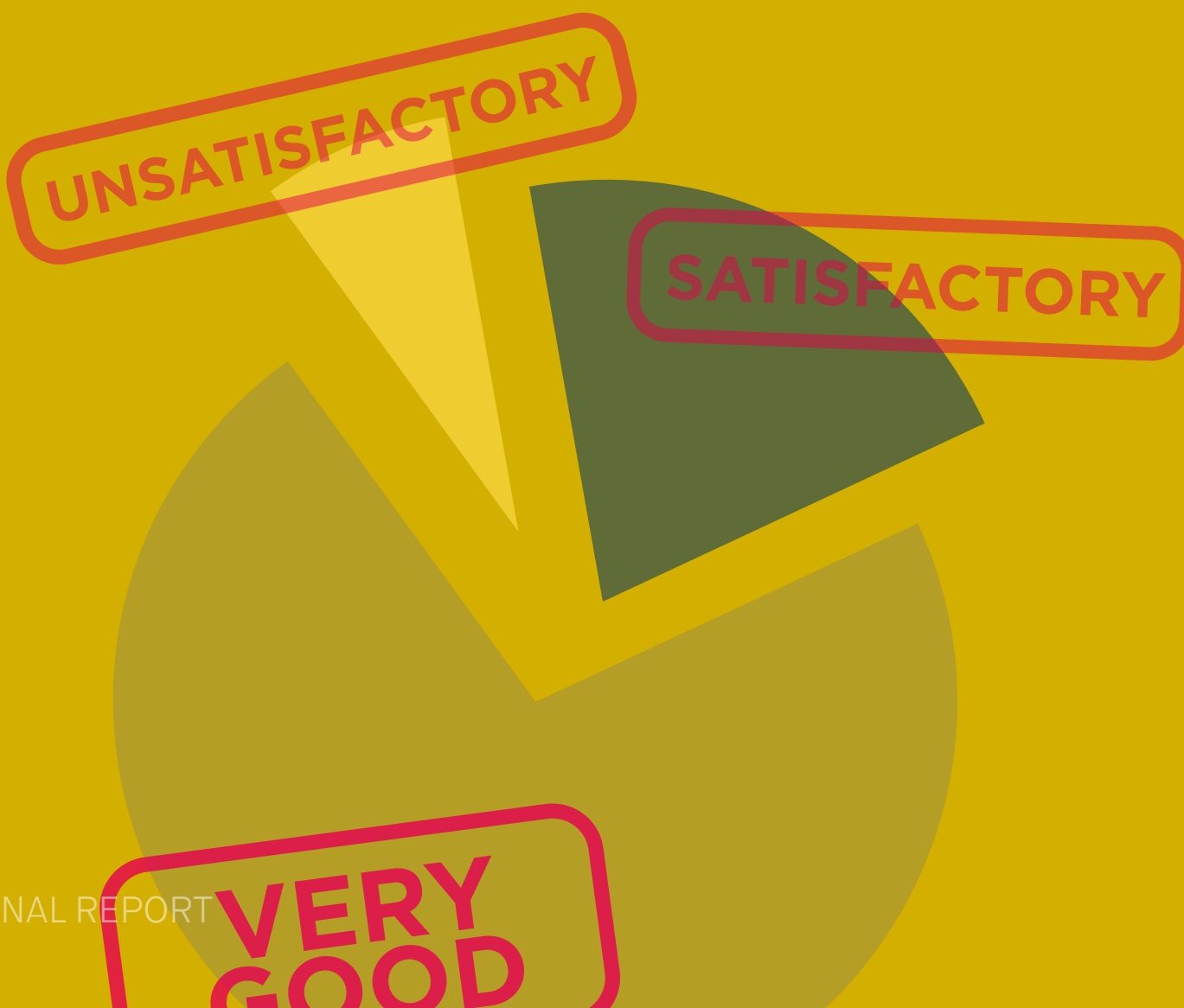
Meeting the **Mark**?

ASSESSING **THE QUALITY OF EVALUATIONS** CONDUCTED BY UN WOMEN IN ASIA AND THE PACIFIC

META-EVALUATION (2010-2016)



FINAL REPORT

Prepared for Yumiko Kanemitsu
Prepared by Jo-Anne Bishop
Designed by UN Women Regional Communications Unit, Pairach Homtong and Yoomi Jun
Infographics by Stephen Tierney

VERSION 2.0
14 MARCH 2017

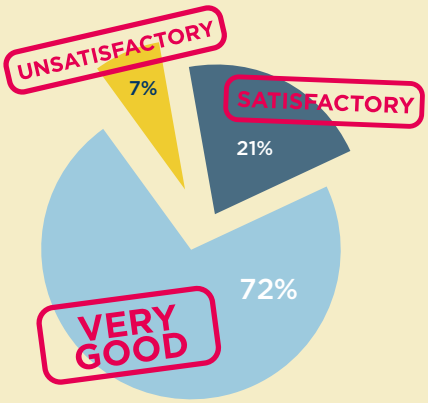
Meeting the **Mark**?

ASSESSING **THE QUALITY OF EVALUATIONS** CONDUCTED BY UN WOMEN IN ASIA AND THE PACIFIC

META-EVALUATION (2010-2016)

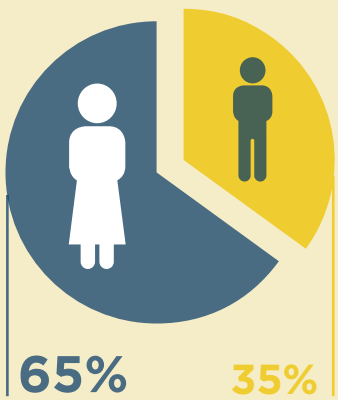
FINAL REPORT

KEY TRENDS AT A GLANCE



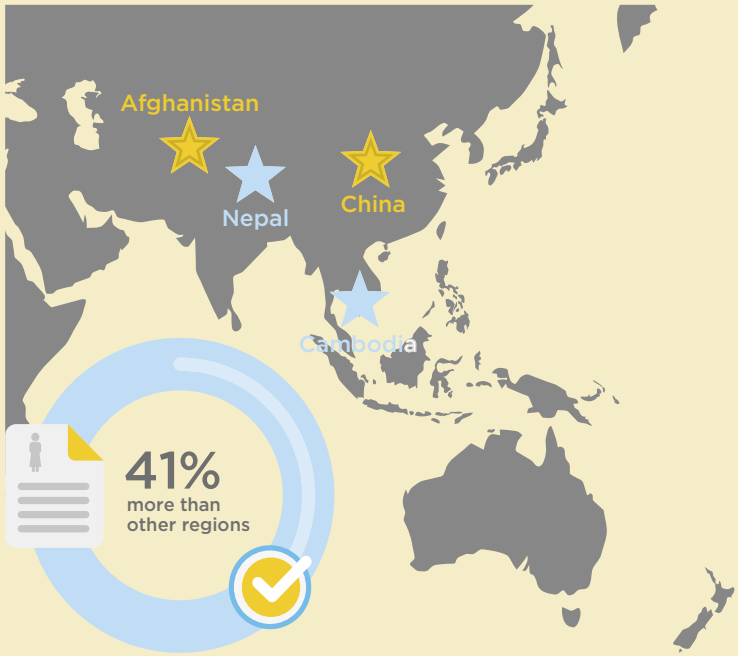
AMONG **34 REPORTS** REVIEWED
21% SATISFACTORY
72% GOOD OR VERY GOOD
7% UNSATISFACTORY

MORE THAN HALF OF 2015-2016 REPORTS **MET UN-SWAP** REQUIREMENTS



THE **AVERAGE SIZE** OF EVALUATION TEAMS **WAS 4.5** WITH A RANGE OF 11 PERSONS TO ONE PERSON.

