UN Women completed a total of 123 evaluations[1]. 34 countries[2] with programme presence were explicitly covered by evaluations in 2017. 5 countries with programme presence were included in evaluations in 2017 that have not previously be covered, with 4 of these in Arab States Region. The most covered countries in 2017 were Egypt and Kenya (5 evaluations each), Moldova (4 evaluations), and Mexico, Jordan and the State of Palestine (3 evaluations each).

The number and percentage of evaluation reports meeting UN standards continues to improve. Over the period of the Strategic Plan 2014-2017, no evaluation has been found to be unsatisfactory according to UNEG/UN Women standards. The level of reports rated Good or Very Good appears increased to 95%, meaning that nearly all evaluations can now be used with confidence. The remaining evaluation reports (5% rated Satisfactory in 2017) can still be used for management, accountability, and learning; but with awareness of the limitations that they face.

---

[1] 146 evaluations have been rated in the four years since GERAAS began, covering 2013-2017
[2] This figure includes countries covered through country case studies for corporate, regional and HQ evaluations

Good quality evaluation evidence is now available at all levels and across all themes. Project and programme evaluations continue to dominate the overall portfolio, although 2017 included 1 corporate (global), 1 global, 3 multi-country, and 5 regional evaluations. The scope of 74% of evaluations is at the country-level; with 93% of these fully meeting UN-Women standards. The only two reports to rate as Satisfactory were country-level project evaluations. As with previous years, the most evaluated impact area was Women's Economic Empowerment (included in 19 evaluations); although the same number of evaluations also covered Ending Violence Against Women in 2017. The least evaluated area was the Normative Framework (included in 7 evaluations). Within the regional architecture, decentralised evaluations have, again, increased in number to reach 34 in 2017 (6 more than in 2016).

Evaluations continue to improve in terms of meeting UN-SWAP standards for integrating gender equality and human rights. The 2017 meta evaluation found a significant increase in performance of evaluation reports with regard to the UN-SWAP Evaluation Performance Indicator (EPI), with the highest recorded performance for UN Women since tracking of this indicator began. Overall, UN Women evaluations were found to be 'meeting requirements' as defined by the UN Evaluation Group, with an average performance rating of 10.08 (out of a maximum score of 12). 13 evaluation reports were rated as exceeding the UN-SWAP standards.

There is increased explanation of overall evaluation designs and approaches in reports, accompanied by emerging examples of innovation. Previous meta evaluations have found a high level of homogeneity in UN-Women evaluation designs (with a high prevalence of qualitative approaches). While this remains the case for project evaluations, the 2017 meta evaluation found many examples of innovation in terms of extending the range and application of evaluation approaches.

2.3 CONCLUSIONS

Conclusion 1: The quality of evaluation reports in 2017 continues to deepen; with more reports more consistently and more fully meeting UN-Women standards across more parameters. Based on this pattern, UN-Women is on track to have 100% of evaluation reports meeting UNEG standards in the next one or two cycles, if it maintains the current level of annual improvements.

Conclusion 2: Good quality evaluation evidence is available at all levels and across all strategic impact areas; but some gaps in geographical coverage of programme presence remain. During the Strategic Plan 2014-2017, most countries with UN-Women office presence or programme presence were directly covered by at least one evaluation. Where evaluative evidence is available, it is of increasingly high levels of quality – meaning that it is reliable and useful for decision-makers. There is also broad coverage across all of the strategic impact areas. The Strategic Plan 2018-2021 represents an opportunity to address and prevent coverage gaps of 11 countries in the future.

Conclusion 3: The while the range of evaluation designs remain mostly qualitative, there are more examples of innovative approaches and methods being used. Two important patterns were identified: (1) reports are more explicitly and transparently discussing the evaluation design (instead of just listing methods), and (2) there are a noticeable number of evaluation reports based on innovative and creative evaluation approaches. These represent useful examples for future evaluations to learn from, and to be inspired by.

Conclusion 4: UN-Women evaluations continue to meet the requirements of the UN-SWAP evaluation performance indicator, and are demonstrating improvements across all criteria. This represents an important achievement; much of which is connected with improvements in the methods and analysis observed in reports during this period. The easiest way to further improve on UN-SWAP performance – and to move closer towards an overall rating of 'Exceeds Requirements' – is to ensure that all reports include an evaluation matrix with indicators that explicitly relate to gender equality and women's empowerment.

Conclusion 5: The use of a weighted quality assessment tool has improved consistency of report ratings; but the GERAAS tool needs more comprehensive revision to still be relevant in the future. The GERAAS assessment tool was updated for 2017 to include weighted parameters and automatic calculation of parameter and overall ratings. This has been successful in terms of consistency; but other important considerations have emerged over the duration of GERAAS that warrant a more fundamental revision.

2.4 RECOMMENDATIONS

Developed based on the findings and conclusions, and validated by IEO.

Recommendation 1: The Independent Evaluation Office is advised to update the GERAAS evaluation quality assessment template/tool to align it with the UN-Women Strategic Plan 2018-2021, and UNEG Norms and Standards 2016.
GERAAS has used the same template and assessment tool throughout the Strategic Plan 2014-2017 cycle; which has provided consistency, but also shows signs of needing to be updated. The Strategic Plan 2018-2021 cycle is an opportunity to do this. This part of the recommendation can be implemented by: (1) aligning with the Strategic Plan 2018-2021 outcomes, (2) revising existing assessment indicators to reduce overlap and to align with the latest UNEG standards, (3) assessing lessons learned separately from other parameters, (4) separating indicators for findings to disaggregate the quality assessment of OECD-DAC criteria, (5) integrating guidance on evaluating humanitarian action.

A consultative process with regional evaluation specialists, corporate evaluation managers, and other stakeholders is recommended to: (1) agree the weighting of each parameter with the Regional Evaluation Specialists and corporate evaluation managers, (2) disaggregating the performance rating of the UN-Women evaluation function and independent evaluators, and (3) differentiating independent from internal evaluations[3].

Recommendation 2: The Independent Evaluation Office is recommended to adjust the process and timing of GERAAS evaluation quality assessment to rolling near-time reviews.
'Near-time' quality assessment has been initiated by several other member entities of the UN Evaluation Group in recent years to improve usefulness. While this approach is sometimes referred to as 'real-time' reporting, contemporary experiences in other UN entities shows that 'near-time' (within 6 weeks) is more feasible. To avoid the need for expensive 'stand-by' capacity to assess report quality in 'real time' (due to the uncertainties of demand), it is therefore recommended that UN-Women consider a transition to 'near-time' quality assessment for the Strategic Plan 2018-2021 cycle.

---

[3] Independent evaluations are those evaluation reports that have been fully quality assured through the involvement of regional evaluation specialists, internal evaluations are those produced directly by business units.

Recommendation 3: The Independent Evaluation Office is recommended to further develop evaluation guidance and report templates, especially for project evaluations.

Project-level evaluations have the smallest teams, and rarely include evaluators with international experience. Based on the consistent quality of country portfolio evaluation reports (which benefit from guidance and templates), it is recommended that a similar set of templates and guides are provided (in multiple languages) for country offices to provide local project evaluators. These should be geared to the low level of resources and time associated with many project evaluations. While the use of these proposed templates can remain optional; they are likely to become used by default if they make producing a high quality report easier.

Recommendation 4: Regional Evaluation Specialists are recommended to jointly develop a shared knowledge product explaining the current successful approach to quality assuring evaluation reports, to help sustain current positive trends in the future.

The meta evaluation reveals a strong positive trend across nearly all indicators since quality assessment began in 2013. Much of this improvement has been achieved at the decentralised level, and a large contribution has been the support provided by regional evaluation specialists. To ensure that this capability – and the positive trend – is maintained, it is strongly recommended that a means is found to systematise the knowledge and practices established by regional evaluation specialists. One option could be a joint knowledge project, capturing the experiences of the current cadre, to be shared with new evaluation specialists as part of the handover process.

Recommendation 5: UN-Women Senior Management, with the support of the Independent Evaluation Office, is recommended to ensure full coverage of all programme presence countries in independent evaluations during the course of the UN-Women Strategic Plan 2018-2021.

The meta evaluation highlights the strengthening resource of good quality evaluation evidence that is available to support UN-Women and its partners. Despite growth in the number of evaluations, and coverage of individual evaluations (e.g. more regional evaluations), 13% of programme countries were not directly covered during the Strategic Plan 2014-2017. It is recommended, therefore, that a consultative but centrally-led strategy is established early in the current strategic plan cycle to ensure that all programme countries are directly included in at least one – but ideally several – evaluations. This coverage should be tracked as a key performance indicator in GERAAS.

Recommendation 5: UN-Women is recommended to link GERAAS to tracking of management response implementation rates as a proxy indicator for use.

While GERAAS provides a proxy indication of evaluation quality, it does not give insight into the utility – or use – of evaluations. However, all evaluations have a management response, the implementation (and tracking) of which is established under UNEG standards as being the prerogative of 'management' (not the evaluation function). The acceptance or rejection of evaluation recommendations, and the completion (or not) of management response commitments, have both been used by other UN entities as proxy indicators for evaluation use.

While this is an imperfect solution to better understanding the use of evaluations, it does provide more insights than is currently available. It is thus recommended that an internal review of the management response system is undertaken, with the aim of assessing the feasibility and value of linking this to GERAAS.

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 average size of evaluation teams during 2017 was 2 people, most typically comprising an international evaluator and a national evaluator. National evaluators were found to be equally likely to be male or female; whereas 75% of international evaluators were female. The average size of an evaluation team at country-level was 1.75 people; whereas at global level it was 2.75 people[4]. Nearly all evaluations are managed by UN Women staff, with only 3 evaluations of joint programmes that appear to have an element of joint management.

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

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

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

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

In response to conclusions from 2016, the GERAAS tool for 2017 introduced explicit weighting between parameters in order to ensure emphasis on the substantive aspects of evaluation quality (and to minimise subjectivity).