65% WERE FEMALE, COMPARED WITH **35% MALE** EVALUATORS.



ASIA-PACIFIC REGION CONDUCTED **41% MORE EVALUATIONS** THAN OTHER REGIONS, AND HAD THE **2ND HIGHEST NUMBER** OF REPORTS RATED AS "GOOD" OR "VERY GOOD"

HIGHEST QUALITY RATINGS: **NEPAL, CAMBODIA, CHINA** AND **AFGHANISTAN**

ACRONYMS

AP	Asia-Pacific
CEDAW	Convention on the Elimination of all Forms of Discrimination Against Women
CO	Country Office
EPI	Evaluation Performance Indicator
EVAW	Ending Violence Against Women
GEOS	Global Evaluation Oversight System
GERAAS	Global Evaluation Reports Assessment and Analysis System
GEEW	Gender Equality and the Empowerment of Women
IEO	Independent Evaluation Office
MCO	Multi-Country Office
M&E	Monitoring & Evaluations
RO	Regional Office
ROAP	Regional Office for Asia and the Pacific
SOAR	Strengths, Opportunities, Aspirations and Results (Framework)
TOC	Theory of Change
TOR	Terms of Reference
UN	United Nations
UNEG	United Nations Evaluation Group
UN SWAP	UN System Wide Action Plan on Gender Equality and the Empowerment of Women
UN WOMEN	UN Women (United Nations Entity for Gender Equality and the Empowerment of Women)

TABLE OF CONTENTS

BACKGROUND

PURPOSE AND OBJECTIVES

The purpose and role of evaluation in UN Women is to contribute to learning on the 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 (GEEW).

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 good quality and credible evaluations particularly at decentralized level, in 2013, the UN Women's Independent Evaluation Office (IEO) introduced the Global Evaluation Reports Assessment and Analysis System (GERAAS).

GERAAS is an approach to rating evaluation reports using UN Women, UN Evaluation Group (UNEG) and UN System Wide Action Plan (SWAP) standards and indicators of report quality. The 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 the oversight system on a quarterly basis and corporate and decentralized evaluations on a yearly basis[1].

This Meta-Evaluation is the second component of a larger review process. The first part included analysis of the Evaluation Oversight System of UN Women in the Asia-Pacific region and the third component will entail a Meta-Analysis to capture key insights from evaluation reports rated as satisfactory or above. Together, all three components will used to share key insights from different evaluation reports in order to develop constructive lessons for future systemic strengthening of programming and organizational effectiveness[2].

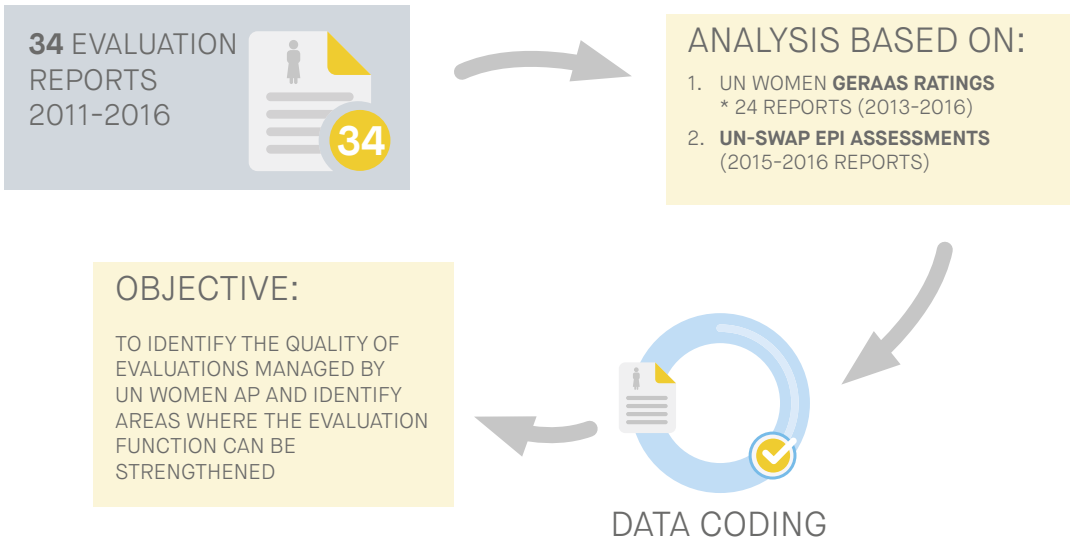
[1] Global Evaluation Report Assessment and Analysis System (GERAAS), UN Women Independent Evaluation Office, January 2015.

[2] Terms of Reference: Evaluation Consultant to Review Evaluation Oversight System (GEOS), Meta-Evaluation and Meta-Analysis (GERAAS) of UN Women Asia and the Pacific in 2010-2016.

The scope of this Meta-Evaluation is **34 evaluation reports**[3] managed by the UN Women Regional Office and Country Offices in the Asia and the Pacific region between 2011 and 2016[4].

The Meta-Evaluation addresses a number of key questions (see Box 1). **The list of questions was modified during the inception phase of the review process and further expanded from the original list included in the Terms of Reference (TOR)**. The list in Box 1 includes questions relevant only for the Meta-Evaluation.

META-EVALUATION APPROACH



MAIN LIMITATION: GERAAS ratings only available for reports after 2013 (comparability challenges between different rating systems – pre-GERAAS ratings generally higher).

[3] The original TOR included 30 evaluation reports; however the final list of reports proposed for inclusion in the Inception Report increased to 41 in order to include reports completed in 2016 and reviewed under GERAAS during the first quarter of 2017. Based on a review of the reports, the final list of 41 was reduced to 34 reports. The seven reports excluded from this meta-Evaluation include: a) three reports completed in 2009 (since this was outside the timeframe of the meta-evaluation); b) three that did not meet the requirement of evaluation reports (i.e. evaluability and training impact assessments) and c) one which was a duplication in the original terms of reference.

[4] Since no evaluation reports were available for 2010, the timeframe for the Meta-Evaluation is now 2011-2016.

BOX 1: META-EVALUATION KEY QUESTIONS

1. How were evaluation reports in the region rated overall by GERAAS (including by year, parameter)? Were there any improvements in the last five years?
2. What was the distribution and rating of evaluations across countries, thematic areas and evaluation types? What trends are evident particularly across the eight UNEG parameter areas?
3. Are there any missing evaluations in terms of thematic areas, types, and countries?
4. What are the main trends related to evaluation management (i.e. structure and composition of evaluation teams and quality of evaluation terms of reference)?
5. What are the strengths and weaknesses of the evaluation reports?
6. What are the recommendations to further strengthen evaluations in region?

METHODS

In aggregating and synthesizing key insights from the portfolio of 34 evaluation reports, a realist synthesis approach [5] was applied which involved distilling all relevant existing evidence in order to enable further analysis of evaluation quality. During the Meta-Evaluation process, data from each of the reviewed evaluation reports was harvested and coded using Excel. Coding for the Meta-Evaluation related to the eight UNEG parameter areas. Identified evaluation good practices were also harvested and coded in order to provide illustrative examples and support further replication.

The primary data source for the Meta-Evaluation was the UN Women GERAAS Executive Review Templates; however as these were only available for reports managed in 2013-2016 (n=24 reports), information from the 2011 and 2012 reports (n=10 reports) was directly harvested and used in the analysis of data. In order to provide more nuanced analysis related to the integration of GEEW in the evaluation reports, UN Women's UN-SWAP EPI assessment ratings and raw numerical scores for 2015 and 201 were also used.

Mixed evaluation methods were applied using quantitative data from the GERAAS ratings and SWAP EPI scores, and qualitative information harvested directly from the reports and GERAAS Executive Review Templates. Deeper analysis was undertaken, where data and information was available, to validate and provide explanations for identified trends. The findings from the analysis of the UN Women Asia-Pacific Evaluation Oversight System were also used to provide further evidence in understanding data trends.