An independent firm was appointed to undertake both a meta-evaluation and meta-analysis of 2017 evaluation reports submitted to GERAAS, including the UN-SWAP scores[5]. A total of 38 reports were included.

---

[4] Overall, this size of evaluation teams are consistent with previous years; and are comparable with similar sized entities in the UN Development Group.
[5] ImpactReady LLP

4. PURPOSE & SCOPE

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

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

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

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

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

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

This assessment considers all 2017 reports submitted to the GERAAS system that were assessed, according to the UN Evaluation Group definition, to be evaluation reports (rather than reviews, evaluability assessments, baselines, studies, etc).

It considers only the evaluation report, as presented on the UN Women GATE system (http://gate.unwomen.org) as a standalone document. The actual evaluation process or utilisation of the evaluation is currently considered outside the scope of this analysis. It should be recognised, therefore, that this report only provides a partial view in answer to the question *"what is the quality of evaluation in UN Women?"*

5. METHODOLOGY

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

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

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

Reports are rated, analyzed and synthesized through a 5-step process that has been refined over the course of GERAAS 2013-2017. This meta evaluation assesses final evaluation reports from fiscal year 2017 that were uploaded in the UN Women Global Accountability and Tracking of Evaluation System (GATE) by January 2018. It includes a rating of all reports using the UN SWAP evaluation performance indicator to the standards agreed by the UNEG working group on gender equality and human rights.

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

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

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

5.1 LIMITATIONS
GERAAS does not measure the quality of evaluation processes. It is designed to assess the quality of *reports* – considered a major output of evaluations – and it does so against a very

specific and prescriptive set of UNEG standards for what an evaluation report should look like.

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

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

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

6. FINDINGS

6.1 OVERALL RATINGS AND FEEDBACK

Finding 1: While evaluation coverage is widespread, there are still some gaps in terms of programme-presence countries that have not been included in evaluations

The meta-evaluation assessed 39 reports from 2017; an increase from the 36 reports from 2016, 27 reports rated from 2015 and 21 from 2014. Over the course of the 2014-2017 Strategic Plan, UN Women completed a total of 123 evaluations[6].

34 countries[7] with programme presence were explicitly covered by evaluations in 2017. 5 countries with programme presence were included in evaluations in 2017 that have not previously be covered, with 4 of these in Arab States Region. The most covered countries in 2017 were Egypt and Kenya (5 evaluations each), Moldova (4 evaluations), and Mexico, Jordan and the State of Palestine (3 evaluations each).

Figure 1: Explicit coverage of countries in 2017 evaluations (including as case studies)
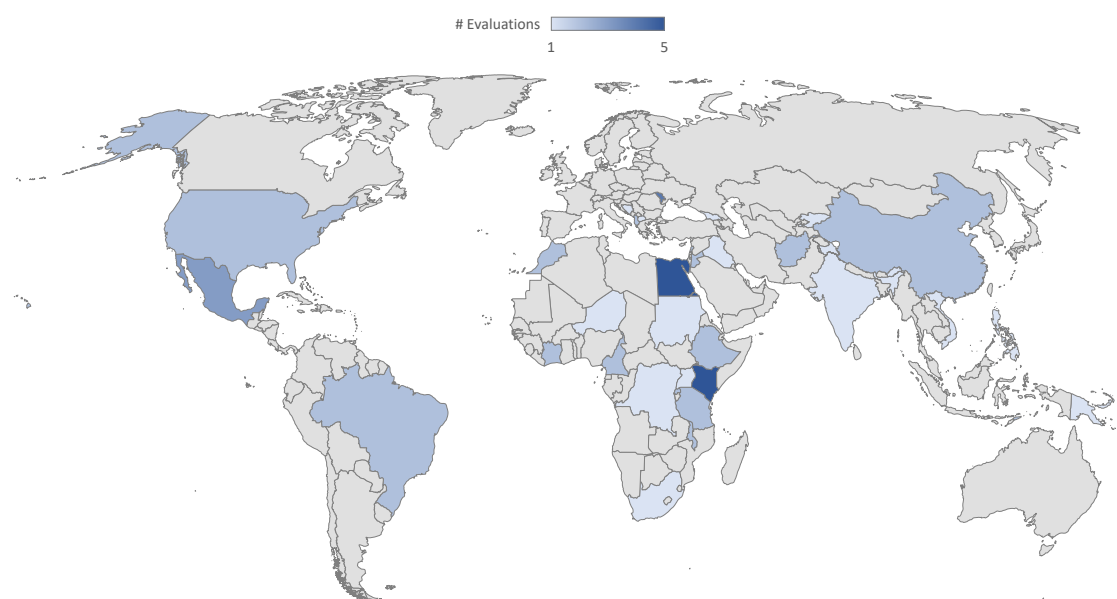
Of the 87 countries with programme presence at the beginning of the 2014-2017 Strategic Plan, 76 have been directly covered in evaluations (either through a country-level evaluation or as a case study); with 11 not covered[8].

---

[6] 146 evaluations have been rated in the four years since GERAAS began, covering 2013-2017
[7] This figure includes countries covered through country case studies for corporate, regional and HQ evaluations
[8] Somalia, Cape Verde, Guinea Bissau, Algeria, Yemen, Mauritania, Maldives, Bhutan, Slovakia, Honduras, Barbados (Caribbean).

Figure 2: Country coverage of UN-Women evaluations 2013-2017

Legend:
- Evaluated 2013–2016
- Evaluated 2017
- Evaluated 2017 & previously
- Programme presence countries (2013) not covered in evaluations
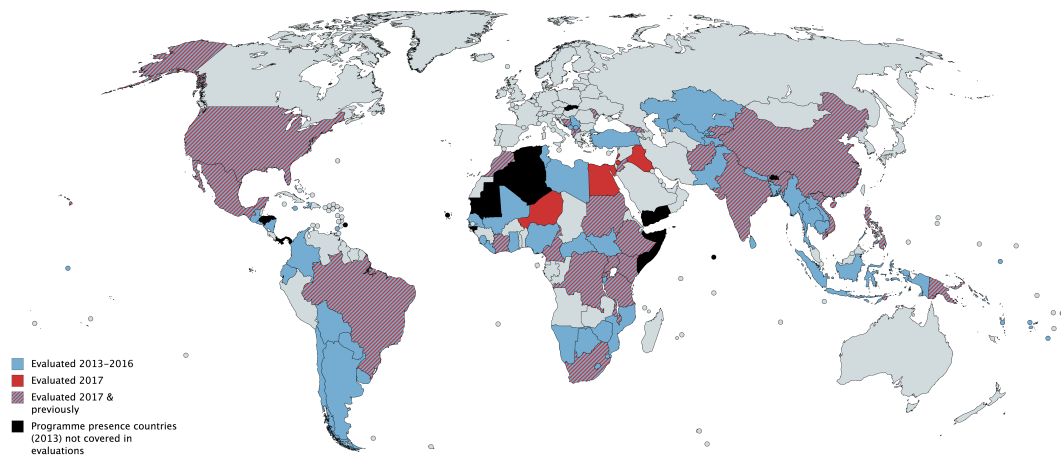
Finding 2: The number and percentage of evaluation reports meeting UN standards continues to improve

Over the period of the Strategic Plan 2014-2017, no evaluation has been found to be unsatisfactory according to UNEG/UN Women standards. The level of reports rated Good or Very Good appears increased to 95%, meaning that nearly all evaluations can now be used with confidence. The remaining evaluation reports (5% rated Satisfactory in 2017) can still be used for management, accountability, and learning; but with awareness of the limitations that they face.

Figure 3: Overall ratings of evaluation report quality 2013-2016

2013	2014	2015	2016	2017
25%	42%	15%	33%	36%
30%	29%	66%	50%	59%
30%	29%	19%	17%	5%
15%				
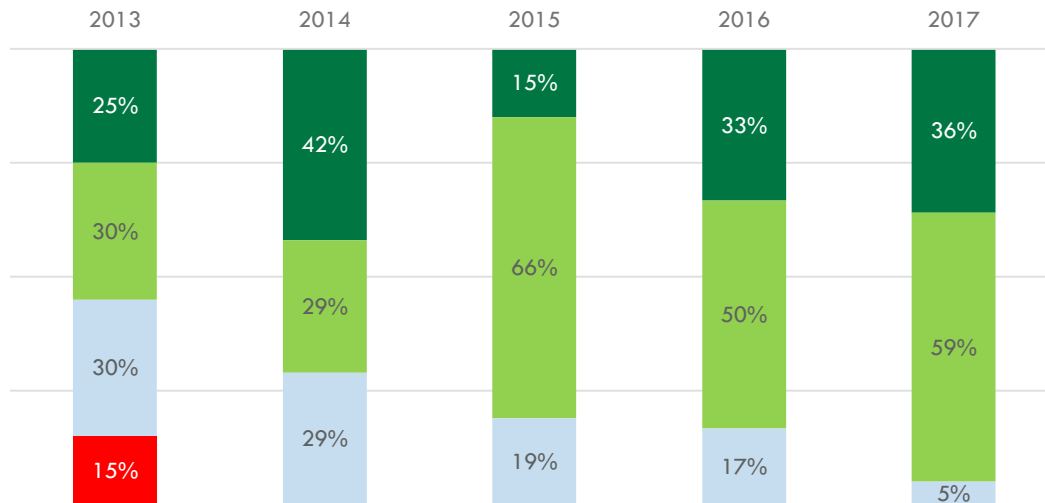
Finding 3: Good quality evaluation evidence is now available at all levels and across all themes

Project and programme evaluations continue to dominate the overall portfolio, although 2017 included 1 corporate (global), 1 global, 3 multi-country, and 5 regional evaluations. There were 11 evaluations that were either Country Portfolio Evaluations or covered an entire thematic area of a country Strategic Note. Of these 'strategic' evaluations, 64% were rated Very Good, and 36% rated Good; meaning that they all fully meet UN-Women standards.

Figure 4: Performance across different levels of evaluation in 2017

	Global	Regional	Multi-country	National	Sub-national
■ Very Good	2	2	1	8	1
■ Good		2	2	15	4
■ Satisfactory				2	
■ Unsatisfactory					
Grand Total	2	4	3	25	5
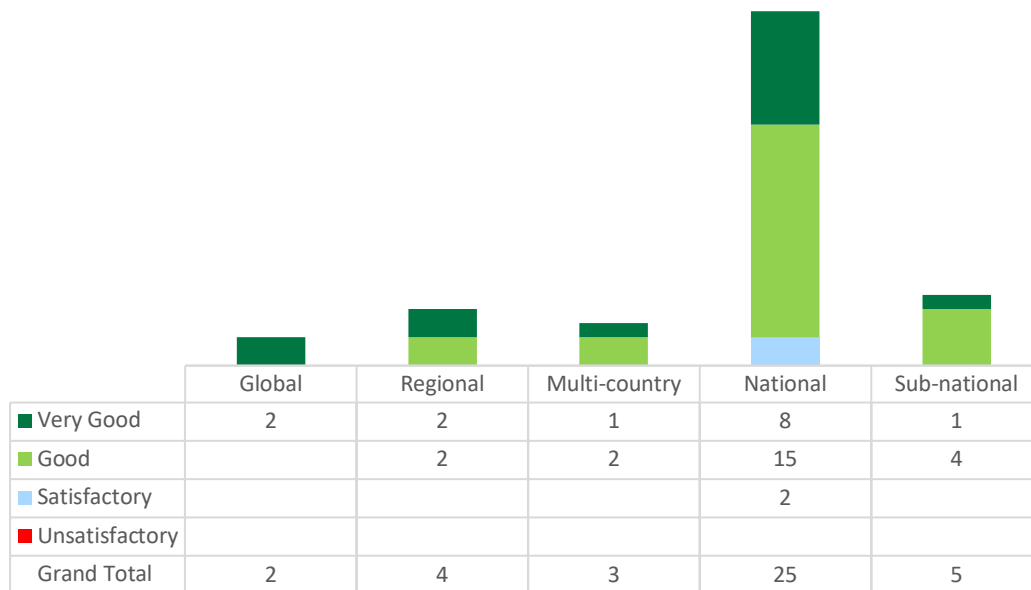
All global evaluations and policy/strategic evaluations rated as Very Good. Of the evaluations of UN-Women Strategic Notes (either Country Programme Evaluations or thematic evaluations), 55% rated as Very Good, and 45% as Good. For regional evaluations, 40% rated as Very Good, and 60% as Good. The scope of 74% of evaluations is at the country-level; with 93% of these fully meeting UN-Women standards. The only two reports to rate as Satisfactory were country-level project evaluations.

Figure 5: Performance across different types of evaluations in 2017

	Policy / Strategic	Country Portfolio / Thematic	Programme	Project
■ Very Good	3	6	2	3
■ Good		5	9	9
■ Satisfactory				2
■ Unsatisfactory				
Grand Total	3	11	11	14
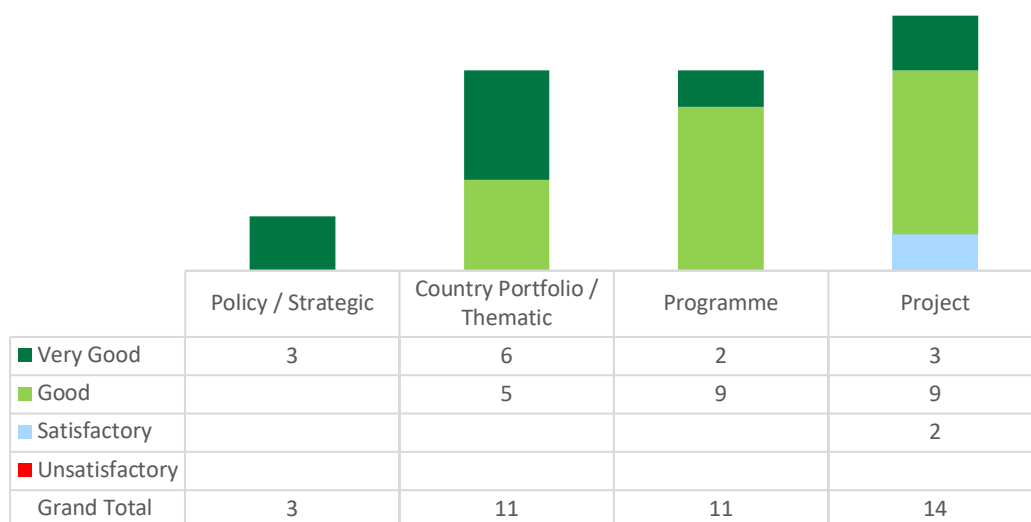
Evaluations cover multiple impact areas under the Strategic Plan 2014-2017. As with previous years, the most evaluated impact area was Women's Economic Empowerment (included in 19 evaluations); although the same number of evaluations also covered Ending Violence Against Women in 2017. The least evaluated area was the Normative Framework (included in 7 evaluations); while the percent of evaluations covering Women Peace and Security increased compared to previous years.

Figure 6: Coverage of UN-Women Strategic Plan 2014-2017 impact areas in 2017 evaluations
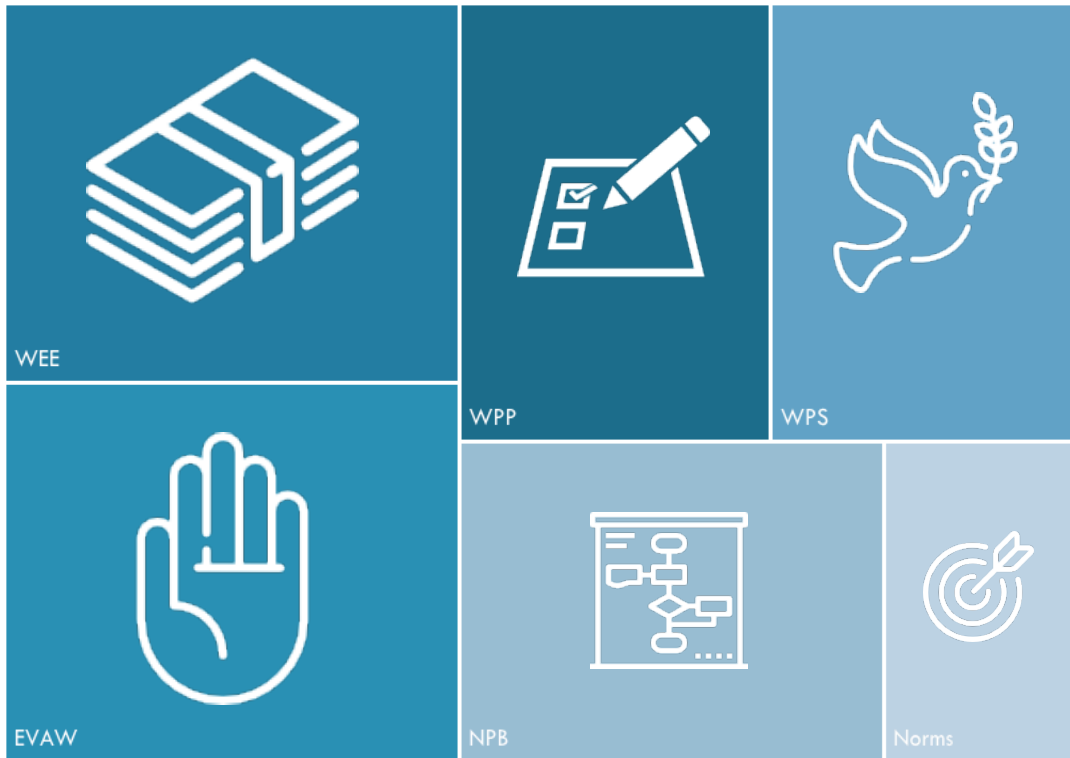
Within the regional architecture, decentralised evaluations have, again, increased in number to reach 34 in 2017 (6 more than in 2016). Of the six UN Women regions, five produced evaluations in 2017, with none submitted from Latin America and the Caribbean Region (although two countries, Mexico and Brazil were included in global and multi-country evaluations). At the global level, there was only one corporate evaluation; with the remaining 4 global evaluations being commissioned by Policy and Programme Division.

Figure 7: Distribution of business units commissioning UN-Women evaluations 2017

Arab States region produced 4 evaluations rated Very Good, and Europe and Central Asia regions produced 3 rated Very Good. The Independent Evaluation Office, Europe and Central Asia, and HQ divisions achieved the highest percent of Very Good rated evaluation reports. As with previous years, Eastern and Southern Africa produced the most evaluations; while Arab States significantly increased the production of evaluations compared to previous years. For the first time, all evaluations from West and Central Africa were rated as meeting UN-Women standards; a significant achievement.