LIMITATIONS

1. The GERAAS ratings were only available for 2013-2016 reports; prior to 2013, different quality assessment approaches were used which likely posed comparability challenges between the different ratings systems, especially since the pre-GERAAS ratings tend to be higher overall.

2. The ability to provide comparisons of quantitative data for each of the eight GERAAS parameter areas was limited by the non-availability of ratings for these areas in the 2011 and 2012 reports.

3. Gaps in data availability remain which meant that not all of the reports could be included in analyses of data (i.e. for 2013-2016 reports, the GERAAS executive template was not available for the 2013 External Evaluation of the EVAW Commission Project in Afghanistan and there was no rating for the 2012 Evaluation of the Programme "Supporting Gender Equality and Women's Rights in Timor-Leste"). Numerical scores for UN-SWAP EPI ratings were only available for 2015 and 2016 so quantitative analysis is only possible for these two years.

[5] Realist synthesis is an approach to reviewing research evidence on complex social interventions, which provides an explanatory analysis of how and why they work (or don't work) in particular contexts or settings (source: betterevaluation.org/sites/default/files/RMPmethods2.pdf).

OVERALL PERFORMANCE

OVERALL RATINGS

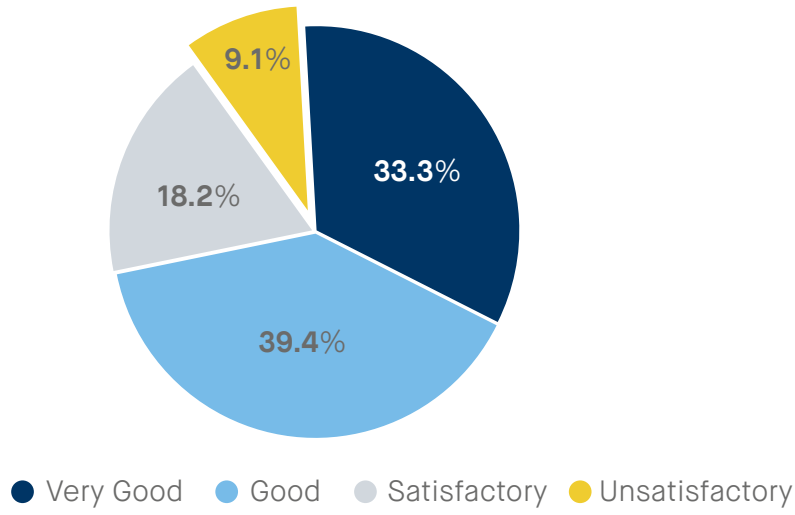
According to the quality ratings for 2011 to 2016, for 33 evaluations[6] , **91 per cent of reports (n=30) were "satisfactory" and above**, indicating that these reports can be used with a reasonable degree of confidence to inform evidence-based planning and programming. Based on the ratings, **73 per cent of evaluation reports in the region were rated in the upper range as "very good"** (33 per cent) or "good" (39 per cent). Only 9 per cent (n=3) were rated as "unsatisfactory".

Nine out of 11 reports (82 per cent) rated as "very good" were managed by M&E Officers in MCOs and COs. This points to the strong capacity of M&E Officers in the region and their important role in supporting quality evaluations.

91 PER CENT

EVALUATION REPORTS ARE RATED **SATISFACTORY** OR ABOVE

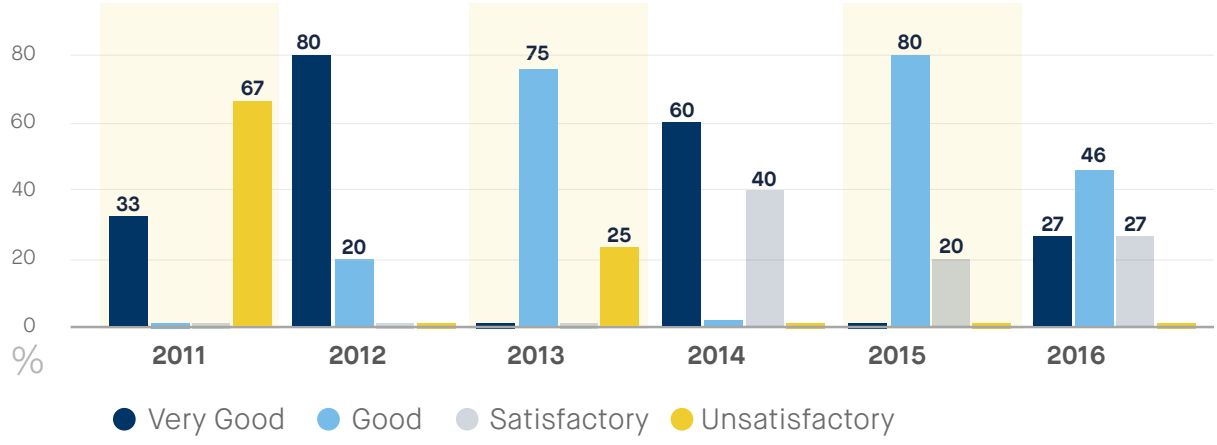
FIGURE 1: QUALITY OF EVALUATION REPORTS (2011-2016)



[6] Due to the fact that one evaluation report was not rated ("Evaluation of Supporting Gender Equality and Women's Rights in Timor-Leste, 2012), only 33 reports are analysed within this section.

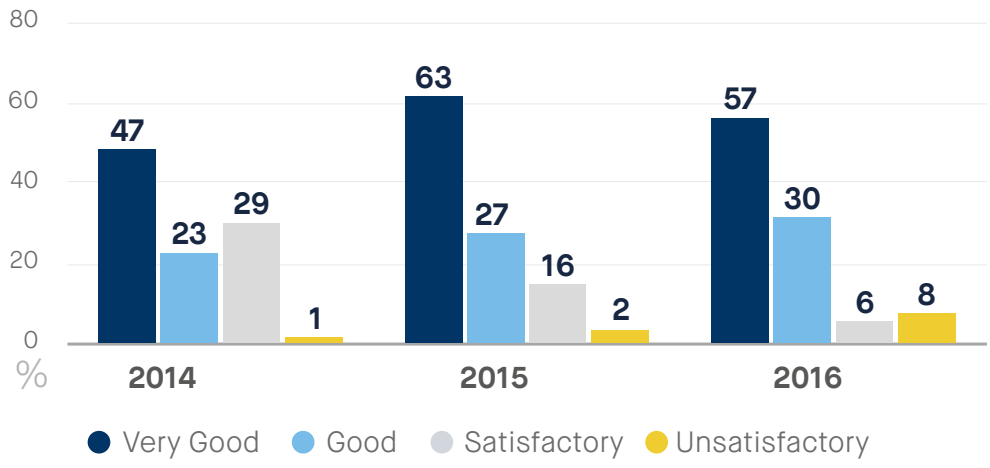
In comparing evaluation report quality between 2011 and 2016, **reports rated as "satisfactory" or above increased from 33 per cent in 2011 to 100 per cent of reports during the last three years** (see Figure 2). Possible reasons for this improvement may be the increased investment in evaluation capacity of the M&E staff as well as the increased financial resources made available for evaluation and the "Quality Assurance Process for Decentralised Evaluation in Asia and the Pacific" set out in the Regional Evaluation Strategy.