Figure 8: Regional disaggregation of overall GERAAS quality ratings for 2017 evaluations

	Arab States	Asia and the Pacific	Eastern and Southern Africa	Europe and Central Asia	Western and Central Africa	HQ	IEO
■ Very Good	4	1	2	3	1	2	1
■ Good	5	4	6	1	5	2	
■ Satisfactory			2				
■ Unsatisfactory							
Grand Total	9	5	10	4	6	4	1
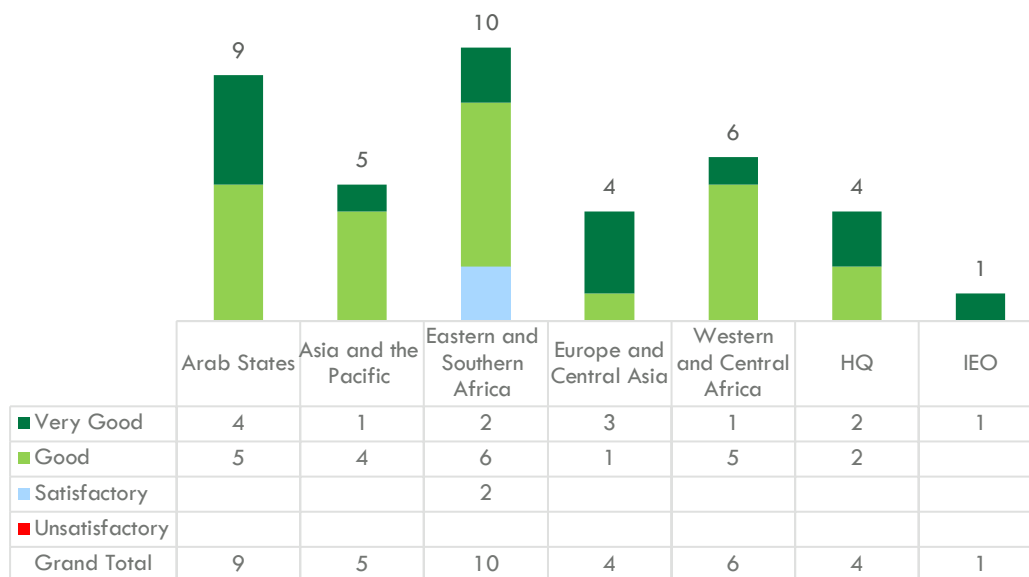
Finding 4: Evaluations continue to improve in terms of meeting UN-SWAP standards for integrating gender equality and human rights

The 2017 meta evaluation found a significant increase in performance of evaluation reports with regard to the UN-SWAP Evaluation Performance Indicator (EPI), with the highest recorded performance for UN Women since tracking of this indicator began.

Figure 9: UN-SWAP Evaluation Performance Indicator for 2017
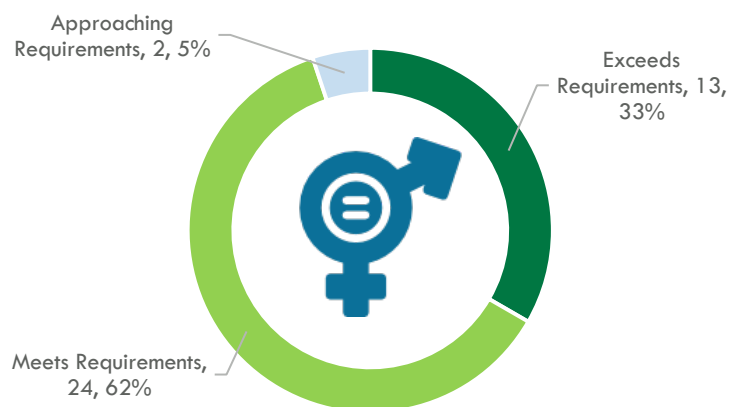
Overall, UN Women evaluations were found to be meeting requirements as defined by the UN Evaluation Group, with an average performance rating of 10.08 (out of a maximum score

of 12). This is a substantive improvement over the 2016 average of 8.31, 2015 average of 6.6 and the 2014 average of 7.5. This means that UN Women has maintained compliance with the UN-SWAP standards. All UN-SWAP criteria now, on average, meet requirements. This reflects, in particular, a improved performance in terms of gender-response evaluation methods.

Table 1: UN-SWAP average scores for 2017 and 2016 evaluation reports

	1. Scope	2. Questions	3. Methods	4. Analysis	Overall
2017	2.21 / 3	2.67 / 3	2.49 / 3	2.72 / 3	10.08 / 12
2016	2.17 / 3	2.25 / 3	1.69 / 3	2.19 / 3	8.31 / 12
Change	+0.04	+0.42	+0.80	+0.53	+1.77

13 evaluation reports were rated as exceeding the UN-SWAP standards. These included six Country Programme Evaluations (Côte d'Ivoire, DRC, Kyrgyzstan, Malawi, State of Palestine, and Sudan), and one Country Office Strategic Note Mid-Term Evaluation (Kenya); two regional evaluations (Evaluation of Humanitarian Action in Arab States, and Evaluation of Gender-Responsive Budgeting in Europe and Central Asia); the Corporate Evaluation on UN-Women Strategic Partnerships; and three country-level evaluations (in Jordan, Egypt and Papua New Guinea). Examples of good practices in reports include:

Corporate Evaluation on Strategic Partnerships for Gender Equality and the Empowerment of Women

Gender equality is included in the objectives and the scope of the evaluation; and in the discussion defining the outcome mapping descriptions (evaluation rubric).

Women's human rights are included both as questions under the standard OECD DAC criteria and as a standalone criterion.

The evaluation used multiple lines and levels of evidence using participatory techniques in both data collection (including social learning) and analysis (including outcome mapping).

Human Rights frameworks were applied to the stakeholder analysis. Section 4.4 of findings deals expressly with gender equality and human rights.

The conclusions and recommendations include a discussion of power relations.

Country Portfolio Evaluation Palestine
Gender equality is integrated into the evaluation objectives and scope from both a normative and programmatic perspective.

The evaluation includes a comprehensive matrix of indicators that also capture both instrumental and transformative contributions.

The evaluation includes a standalone criterion on women's human rights and mainstreams gender into all of the evaluation criteria through dedicated questions.

The evaluation uses participatory processes to include the voice of different stakeholders. Contribution analysis is relevant to identify gender dimensions.

Analysis addresses gender in terms of both structural and social-norms; at the institutional and the normative levels.

Evaluation of Equality for Progress (E4P) and Planim Save Kamap Strongpela (Plant Knowledge, Grow Strong), Papua New Guinea
Gender equality is included explicitly within the objectives of the evaluation.

Evaluation questions under relevance and effectiveness explicitly address gender equality.

The qualitative design of the evaluation triangulated multiple lines and levels of evidence. A validation and conclusions workshop ensured the participation of stakeholders in data analysis as well as in data collection.

The findings, conclusions and the recommendations address both the project process and the outcomes for women. While this is mostly restricted to social norms at the community level (rather than structural drivers), this is a result of the level of the evaluation rather than a gap.

Evaluation of Humanitarian Action in Arab States
Gender equality is clearly included in the scope, objectives and the indicators for the evaluation (especially around effectiveness).

The evaluation includes a specific section on women's human rights as well as mainstreaming in questions throughout the evaluation matrix.

The evaluation applied a predominantly qualitative design, based on participatory data collection and surveys.

The stakeholder analysis included identification of human rights roles, and sampling considered gender groups.

Case studies explicitly created opportunities for participation in data analysis and validation of findings.

The findings, conclusions and recommendations all explicitly address gender-equality in structural terms as well as instrumental.

Finding 5: There is increased explanation of overall evaluation designs and approaches in reports, accompanied by emerging examples of innovation

Previous meta evaluations have found a high level of homogeneity in UN-Women evaluation designs (with a high prevalence of qualitative approaches). While this remains the case for project evaluations, the 2017 meta evaluation found many examples of innovation in terms of extending the range and application of evaluation approaches. These individual examples not only represent interesting practices for futures evaluations to consider, but are also starting to contribute to a more diverse overall portfolio of evaluative evidence. The quality assessments noted innovations emerging in five main aspects of evaluation design:

Analytical Frameworks: Beyond the evaluation framework, a few evaluations started to articulate (or even develop) analytical frameworks that are specific to the intervention being evaluated. The clearest example of this was the development and use of a Capacity Development Process Framework for the evaluative analysis in the Eastern and Southern Africa Regional Office Evaluation of capacity development initiatives during the Strategic Note period 2014-17[9]. Another example was from Egypt, where a thematic evaluation[10] drew on elements of the Inclusive Systemic Evaluation (ISE) approach, and used a 'Gender Results Effectiveness Scale' (GRES) to enable deeper analysis of gender changes[11].

Performance assessment: A challenge for some evaluations is to address gaps in the monitoring and results data collected by projects and programmes. Several evaluations used alternative approaches to mitigate this challenge. In China, for example, a hypothesis-based approach was used to mitigate the absence of programme indicators[12]. Alternatively, the use of rubric establishes clear parameters for assessing the performance of an initiative in advance of collecting evaluation evidence. Several examples of rubrics were observed, including a global programme evaluation of ending violence against women and girls[13] (which used rubric for rating each criterion), and the use of outcome mapping by the corporate evaluation of strategic partnerships. The Country Portfolio Evaluation of Malawi demonstrated that these approaches can be simplified and used at the country-level, by – for example – using a checklist assessment to systematically examine Results Based Management performance.

Sampling: While transparency around sampling approaches and sample frames is an area that requires strengthening overall (see later analysis), several evaluations demonstrated well justified or innovative approaches to sampling. In terms of meeting UN standards, the Country Portfolio Evaluation for DRC is an example of clearly justifying sampling and the level of stakeholder participation through a reference group process. An example of good practice in sampling was noted in Kenya[14], with the design of a rapid assessment combining elements of probabilistic, stratified sampling and non-probabilistic sampling: key informant interviews were lined up with a weighted, purposive selection of those partner agencies seen as the most relevant partners. An example of

---

[9] Evaluation of UN Women ESARO Capacity Development Initiatives during the Strategic Note period 2014-17
[10] Thematic Evaluation of UN Women's Contribution to Women's Leadership and Political Participation in Egypt
[11] The evaluation classified data on results against 3 scales from the 5 categories found in the GRES scale, namely: gender targeted, gender responsive and gender transformative.
[12] Final Evaluation of the 2nd Phase EVAW Programme: Promote Efforts for National Legislation on Domestic Violence and Upscale the Multi-sector Model in China
[13] Preventing and addressing violence against women and girls in Albania, Mexico and Timor Leste
[14] UN Women Kenya Country Office Strategic Note Mid-Term Evaluation

innovation was found in a regional programme evaluation in Arab States[15], which given the varying backgrounds of project participants (with respect to nationality, marital status, age and the components they participated in) used a maximum variation approach to ensure inclusiveness.

Qualitative analysis and process: While all evaluations used qualitative methods, many of these were a combination of key informant interviews, group interviews (labeled as focus group discussions), and document analysis. A few evaluations, however, expanded the range of techniques. Two examples include the use of a World Café process for both data collection and analysis in Ethiopia[16], and the use of Q-methodology to investigate different perspectives of participants on an issue by ranking and sorting a series of statements (also known as Q-sort) in Kyrgyzstan[17].

Quantitative analysis: Overall, few UN-Women evaluations make extensive use of quantitative data and analysis techniques. However, there were some examples of analysing 'quantified' data (i.e. transforming qualitative data into numbers) to expand the range of evaluation methods. In Kenya[18], for example, in order to investigate patterns in numeric and categorical data, SPSS v20 functionalities were used, including cross tabulation, exploratory techniques, frequency tables, measures of central tendency (especially mode) and summary statistics. In the global evaluation of the flagship Progress of the World's Women report[19], the evaluation disaggregates and evaluates the staff time for each position that contributed to Progress. At the regional-level, in Eastern Europe and Central Asia a thematic evaluation[20] attempted to apply an version of Social Return on Investment (SRoI) adapted to time-based accounting.

---

[15] Economic Empowerment of Syrian Women Refugees and Host Communities in the Arab Region – Final Evaluation of UN Women Regional Project (2014- 2016)
[16] Evaluation of Phase II of the Joint Programme On Gender Equality And Women's Empowerment In Ethiopia
[17] Kyrgyzstan Country Programme Evaluation
[18] Report of Final Evaluation Integrating Gender in Peace Support Operations
[19] PROGRESS OF THE WORLD'S WOMEN: Evaluation of UN Women's Flagship Report
[20] Evaluation of UN Women's Contribution to Gender-Responsive Budgeting in the Europe and Central Asia Region

6.2 TRENDS BY QUALITY ASSESSMENT PARAMETERS

Across all parameters that are assessed by GERAAS, the average quality of evaluations improved in 2017; with more evaluations meeting UN-Women standards across more parameters than previous years. Findings were indicated as particularly strong in 2017. The only area with substantive remaining gaps is 'Lessons Learned' (which is combined with 'Conclusions' in the assessment template).

Figure 10: Disaggregation by GERAAS quality parameter of ratings for 2017 evaluations

	Background	Purpose	Methods	Findings	Conclusions	Recommendations	UN-SWAP	Presentation
■ Very Good	36%	46%	23%	54%	38%	28%	33%	46%
■ Good	49%	46%	64%	44%	51%	69%	62%	49%
■ Satisfactory	15%	8%	13%	3%	5%	3%	5%	5%
■ Unsatisfactory	0%	0%	0%	0%	5%	0%	0%	0%
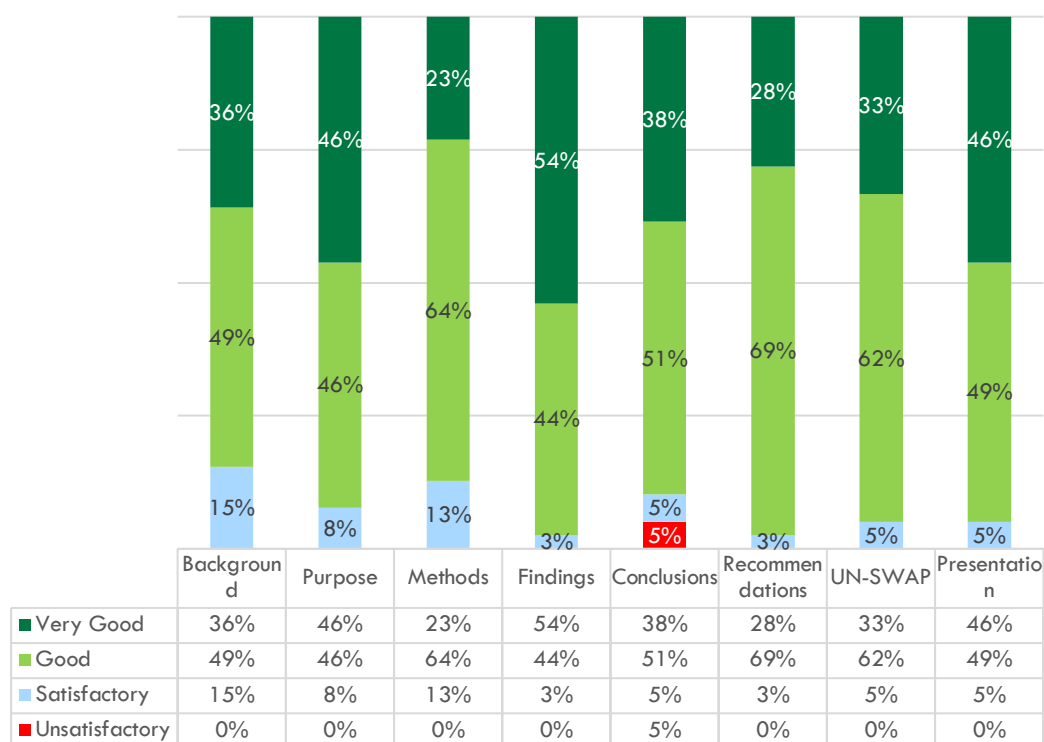
Parameter 1: Object and Context of the Evaluation

The 2017 portfolio of evaluations maintains, and builds on, the patterns of quality observed in 2016. For each indicator, over half of reports fully meet UN-Women standards. The strongest aspect of reports is providing information on the current implementation status of the intervention or approach that is being evaluated.

In 58% of reports, standards for stakeholder analysis are fully met; with the best reports undertaking a human rights based analysis of the stakeholders, and their main roles in the system being evaluated. Where evaluation reports needed strengthening in this regard, GERAAS assessments tended to recommend that evaluation managers review the Better Evaluation resource page on understanding and engaging stakeholders[21].

The percentage of reports meeting UN-Women standards for presenting the logic model of the evaluation object improved compared to 2016, but 14% of reports still require substantial improvement in this regard. The reviews noted some emerging weaknesses in the context analysis of evaluations; with some reports only providing information on the 'internal' UN-Women programming context, or national policy context; whereas the best reports also identify the key political, social, economic, cultural and institutional contexts.

---

[21] http://www.betterevaluation.org/en/plan/manage/identify_engage_users

Examples of good practices in reports include:
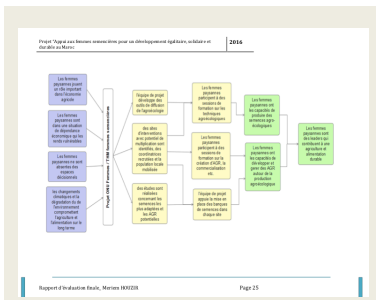
Évaluation du projet "appui aux femmes semencières pour un développement égalitaire, solidaire et durable au Maroc"

Evaluation of UN Women's Contribution to Gender-Responsive Budgeting in the Europe and Central Asia Region

Sudan Country Portfolio Evaluation Strategic Note 2014 - 2016

The report presents a complete description of the theory of change that clearly explains the inputs, outputs, outcomes as well as the expected impact of the project.

The context of women farmers in Morocco is also discussed in detail, as is the situation of gender and human rights.
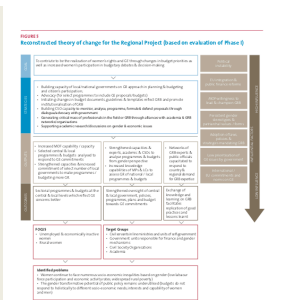
The report includes specific information about the initiative's scale, budget, objectives, and theory of change. It describes the international and national policy and socio-economic environments around gender-responsive budgeting.

Stakeholders and their contributions (including rights holders and duty-bearers) are explicitly identified within a dedicated sub-section.

The report describes the country programme evolution, financial status, geographic reach, results targets, programming areas, and key initiatives and partnerships.

The report includes information on national political frameworks, the current status of women's rights, and UN Women's national positioning.

Parameter 2: Purpose, Objectives and Scope

Consistent with previous meta evaluations, it was found that most reports take the statements of purpose, objectives and scope directly from the terms of reference. The objectives were found to be highly consistent across the entire portfolio of evaluations; reflecting both the UN-Women guidance on management gender responsive evaluations and the contributions of Regional Evaluation Specialists to developing terms of reference.

Most reports included a full evaluation matrix in the annexes. However, not all reports explicitly defined the evaluation criteria. While the criteria could often be derived from the evaluation objectives or questions, explicitly naming and defining them is necessary to be fully compliant with UN Evaluation Group standards.

The assessments indicate that 29% of reports would have benefited from more detailed explanation of the evaluation scope – clearly defining the boundaries of the evaluation, including naming and justifying any excluded issues or areas. While many evaluation reports

elaborated indicators for each evaluation question[22], few established rubric (pre-determining levels of expected performance to ensure transparent assessment of evidence). Several resources on the use of rubric are included on the Better Evaluation platform[23].

A number of evaluations were found to explicitly include human rights and gender equality as clear objectives for the evaluation; and to explore the analytical framework used to achieve this. These represent a good practice that can guide other evaluations in the future.

Examples of good practices in reports include:

Country Portfolio Evaluation Malawi	***Country Portfolio Evaluation Palestine***	***Strengthening women's capacity in disaster risk reduction to cope with climate change in Viet Nam***
The purpose is covered in several places, and includes explicit analysis of the primary intended users and uses.	*Evaluation criteria are defined and justified, and evaluation questions are prioritised.*	*The report is particularly strong at identifying both primary and secondary evaluation users.* *The evaluation uses standard OECD/DAC evaluation criteria and lists the evaluation questions.*

Parameter 3: Methodology

Evaluation reports are continuing to strengthen in terms of describing the methodological approach. The quality assessments noted that many more reports are now defining the overall design and approach of the evaluation, as well as the individual data collection and analysis methods. Assessments of evaluations that did not elaborate a coherent design used the Better Evaluation list of approaches[24] to propose options for future evaluations of a similar nature.