FIGURE 2: COMPARISON OF QUALITY RATINGS FOR THE AP REGION (2011-2016)



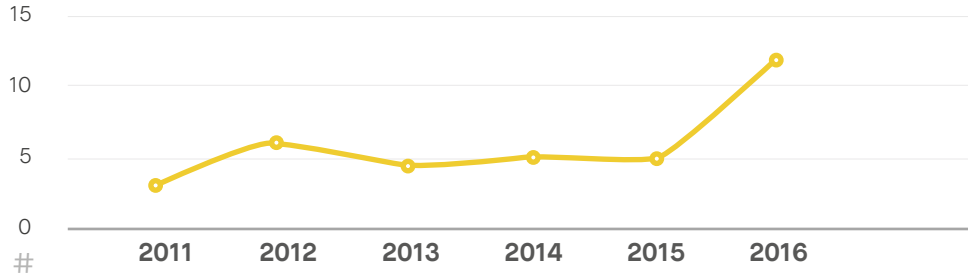
Despite this improvement, **three reports were assessed only as "satisfactory" in 2016 and the percentage of reports assessed as "very good" or "good" dropped from 80 per cent in 2015 to 73 per cent in 2016**, underlining the need for continued investment in strengthened evaluation capacity. When examining evaluation expenditure for the region, financial data shows a progressive decrease in investment in evaluation capacity building from 29 per cent in 2014 to only 6 per cent in 2016 (see Figure 3).

FIGURE 3: EVALUATION EXPENDITURE IN THE AP REGION (2014-2016)



Another contributing factor to this trend could be the significant change in the number of evaluations conducted from an average of five per year between 2011 and 2015 to 11 in 2016 (see Figure 4). It is plausible that this increased volume of evaluations could have placed additional burden on the Regional and Evaluation Specialist and M&E staff and thus affected their ability to provide high quality assurance. Other influencing factors could be project design quality and lack of baseline data, limited gender analysis and absence of a Theory of Change resulting in lower evaluability.

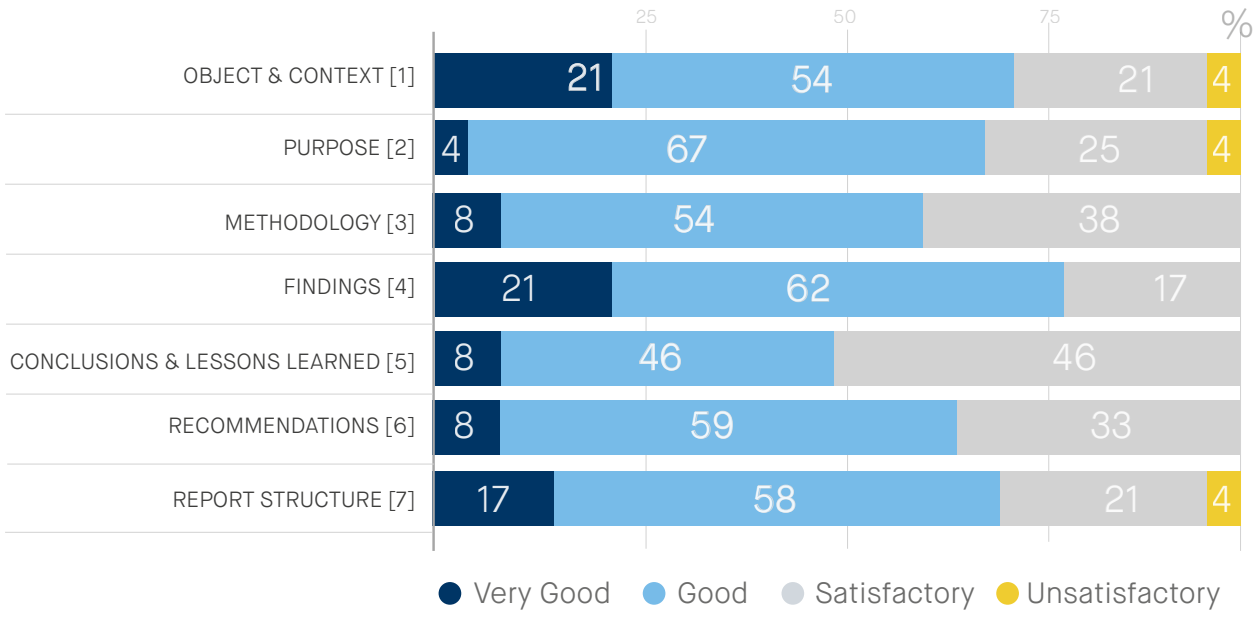
FIGURE 4: NUMBER OF EVALUATIONS CONDUCTED IN THE AP REGION (2011-2016)



SUMMARY OF GERAAS RATINGS BY PARAMETER[7]

Report findings (parameter 4) were given the highest quality rating with 83 per cent of reports rated as "very good" or "good" in this aspect. The other highly rated areas were object and context of the evaluation (parameter 1) and the report structure (parameter 8). **The lowest quality ratings were for conclusions and lessons learned, followed by methodologies (parameters 5 and 3) with 46 per cent and 38 per cent of reports assessed as "satisfactory"** (See Figure 5).

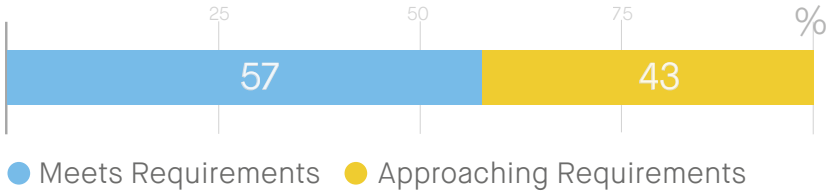
FIGURE 5: GERAAS RATINGS PER PARAMETER (2013-2016 REPORTS)



[7] Due to the fact that one evaluation report was not rated ("Evaluation of Supporting Gender Equality and Women's Rights in Timor-Leste, 2012), only 33 reports are analysed within this section.

According to UN-SWAP EPI ratings for 2014-2016 evaluation reports[8], **more than half of the reports (57 per cent) were assessed as "meeting requirements" for integration of GEEW whilst 43 per cent were rated as "approaching requirements"**. No reports were rated as "exceeding requirements" or "missing requirements". The ratings also show **a significant improvement in integrating GEEW and as of 2016, all reports were "meeting requirements" of the UN SWAP Evaluation Performance Indicator**.

FIGURE 6: PARAMETER 7 RATINGS FOR UN-SWAP EPI (2014-2016 REPORTS)



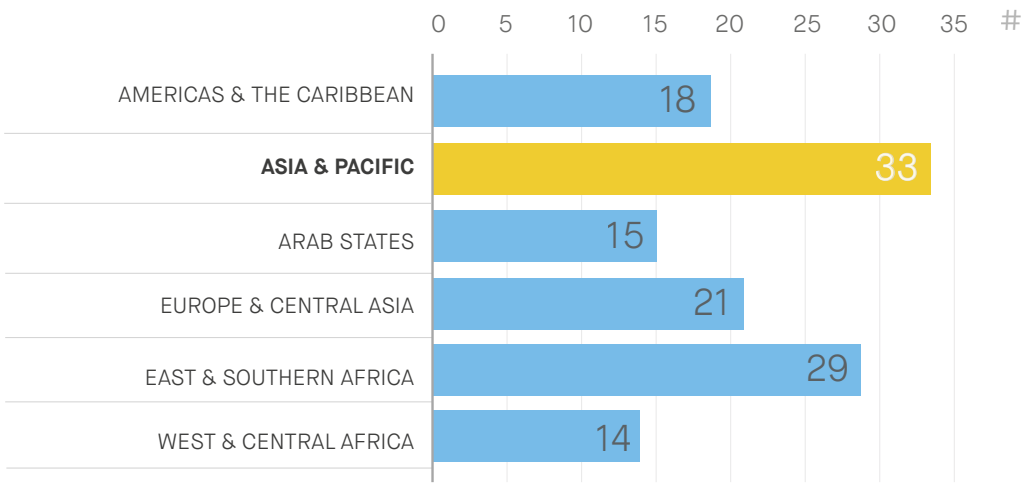
[8] SWAP EPI ratings were applying within GERAAS from 2014 onwards.

GLOBAL AND REGIONAL VIEW

GLOBAL COMPARISON

Out of six regions, the Asia-Pacific region had the **highest number of evaluations completed between 2011 and 2016 with 41 per cent more evaluations conducted than other regions** (see Figure 7). This corresponds with the overall level of investment in the region compared to other regions[9].

FIGURE 7: GLOBAL COMPARISION OF NUMBER OF EVALUATIONS COMPLETED

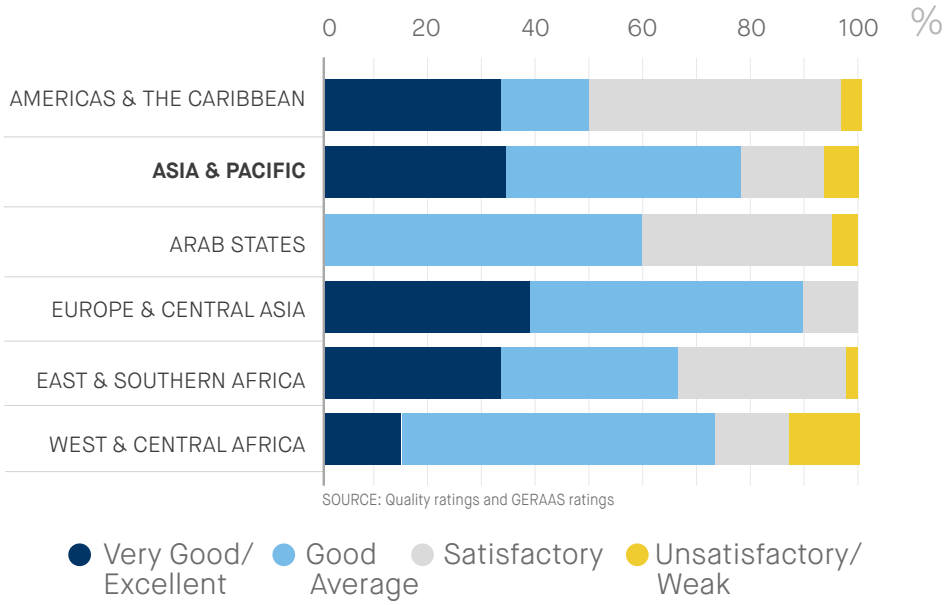


SOURCE: GEOS January - December 2016

[9] Source: UN Women Planning and Programme Coordination Unit, Meta-Analysis of UN Women 2017 Field Annual Work Plans.

The Asia-Pacific region also had the **second highest percentage of reports rated as "very good" and "good"** (see Figure 8).

FIGURE 8: COMPARISON OF QUALITY RATINGS ACROSS REGIONS (2011-2016)



A contributing factor to these higher numbers is that, compared with other regions, the Asia-Pacific region has the second highest number of M&E staff overall[10] and the third highest ratio of full-time M&E officers[11]. It is interesting to note that the office with the highest ratio of full-time staff (the Europe and Central Asia Regional Office) also had the highest percentage of reports rated as "very good" and "good[12]" so the correlation between these factors appears strong.

The Asia-Pacific region's higher quality ratings can also be attributed to high financial investment in evaluation. In 2014, the Asia-Pacific region had the highest investment with 1.3 per cent of its overall budget invested in evaluation and during the same year[13], it also had the highest number of evaluations rated as "very good"[14].

[10] The West and Central Africa region has the highest number of M&E staff (15).

[11] The Europe and Central Asia region has the highest ratio of full-time M&E officers (57 per cent) followed by the West and Central Africa region (56 per cent).

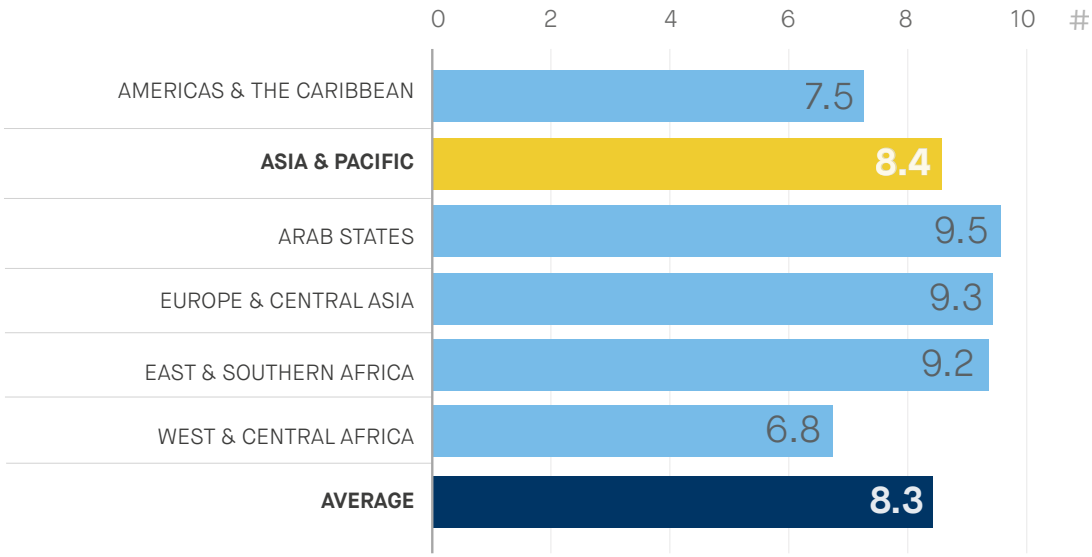
[12] See the report "Review of UN Women's Oversight System in Asia and the Pacific" for more detail.

[13] Source: GEOS data for 2014.

[14] Global data to compare financial investment in evaluation is only available for 2014, so this assertion is based on limited data.

[15] Source: Data provided by the UN Women Independent Evaluation Office on 19 January 2017.

FIGURE 9: REGIONAL COMPARISION OF UN SWAP EPI SCORES FOR 2016 EVALUATION REPORTS



SOURCE: UN Women UN SWAP 2016

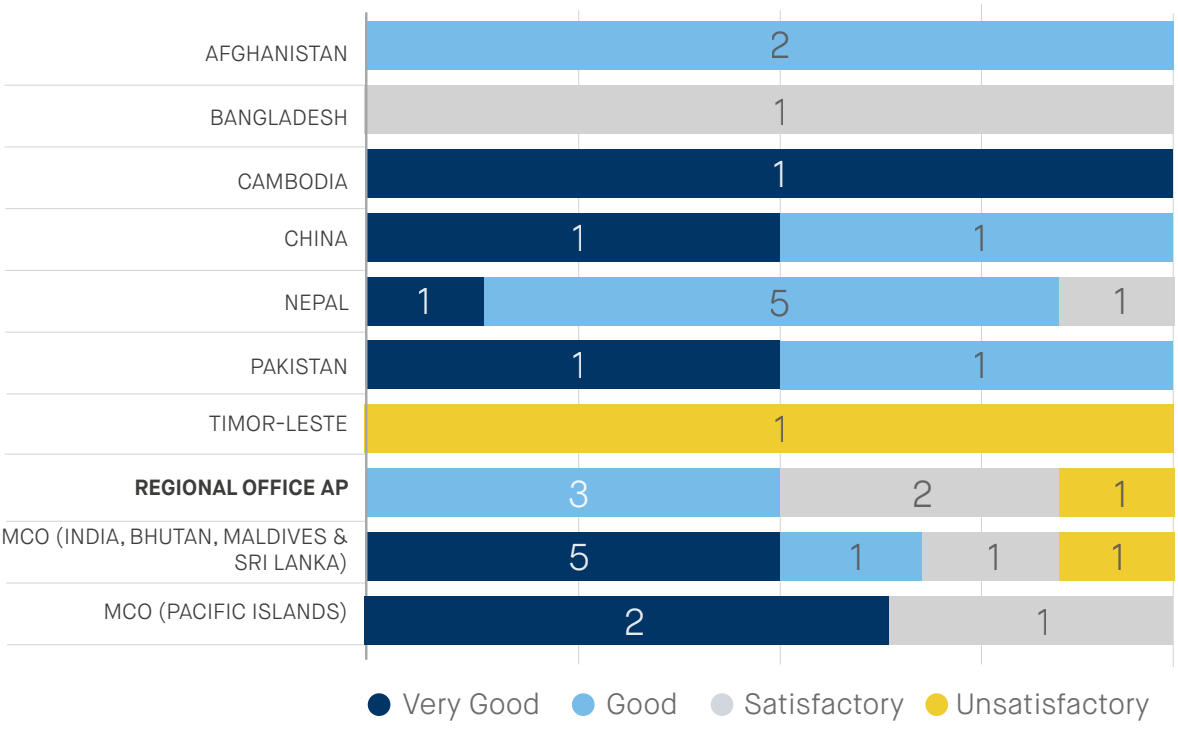
In relation to integration of GEEW and human rights, independent assessments of the 2015-2016 UN SWAP EPI evaluations show that, compared with other regions, the Asia-Pacific region ranked fourth in terms of its average score (see Figure 9). Although the region had the highest number of staff completing the UN Women e-learning course on gender responsive evaluation, the completion rate of M&E officers and focal points was 0 per cent[15] which may explain the discrepancy.