Particular strong points include specifying appropriate data collection methods, and justifying data sources. 55% of reports would benefit from having more detail on the tools used to collect, store, and analyse data to ensure the preservation of quality. There is a tendency for only evaluations that include a strong element of quantitative methods to explicitly name the software tools used to collate and analyse data.

---

[22] The meta evaluation did note that the UN-Women Country Programme Evaluation templates and guidance needs to be updated to explicitly include indicators
[23] http://www.betterevaluation.org/en/blog/introducing_rubrics
[24] http://www.betterevaluation.org/en/approaches

Levels of stakeholder participation, and the application of methods appropriate to identifying human rights and gender equality issues, have become areas of strength in 2017 evaluation reports. This is also reflected in the improved performance of UN-SWAP evaluation criteria.

While 39% of reports fully meet UN-Women standards for ethics, 16% of reports did not reference ethics (similar proportions to 2016). Therefore, this continues to be an issue that warrants special attention for future evaluations.

Another indicator with similar performance characteristics to ethics, is the requirement to present a sampling frame and appropriate sampling strategy. While 34% of reports fully met this standard, most evaluations would benefit from greater transparency around sampling; this is especially the case for qualitative evaluations that rely on convenience sampling, which tend not to describe the sampling universe or justify the chosen sampling strategy.

Examples of good practices in reports include:

Évaluation du portefeuille pays - République Démocratique du Congo (2014-2017)

The methodology used is highly participatory and the report describes how the methods used seek to specifically answer evaluative questions.

The sampling strategies used are duly justified as well as the level of stakeholder participation through a reference group process covering the entire scope of the evaluation.

The evaluation incorporates ethical considerations into its methodology as well as gender and equity issues.

Kenya Country Office Strategic Note Mid-Term Evaluation

Like many reports, the evaluation uses a qualitative approach; but it fully acknowledges the limitations of this and explains the justification. Useful levels of detail are provided for each of the elements of the method.

Progress of the World's Women: Evaluation of UN Women's Flagship Report

The sequencing of multiple lines and levels of evidence are matched to sampling for each method that is intentional and fully justified.

The use of peer reviews of the flagship reports is an interesting and innovative method.

Parameter 4: Findings

The findings sections of reports were found to be stronger in 2017 than in previous years. In particular, 92% of reports fully addressed all of the evaluation criteria specified in the evaluation objectives; with a lower occurrence of evaluations inappropriately claiming to be able to assess impact when the design and level-of-effort precluded such analysis. This indicates better planning of evaluations, and more discipline in setting criteria and objectives.

The majority of reports were fully compliant with UN Evaluation Group standards for undertaking systematic analysis (68%), based on objective use of all the available evidence (79%), to explain the reasons that led to indicated levels of performance (84%). No evaluation completely failed to address, at least partly, any of the GERAAS indicators relating to findings.

Where there is room for growth is in relation to ensuring full transparency in relation to the level of evidence, and any gaps, in regard to each evaluation finding. The best reports fully explained the specific evidence that contributed to each finding, including any refuting or missing data; however, 21% of reports only partly achieved this – exhibiting a tendency to present only the final analysis, without fully explaining what lay underneath this.

Examples of good practices in reports include:

Evaluation finale du projet "Mainstreaming gender into humanitarian response and protection of women and girls affected by Boko Haram terrorism in the far north of Cameroon"	*Final Evaluation of the 2nd Phase EVAW Programme: Promote Efforts for National Legislation on Domestic Violence and Upscale the Multi-sector Model in China*	*Evaluation of Humanitarian Action in Arab States*
The evaluation comprehensively presents both good results as well as the aspects that presented challenges. The results are presented according to the evaluation criteria.	The findings are arranged in a novel way, through sections that are most useful to users. They also consolidate the key findings and evidence at the end of the report around the criteria and the hypotheses. This approach strikes a good balance between usability and systematic use of the evaluation framework.	Findings systematically address the evaluation questions, and marshal multiple sources of evidence to do so. The report is notable for carefully defining what level of data constitutes 'evidence'.

Parameter 5: Conclusions and Lessons Learned

In general, conclusions remain a strong parameter of evaluation reports. 71% of reports fully met all aspects of the UN-Women standards for conclusions, with only 5% of reports (i.e. 2 evaluations) failing to deepen the analysis of the findings or provide answers to the evaluation questions. Opportunities to further strengthen the development of conclusions in 29% of reports include being sure to identify potential solutions to the weakness, challenges, and opportunities identified throughout the evaluation.

The elaboration of generalised lessons learned was found to be dichotomous: with reports either fully meeting UN-Women standards, or not including lessons learned (even when 'lessons' were identified as an objective of the evaluation). While the percent of reports included lessons is similar to 2016, where lessons have been included these were found to be consistently better than in previous years. This is an important achievement. Nevertheless, for at least 32% of evaluations, the GERAAS assessments recommended that evaluation managers consider using, in future, the ILO-published templates for lessons learned and good practices[25].

The meta evaluation notes that overall performance of the 'conclusions and lessons learned' parameter is heavily influenced by the presence, or absence, of lessons learned: indicating that future revisions to GERAAS should consider separating these elements of the UN Evaluation Group standards.

Examples of good practices in reports include:

Evaluation of Phase II of the Joint Programme on Gender Equality and Women's Empowerment in Ethiopia	*Final Evaluation of the Global Programme: Women's Economic Empowerment and Integration into the Value Chain of the Coca-Cola Company in Brazil, Egypt and South Africa*	*Corporate Evaluation on Strategic Partnerships for Gender Equality and the Empowerment of Women*
The conclusions develop strong and balanced insights based directly on the evidence from findings. These provide a useful steer for users on the direction of the JP, which is directly linked to the purpose of the evaluation.	*The conclusions address each of the evaluation criteria, and are clearly derived from the relevant findings sections.* *A wide set of common and country-specific lessons are provided.*	*The report provides eight (8) concrete conclusions that provide key insights and identify solutions to important problem areas.* *The lessons learned can be used to inform similar initiatives in other contexts.*

---

25 http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_165981.pdf

Parameter 6: Recommendations

Similarly to 2016, evaluation reports for 2017 were found to consistently present relevant and evidence-based recommendations that are actionable. This is a key factor for the utility of evaluations. In terms of the way recommendations are written, there were some good examples of presenting recommendations so that the specific targeting and prioritisation of actions is clear; and these can serve as a model for future reports (see below).

Once again few evaluation reports explicitly describe the process of how recommendations were developed (only 34%, which is the same proportion as 2016); although for a further 26% of reports this process could be deduced from the methods section. Transparency on who was involved in developing and validating the recommendations is a requirement of the UN-Women standards; and many GERAAS assessments indicated that future reports should refer to the role of evaluation reference groups, or other participatory processes in ensuring the relevance and utility of recommendations.

Examples of good practices in reports include:

External Evaluation of the Project "Enhancing women's political representation through improved capacity and enhanced support in Moldova"	*Country Portfolio Evaluation Cameroon.*	*Kyrgyzstan Country Programme Evaluation*
The format of a table ensure that the recommendations, which are targeted and prioritised, are clearly linked to the findings from which they are derived.	*The recommendations are clearly derived from the findings, and are all marked according to urgency, difficulty, and potential impact.*	*Recommendations are grounded in the conclusions; are rated according to (1) Urgency, (2) Impact, and (3) Difficulty; and provide proposals for how they can be implemented.*

Parameter 8: The Report Structure

All evaluation reports were found to be logically structured and well written, with 80% fully meeting UN-Women standards for layout and presentation. Executive summaries were also an overall strength — which is important for usability — with 71% fully meeting UN-Women standards; but are the area in which further improvement is likely to bring the greatest benefits[26]. In a similar observation to 2016, 39% of reports were found to require additional material in the annexes; although 61% of reports fully met this standard. Several reports had been formatted in the UN-Women 'house-style'; and these were all found to be clear, accessible, and fully compliant with required standards. Examples of innovative presentation include:

---

[26] 3 reports completely failed to meet the expected standard

Thematic Evaluation of UN Women's Contribution to Women's Leadership and Political Participation in Egypt

For such a comprehensive report, this is surprisingly concise – a feature which supports usability.

Malawi Gender Based Governance Programme Midterm Evaluation

The use of 'opportunities' and 'challenges' to synthesise issues throughout the report is novel and helpful.

Moldova Country Programme Evaluation

The use of visualisations to illustrate both data/evidence, and conceptual ideas
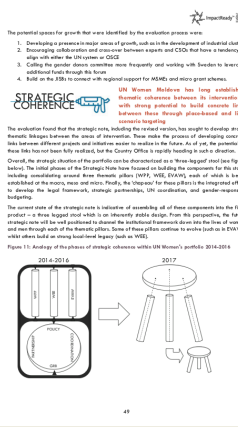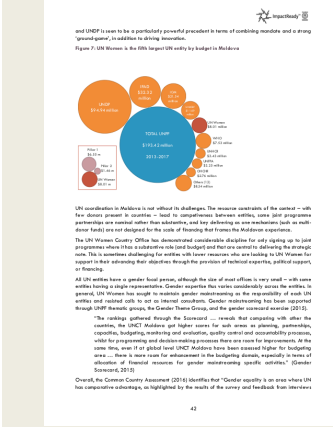
7. CONCLUSIONS

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

Conclusion 1: The quality of evaluation reports in 2017 continues to deepen; with more reports more consistently and more fully meeting UN-Women standards across more parameters.

The most important 'boundary' between report ratings is between 'Good' and 'Satisfactory' classifications, since this denotes whether reports sufficiently meet UN Evaluation Group (UNEG) standards, or whether they are 'approaching' these standards. The percentage of reports that are rated as Good or above has steadily increased from 55% in 2013, to 95% in 2017. Based on this pattern, UN-Women is on track to have 100% of evaluation reports meeting UNEG standards in the next one or two cycles, if it maintains the current level of annual improvements.

The pattern between reports rating as 'Good' or 'Very Good' (i.e. exceptional across most parameters) demonstrates less consistent pattern. However, the introduction of automatically aggregated scoring should help to normalise this ratio for future cycles. During the current cycle, a high-requirement was set for reports to be classified as Very Good (fully meeting all indicators across the majority of parameters). This resulted in a similar proportion of reports being rated Very Good as in 2016, which suggests that it is appropriate and should be retained. At the other end of the scale, no report was rated as Unsatisfactory for the fourth year in a row.

All of these patterns – the increasing overall quality, the high standard of Very Good evaluations, and the avoidance of unsatisfactory reports – indicate that the evaluation capacity established by UN-Women is being effective, and its maintenance is required to ensure continued improvement.

Conclusion 2: Good quality evaluation evidence is available at all levels and across all strategic impact areas; but some gaps in geographical coverage of programme presence remain.

During the Strategic Plan 2014-2017, most countries with UN-Women office presence or programme presence were directly covered by at least one evaluation. However, 11 programme countries (13%) were not directly covered by evaluation, and this represents an important gap. In addition, for 2017, no evaluations were commissioned within Latin America and the Caribbean Region; which is the first time this has been observed. The Strategic Plan 2018-2021 represents an opportunity to address and prevent such gaps in the future.

Nevertheless, where evaluative evidence is available, which is in most countries and at all levels (global, regional, national, and subnational), it is of increasingly high levels of quality – meaning that it is reliable and useful for decision-makers. There is also broad coverage across all of the strategic impact areas. This has been helped by a steady shift from project evaluations, to more comprehensive programme, thematic, and country-portfolio evaluations.

Conclusion 3: The while the range of evaluation designs remain mostly qualitative, there are more examples of innovative approaches and methods being used.

A continued observation of meta evaluations 2013-2016 was of the limited range of designs evident in UN-Women evaluation reports; which has important implications for the type and level of evidence available to the entity. The findings for 2017 indicate an important break from this pattern. While most reports are still predominantly qualitative, two important patterns were identified: (1) reports are more explicitly and transparently discussing the evaluation design (instead of just listing methods), and (2) there are a noticeable number of evaluation reports based on innovative and creative evaluation approaches. These represent useful examples for future evaluations to learn from, and to be inspired by.

Conclusion 4: UN-Women evaluations continue to meet the requirements of the UN-SWAP evaluation performance indicator, and are demonstrating improvements across all criteria.

Strengthening the gender responsiveness of UN-Women evaluations has been a consistent recommendation of previous meta evaluations; however, the performance of evaluation reports in terms of the four UN-SWAP criteria now consistently meets the required standards, and has steadily improved over the last three years. This represents an important achievement; much of which is connected with improvements in the methods and analysis observed in reports during this period.

The easiest way to further improve on UN-SWAP performance – and to move closer towards an overall rating of 'Exceeds Requirements' – is to ensure that all reports include an evaluation matrix with indicators that explicitly relate to gender equality and women's empowerment. Updating the current Country Portfolio Evaluation guidance and templates to include 'model' indicators is one activity that could assist in achieving this.

Conclusion 5: The use of a weighted quality assessment tool has improved consistency of report ratings; but the GERAAS tool needs more comprehensive revision to still be relevant in the future.

The GERAAS assessment tool was updated for 2017 to include weighted parameters and automatic calculation of parameter and overall ratings. This has been successful in terms of consistency; but other important considerations have emerged over the duration of GERAAS that warrant a more fundamental revision. These include:

1. Ensuring that GERAAS fully aligns with the revised UNEG Norms and Standards (2016)
2. Ensuring that GERAAS fully aligns with the UN-Women Strategic Plan 2018-2021
3. Agreeing the weighting of each parameter with the Regional Evaluation Specialists and corporate evaluation managers
4. Reducing the overlap of some indicators
5. Disaggregating in overall ratings the analysis of parameters influenced by the UN-Women evaluation function (such as purpose and scope), from the parameters influenced by independent evaluators (such as findings, conclusions and recommendations)
6. Including lessons learned in a separate parameter to conclusions.

8. RECOMMENDATIONS

The following recommendations were developed based on the conclusions, and validated in discussion with UN Women Independent Evaluation Office.

Recommendation 1: The Independent Evaluation Office is advised to update the GERAAS evaluation quality assessment template/tool to align it with the UN-Women Strategic Plan 2018-2021, and UNEG Norms and Standards 2016.

GERAAS has used the same template and assessment tool throughout the Strategic Plan 2014-2017 cycle; which has provided consistency, but also shows signs of needing to be updated. The Strategic Plan 2018-2021 cycle is an opportunity to do this. This part of the recommendation can be implemented by:

1. Revising the categorisation of evaluation coverage to align with the Strategic Plan 2018-2021 outcomes.
2. Consolidating and revising existing assessment indicators to reduce overlap; and to align with the UN Evaluation Group Norms and Standards 2016, and guidance on Integrating Human Rights and Gender Equality in Evaluation.
3. Assessing lessons learned separately from other parameters.
4. Differentiate assessment of OECD-DAC evaluation criteria under findings (i.e. separate indicators to assess the quality of efficiency, effectiveness, relevance, sustainability and impact.
5. Review the tool to ensure that it is compatible with assessing the application of any guidance on evaluating humanitarian action that is recognised by UN-Women.

In the process of revising the GERAAS evaluation quality assessment, it is recommended that the following issues are addressed through a consultative process with regional evaluation specialists, corporate evaluation managers, and other stakeholders:

1. Agreeing the weighting of each parameter with the Regional Evaluation Specialists and corporate evaluation managers.
2. Disaggregating in the final rating of each report the performance of parameters influenced by the UN-Women evaluation function (such as purpose and scope), from the parameters mainly influenced by independent evaluators (such as findings, conclusions and recommendations).
3. Formally differentiating evaluation reports that have been fully quality assured through the involvement of regional evaluation specialists, from those produced directly by business units[27]; to better allow for the direct impact of the Independent Evaluation Office on quality to be assessed.

Recommendation 2: The Independent Evaluation Office is recommended to adjust the process and timing of GERAAS evaluation quality assessment to rolling near-time reviews.

To enhance the utility of the GERAAS quality assessment, it is optimal for management to have access to the ratings and feedback as soon as feasible after the report is published. For this reason, 'near-time' quality assessment has been initiated by several other member entities of the UN Evaluation Group in recent years.

---

[27] ILO calls these different types of evaluations 'independent evaluations' and 'internal evaluations'. See: http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/policy/wcms_603265.pdf

While this approach is sometimes referred to as 'real-time' reporting, contemporary experiences in other UN entities shows that 'near-time' (within 6 weeks) is more feasible. This is due to process factors including: (1) unknown pipelines of report finalisation, (2) timeliness of submission of the finalised reports by business units, (3) initial checks to assess correct classification as an evaluation and uploading to online systems, (4) submission for independent assessment, (5) independent assessment including quality assurance activities, (6) reporting.

To avoid the need for expensive 'stand-by' capacity to assess report quality in 'real time' (due to the uncertainties of demand), it is therefore recommended that UN-Women consider a transition to 'near-time' quality assessment for the Strategic Plan 2018-2021 cycle.

Recommendation 3: The Independent Evaluation Office is recommended to further develop evaluation guidance and report templates, especially for project evaluations.
Strategic evaluations (global, regional, corporate, and country-portfolio) consistently achieve the highest ratings for quality. These benefit from the most substantive bodies of guidance. It is recommended that the Independent Evaluation Office continues to develop and update this library of material, but with a stronger focus on project evaluations.

Project-level evaluations have the smallest teams, and rarely include evaluators with international experience. Based on the consistent quality of country portfolio evaluation reports (which benefit from guidance and templates[28]), it is recommended that a similar set of templates and guides are provided (in multiple languages) for country offices to provide local project evaluators. These should be geared to the low level of resources and time associated with many project evaluations. While the use of these proposed templates can remain optional; they are likely to become used by default if they make producing a high quality report easier.

Recommendation 4: Regional Evaluation Specialists are recommended to jointly develop a shared knowledge product explaining the current successful approach to quality assuring evaluation reports, to help sustain current positive trends in the future.
The meta evaluation reveals a strong positive trend across nearly all indicators since quality assessment began in 2013. Much of this improvement has been achieved at the decentralised level, and a large contribution has been the support provided by regional evaluation specialists. To ensure that this capability – and the positive trend – is maintained, it is strongly recommended that a means is found to systematise the knowledge and practices established by regional evaluation specialists. One option could be a joint knowledge project, capturing the experiences of the current cadre, to be shared with new evaluation specialists as part of the handover process.

Recommendation 5: UN-Women Senior Management, with the support of the Independent Evaluation Office, is recommended to ensure full coverage of all programme presence countries in independent evaluations during the course of the UN-Women Strategic Plan 2018-2021.
The meta evaluation highlights the strengthening resource of good quality evaluation evidence that is available to support UN-Women and its partners. Despite growth in the number of evaluations, and coverage of individual evaluations (e.g. more regional evaluations), 13% of programme countries were not directly covered during the Strategic Plan 2014-2017. It is recommended, therefore, that a consultative but centrally-led strategy is established early in

---

[28] The meta evaluation notes, however, that the CPE guidance requires updating with model indicators in the evaluation framework.

the current strategic plan cycle to ensure that all programme countries are directly included in at least one – but ideally several – evaluations. This coverage should be tracked as a key performance indicator in GERAAS.