REGIONAL COMPARISON

Between 2011 and 2016, **almost half of the 34 evaluations conducted (44 per cent) were managed by two offices - the MCO for India, Bhutan, Maldives and Sri Lanka and the Nepal CO**, both of which are well-resourced with full-time M&E staff. **Evaluations of these two offices also had high quality rankings with 86 per cent of reports managed by the Nepal CO rated as "very good" or "good' and 75 per cent for the MCO** (see Figure 10). This points to the strong capacity of M&E Officers in the MCOs and their important role in supporting quality evaluations. Conversely, the Regional Office for Asia and the Pacific had the third highest number of evaluations with only 50 per cent rated as "very good" or "good". A possible factor contributing to this difference could be the limited number of M&E staff at the regional level to manage these evaluations.

For the majority of Country Office-led evaluations, 81 per cent were rated as "very good" or "good" and, with the exception of the Nepal CO, no CO conducted more than two evaluations during the six year timespan. **The COs in Nepal, Cambodia and Afghanistan as well as the China Project Office had the highest quality ratings and the lowest were for the Bangladesh and Timor-Leste COs**.

FIGURE 10: BREAKDOWN OF QUALITY RATINGS BY OFFICE



TRENDS BY TYPE AND SCOPE OF EVALUATION

TYPE OF EVALUATIONS AND INTERVENTIONS

Most evaluations (88 per cent) were programme evaluations focused on outcome-level results. Thematic evaluations constituted 9 per cent of evaluations and there was one meta-evaluation. There were no evaluations categorized as project evaluations in the UN Women GATE System.

The majority of evaluations were final and summative (71 per cent) compared with only 21 per cent formative mid-term evaluations (see Figure 11). Out of the 34 evaluations, seven were joint evaluations managed by UN Women and partners. **Quality ratings for mid-term evaluations were generally lower** with 50 per cent of reports rated as "very good" of "good" compared with 84 per cent of final evaluation reports. This could be attributable to a lower allocation of funds for mid-term evaluations (compared to final evaluations)[16]resulting in lower quality of evaluation consultants.

In terms of geographical focus, **41 per cent of evaluations were conducted at the national-level** whilst the proportion of regional/multi-country, and sub-national-level evaluations was generally balanced (see Figure 12). **The highest quality ratings were for sub-regional evaluations with 80 per cent of reports rated as "very good" or "good" compared with 70 per cent for regional/ multi-country evaluation reports and 60 per cent for evaluations at the national level**[17].

FIGURE 11: TYPE OF EVALUATIONS

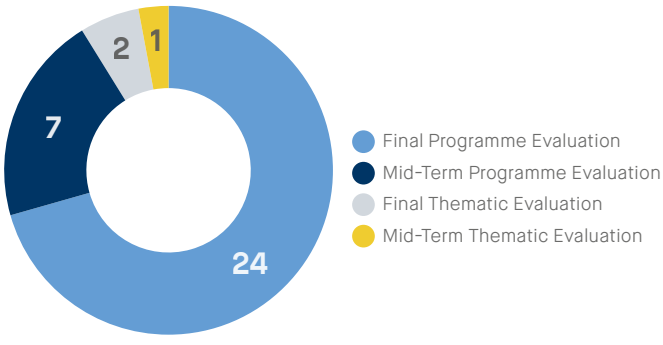
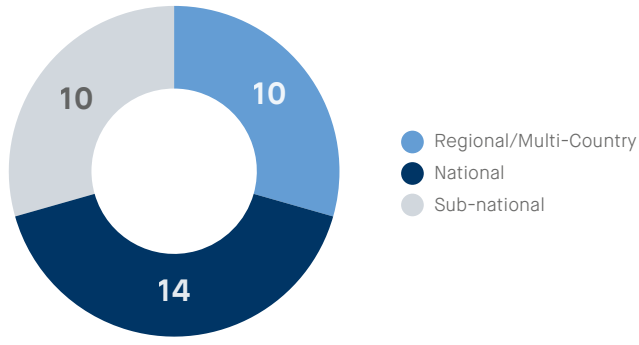


FIGURE 12: GEOGRAPHIC FOCUS



[16] Source: UN Women Regional Office for Asia and the Pacific, 2009-2015 Evaluation Summary.

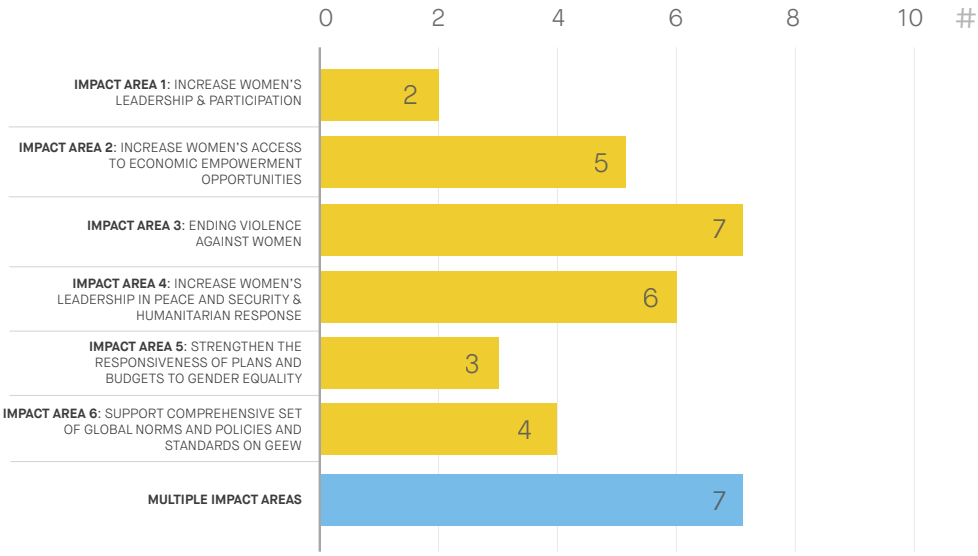
[17] The report "Review of UN Women's Oversight System in the Asia-Pacific Region" cited quality of consultants as one of the main factors affecting overall quality of evaluation reports.

THEMATIC FOCUS

The thematic scope of evaluations conducted between 2011 and 2016 included all Strategic Plan Impact Areas with **the highest number of evaluations focused on EVAW (21 per cent)** and multiple impact areas (21 per cent). Women's leadership in peace and security and humanitarian response was also a major focus with 26 per cent of reports addressing this theme. **The lowest number of reports (6 per cent) corresponded with women's leadership and participation followed by gender responsive plans and budgets (9 per cent).** This range in thematic focus area generally mirrors UN Women's global investment across each area[18].

Whilst most themes received an average rating of "good" or above, the exception to this was the **evaluation reports on global norms, policies and standards on GEEW where only one report out of four was rated as "good" or above.** A possible factor contributing to this may be the complexity of policy evaluations compared to programme evaluations.

FIGURE 13: THEMATIC FOCUS OF 2011-2016 EVALUATION REPORTS



SOURCE: GEOS January - December 2016

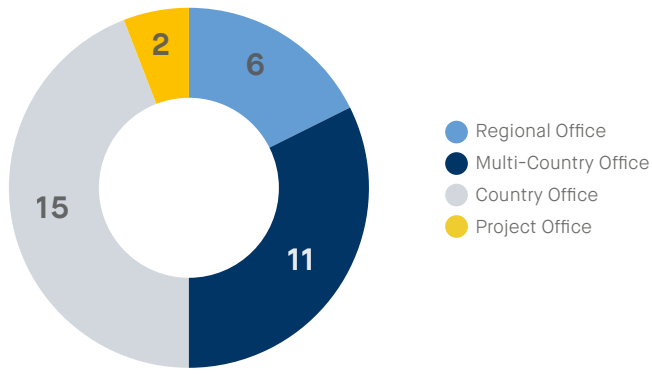
[18] Source: UN Women Planning and Programme Coordination Unit, Meta-Analysis of UN Women 2017 Field Annual Work Plans (AWP).

EVALUATION MANAGEMENT TRENDS

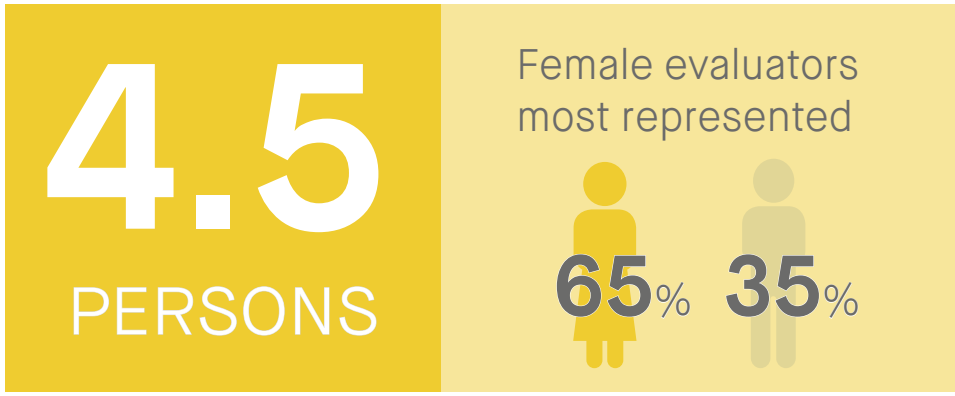
Almost half of the evaluations conducted between 2011 and 2016 were commissioned by Country Offices (44 per cent) followed by MCOs (32 per cent). UN Women directly managed 91 per cent of evaluations. Two out of seven joint evaluations were co-managed with partners (one with a donor and the other with a regional body) and one was managed by the Resident Coordinator's Office. For the two jointly managed evaluations (both of which were multi-country in scope and conducted in 2011), one report rated "unsatisfactory" and one rated "very good".

The majority of evaluations were conducted by companies or research centers (62 per cent) and 38 per cent by independent consultants. **Quality ratings for companies were higher with 81 per cent of reports rated as "very good" or "good"** compared to 67 per cent for independent consultants. For the two offices with the largest number of evaluations, a high reliance on companies can be seen with 86 per cent of evaluations for the Nepal CO conducted by companies and 75 per cent for the MCO India, Bhutan, Maldives and Sri Lanka. Given that most evaluations for these offices also received high quality ratings, there may be value in the use of companies with an established quality record to support offices with a high volume of evaluations.

FIGURE 14: TYPE OF OFFICE COMMISSIONING THE EVALUATION



The average size of evaluation teams was 4.5 with a range of 11 persons to one person. **65 per cent of evaluators were female, compared with 35 per cent male evaluators**.

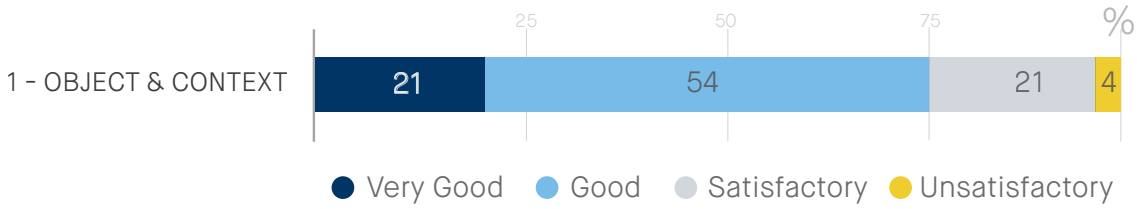


TRENDS BY QUALITY PARAMETERS

PARAMETER 1: OBJECT AND CONTEXT OF THE EVALUATION

This parameter area ranked highly with 75 per cent of reports rated as "very good" or "good". Overall, most evaluations included a good balance between context and programme information.

FIGURE 15: RATINGS FOR PARAMETER 1 (2013-2016 REPORTS)



Descriptions of the logic model and/or results framework were included in the majority of reports although a limited number adequately described the chain of results and highlighted the causal pathways between the levels of results. Only six reports included comprehensive analysis and re-construction of programme Theories of Change.

Most evaluation reports contained detailed analysis of the context in which the programmes operated including key social, economic and political factors; however comprehensive gender analysis was missing from most reports (see Parameter 7 for more detail).

Almost half of the reports lacked adequate identification of stakeholding groups and their roles was generally limited across most reports.

BOX 2: FEATURED GOOD PRACTICE: EVALUATION OBJECT AND CONTEXT

FINAL EVALUATION OF THE PROGRAMME "MAKING POLITICS WORK WITH WOMEN IN NEPAL" (2013)

The evaluation report contains a robust background analysis of the socio-political context relevant to the programme with strong gender analysis included. A particular good practice is the analysis of how the intervention was designed to support implementation of relevant gender equality and human rights normative frameworks, including CEDAW, the Beijing Platform of Action and UN Security Council Resolution 1325 as well as broader human rights frameworks such as ILO Convention 169 and the UN Declaration on the Rights of Indigenous Peoples. The analysis also examines trends related to women's political participation including potential barriers and opportunities. The programme is clearly and succinctly described in brief summary tables which includes a summary of expected results and indicators as well as a summary of results achieved.

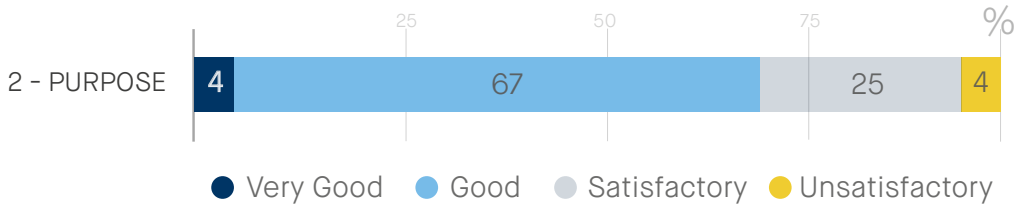
FINAL EVALUATION OF UN WOMEN'S ANTI-HUMAN TRAFFICKING PROGRAMME IN INDIA (2014)

The Theory of Change is described in detail and provides important evidence of understanding the project design. Analysis of the TOC is used to support a theory and evidence-based approach to assessing results and enabled the evaluation process to focus on causes, linkages and assumptions related to the programme interventions.