Recommendation 5: UN-Women is recommended to link GERAAS to tracking of management response implementation rates as a proxy indicator for use.
While GERAAS provides a proxy indication of evaluation quality, it does not give insight into the utility – or use – of evaluations. Full assessments of 'use' require detailed audits of evaluation processes, and ex-post assessments of changes that have been contributed-to by an evaluation: both of which are time consuming and impractical at scale.

However, all evaluations have a management response, the implementation (and tracking) of which is established under UNEG standards as being the prerogative of 'management' (not the evaluation function). The acceptance or rejection of evaluation recommendations, and the completion (or not) of management response commitments, have both been used by other UN entities as proxy indicators for evaluation use.

While this is an imperfect solution to better understanding the use of evaluations, it does provide more insights than is currently available. It is thus recommended that an internal review of the management response system is undertaken, with the aim of assessing the feasibility and value of linking this to GERAAS.

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

---

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

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 <u>UN Women Global Accountability and Tracking of Evaluation System – the GATE system</u>. 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

Evaluation	Region	UN-SWAP	GERAAS
Thematic Evaluation of UN Women's Contribution to Women's Leadership and Political Participation in Egypt	Arab States	Exceeds Requirements	Very Good
MTE of Securing Rights and Improving Livelihoods of Women (SRILW) Programme	Arab States	Meets Requirements	Good
ÉVALUATION DU PROJET "APPUI AUX FEMMES SEMENCIÈRES POUR UN DÉVELOPPEMENT ÉGALITAIRE, SOLIDAIRE ET DURABLE AU MAROC"	Arab States	Meets Requirements	Good
Evaluation du degré d'intégration du genre dans Les Plans Communaux de Développement	Arab States	Meets Requirements	Good
Creating one-stop-shop for sustainable businesses final evaluation	Arab States	Meets Requirements	Good
Country Portfolio Evaluation Palestine	Arab States	Exceeds Requirements	Very Good
Evaluation of Humanitarian Action in Arab States	Arab States	Exceeds Requirements	Very Good
Final Evaluation of UN Women's Project "Rural Women's Food Security in Jordan"	Arab States	Exceeds Requirements	Very Good
Economic Empowerment of Syrian Women Refugees and Host Communities in the Arab Region – Final Evaluation of UN Women Regional Project (2014- 2016)	Arab States	Meets Requirements	Good
EVALUATION OF UN WOMEN AFGHANISTAN'S PORTFOLIO ON ENDING VIOLENCE AGAINST WOMEN (2014 -2016)	Asia and the Pacific	Meets Requirements	Good
Final Evaluation of the 2nd Phase EVAW Programme: Promote Efforts for National Legislation on Domestic Violence and Upscale the Multi-sector Model in China	Asia and the Pacific	Meets Requirements	Good
Evaluation of Equality for Progress (E4P) and Planim Save Kamap Strongpela (Plant Knowledge, Grow Strong)	Asia and the Pacific	Exceeds Requirements	Good
Final Report "Institutional Strengthening of Gender Equality and Women's Empowerment in Timor Leste" 2013-2016	Asia and the Pacific	Meets Requirements	Good
Strengthening women's capacity in disaster risk reduction to cope with climate change in Viet Nam	Asia and the Pacific	Meets Requirements	Very Good
Corporate Evaluation on Strategic Partnerships for Gender Equality and the Empowerment of Women	Corporate (HQ)	Exceeds Requirements	Very Good
PROGRESS OF THE WORLD'S WOMEN: Evaluation of UN Women's Flagship Report	Corporate (HQ)	Meets Requirements	Very Good
Final Evaluation of the Global Programme:	Corporate	Meets	Good

Women's Economic Empowerment and Integration into the Value Chain of the Coca-Cola Company in Brazil, Egypt and South Africa	(HQ)	Requirements	
Preventing and addressing violence against women and girls in Albania, Mexico and Timor Leste	Corporate (HQ)	Meets Requirements	Very Good
Promoting and Protecting Women Migrant Workers' Labour and Human Rights: Engaging with International, National Human Rights Mechanisms to Enhance Accountability	Corporate (HQ)	Meets Requirements	Good
Evaluation of Phase II of the JOINT PROGRAMME ON GENDER EQUALITY AND WOMEN'S EMPOWERMENT IN ETHIOPIA	Eastern and Southern Africa	Meets Requirements	Good
UN Women Kenya Country Office Strategic Note Mid-Term Evaluation	Eastern and Southern Africa	Exceeds Requirements	Good
ENGAGING WOMEN IN PREVENTING AND COUNTERING EXTREMIST VIOLENCE IN KENYA	Eastern and Southern Africa	Meets Requirements	Good
Report of Final Evaluation Integrating Gender in Peace Support Operations	Eastern and Southern Africa	Meets Requirements	Good
MALAWI GENDER BASED GOVERNANCE PROGRAMME MIDTERM EVALUATION	Eastern and Southern Africa	Meets Requirements	Good
Country Portfolio Evaluation Malawi	Eastern and Southern Africa	Exceeds Requirements	Very Good
Evaluation of UN Women ESARO Capacity Development Initiatives during the Strategic Note period 2014-17	Eastern and Southern Africa	Approaching Requirements	Good
Report on the Final Evaluation of the Project for the National Scale Up of the Isange One Stop Center Model in Rwanda	Eastern and Southern Africa	Meets Requirements	Satisfactory
Sudan Country Portfolio Evaluation Strategic Note 2014 - 2016	Eastern and Southern Africa	Exceeds Requirements	Very Good
Final Project Evaluation: Wanawake-Wanaweza Project on Women Leadership and Political Participation in Tanzania	Eastern and Southern Africa	Meets Requirements	Satisfactory
KRYGYZSTAN COUTRY OFFICE COUTRY PROGRAMME EVALUATION	Europe and Central Asia	Exceeds Requirements	Very Good
External Evaluation of the Project "Enhancing women's political representation through improved capacity and enhanced support in Moldova"	Europe and Central Asia	Meets Requirements	Very Good
Country Portfolio Evaluation of Moldova Strategic Note 2014 - 2017	Europe and Central Asia	Meets Requirements	Good

Evaluation of UN Women's Contribution to Gender-Responsive Budgeting in the Europe and Central Asia Region	Europe and Central Asia	Exceeds Requirements	Very Good
Evaluation finale du projet "Mainstreaming gender into humanitarian response and protection of women and girls affected by Boko Haram terrorism in the far north of Cameroon"	Western and Central Africa	Meets Requirements	Good
Country Portfolio Evaluation Cameroon.	Western and Central Africa	Meets Requirements	Good
Évaluation du Portefeuille-Pays Côte d'Ivoire	Western and Central Africa	Exceeds Requirements	Good
ÉVALUATION FINALE DU PROJET "PREVENTION ET ANTICIPATION DE LA VULNERABILITE DES FEMMES ET DES FILLES FACE AU TERRORISME A TRAVERS LEUR AUTONOMISATION, LE DIALOGUE COMMUNAUTAIRE ET L'EDUCATION DANS LE NORD DE LA COTE D'IVOIRE"	Western and Central Africa	Meets Requirements	Good
ÉVALUATION DU PORTEFEUILLE PAYS - RÉPUBLIQUE DÉMOCRATIQUE DU CONGO (2014-2017)	Western and Central Africa	Exceeds Requirements	Very Good
Projet "Action Humanitaire Sensible au Genre et Aide aux Femmes et Filles touchées par le Terrorisme de Boko Haram à Diffa - Niger"	Western and Central Africa	Approaching Requirements	Good

9.3 REPORT REVIEW FORMAT

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

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

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

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

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

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

PARAMETER 6: RECOMMENDATIONS			Guiding Question	Constructive feedback for future reports
GUIDING POINTS/EXPLANATORY NOTE	RATING	Remarks	**Are the recommendations relevant to the object and purpose of the evaluation and clearly presented in a priority order?**	Including how to address weaknesses and maintaining good practice
Recommendations				
6.1 Recommendations are supported by evidence and conclusions, and were				

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

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

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

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

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

9.4 GERAAS METHOD

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

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

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.

Meta-evaluation

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

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

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

PART 2: ASSESSMENT OF THE QUALITY AGAINST EACH OF THE EIGHT 'PARAMETERS'
The final review template is composed of *8 Parameters* (Object and Context of Evaluation; Evaluation Purpose, Objectives and Scope; Evaluation Methodology; Findings; Conclusions and

Lessons Learned; Recommendations; Gender and Human Rights Considerations; and the report structure). The Eight Parameters are further defined by 39 Guiding Points.

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

Figure 1: Meta-evaluation process

Filtering by EO

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

Filtering by Independent Reviewer

Reviewer checks that all submitted reports are evaluations	No reports removed

First Review

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

Second Review

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

Feedback

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

Meta-evaluation draft report

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

Comments

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

Meta-analysis draft report

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

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

Very good
Good
Satisfactory
Unsatisfactory

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

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

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

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

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

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

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

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

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

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

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

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

Changes Made From Previous Years
In response to conclusions from 2016, the GERAAS tool for 2017 introduced explicit weighting between parameters in order to ensure emphasis on the substantive aspects of evaluation quality (and to minimise subjectivity).

Table 2: Parameter weightings introduced for 2017

Parameter 1: Object and context	5%	Parameter 5: Conclusions and lessons	15%
Parameter 2: Purpose and scope	10%	Parameter 6: Recommendations	15%
Parameter 3: Methodology	15%	Parameter 7: UN-SWAP	15%
Parameter 4: Findings	20%	Parameter 8: Presentation	5%

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

UN entities are expected to meet UN SWAP performance standards by 2017, with an extended timeframe to 2019 for those entities with a mainly technical focus. The ultimate goal is that all UN system entities "meet requirements" related to the Evaluation Performance

Indicator. However, achieving this is only considered a starting point, and UN entities should continually strive to "exceed requirements" if the UN system is to truly benefit from gender responsive evaluation practice.

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

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

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

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

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

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

9.5 PARAMETER RATINGS