PARAMETER 2: PURPOSE, OBJECTIVES AND SCOPE

The majority of reports were rated as "good" or "satisfactory" in terms of meeting requirements under Parameter 2.

FIGURE 16: RATINGS FOR PARAMETER 2 (2013-2016 REPORTS)



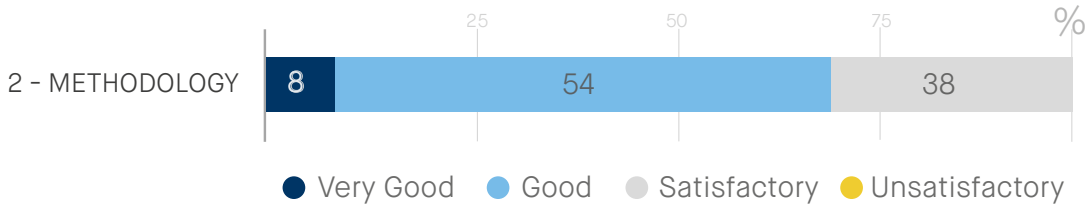
Most reports aptly described the objectives of the evaluation and included an overview of the evaluation criteria and key questions; however the discussion of purpose and scope was limited. In explaining the purpose, many reports lacked detail or deeper insights about why the evaluation was needed and how it will be used. Discussion of scope was either missing or limited in many reports with insufficient detail about what is to be covered by the evaluation in chronological and geographical terms and what issues were excluded from the evaluation.

Whilst the majority of reports did not include explicit mention of how gender and human rights were included in the evaluation objectives and scope, for reports commissioned in 2016, there was a clear improvement with more explicit reference to integration of GEEW and human rights in the objectives, scope, criteria and questions (see Parameter 7 for more detail).

PARAMETER 3: METHODOLOGY

Parameter 3 was the second lowest rated parameter areas with 38 per cent of reports rated as "satisfactory". A contributing factor was that 65% of reports did not address the extent to which the evaluation design included ethical safeguards or implemented measures to ensure that the evaluation process observed ethical standards.

FIGURE 17: RATINGS FOR PARAMETER 3 (2013-2016 REPORTS)



Another general weakness was the over-focus on qualitative methods by more than a third of evaluation reports, even in instances where use of mixed methods was required by the evaluation terms of reference. Whilst a number of good practices can be found where quantitative information from monitoring reports or survey data was collected and used; overall, there was limited application of mixed methods, based on qualitative and quantitative data, in triangulating findings.

Most reports also lacked discussion of whether methods were appropriate for analysing GEEW and human rights and how data collection approaches and tools were designed to ensure maximum stakeholder inclusion and mitigate potential barriers to participation (see Parameter 7 for more detail).

Despite these weaknesses, between 2011 and 2016, an increasing number of good practices were present across a number of areas including introductive of innovative and gender-responsive evaluation methods (see Box 3 below).

BOX 3: FEATURED GOOD PRACTICE: EVALUATION METHODS

FINAL EVALUATION OF UN WOMEN'S ANTI-HUMAN TRAFFICKING PROGRAMME IN INDIA (2014)

The evaluation employed a quasi-experimental design approach using before-and-after as well as 'case' and 'control' group comparisons to establish a counterfactual. Data was collected through a mixed-methodology approach where quantitative and qualitative information were collected from the field through relevant questionnaires, tools and templates.

EVALUATION OF THE PROGRAMME "EMPOWERING WOMEN IN RURAL AND INFORMAL SETTINGS THROUGH CAPACITY DEVELOPMENT", INDIA (2014)

The report includes an attempt at explicitly using mixed methods (including a quantitative survey and qualitative interviews) combined through an appreciative inquiry approach (involving a reflection exercise with project staff). The report also provides a good justification of the design and methods used and the evaluation framework specifies the tools used to answer each question.

END OF TERM EVALUATION OF THE PROJECT "TOWARDS GENDER EQUALITY: WOMEN'S ECONOMIC EMPOWERMENT HOME-BASED WORKERS, PHASE II, PAKISTAN (2016)

Eleven case studies were developed on success stories gleaned from Key Informant Interviews and/or Focus Group Discussions where in depth interviews were conducted with the relevant women home-based workers and with one male change maker and one male government official.

MID-TERM EVALUATION OF STRENGTHENING IMPLEMENTATION OF WOMEN, PEACE AND SECURITY AGENDA IN NEPAL (2015)

The report presents a series of data quality assurance mechanisms to make the evaluation findings credible.

FINAL EVALUATION OF THE MAKING POLITICS WORK WITH WOMEN PROJECT, NEPAL (2013)

In order to ensure informed consent, a consent form was prepared and attached in the front page of the evaluation questionnaire. The form included the objective of the evaluation and assured confidentiality of responses. The consent form was read out and agreement of the sampled respondents was obtained before administering the questionnaire.

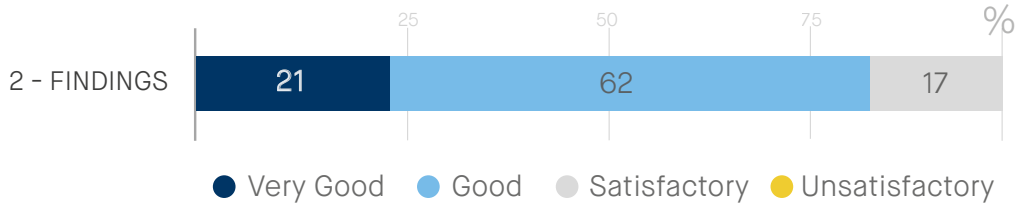
EVALUATION OF UN WOMEN SUPPORT TO PWN+ ON HIV AND AIDS PROJECT IN INDIA (2012)

Ethical safeguards were applied to ensure that the evaluation process observed ethical standards. Given the sensitive nature of the evaluation subject, anonymity of informants was ensured by assigning codes to the respondents and groups. To the extent possible, each interview was conducted without others present except on the request of the respondent. Quotations from the interviews used in the report were kept anonymous.

PARAMETER 4: FINDINGS

The parameter of report findings received the highest quality rating with 83 per cent of reports rated as "very good" or "good" in this aspect.

FIGURE 18: RATINGS FOR PARAMETER 4 (2013-2016 REPORTS)



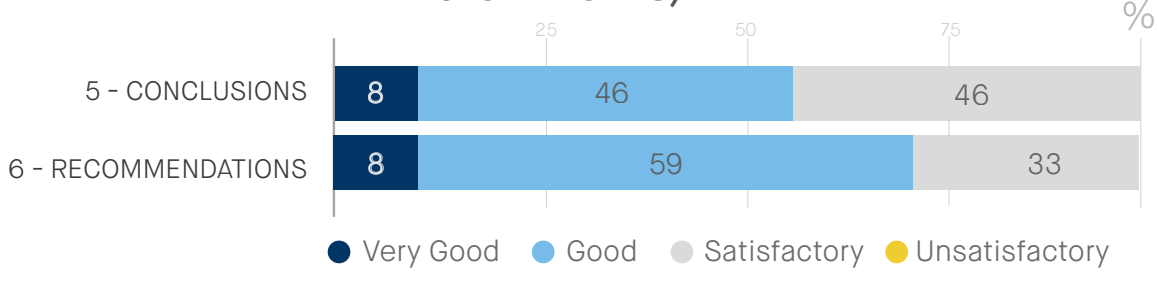
Overall, the findings directly responded to the criteria and in some instances also to each individual question. Most findings sections included a discussion of unexpected results, challenges and opportunities. Findings generally reflected a systematic and appropriate analysis of data; although most findings relied on qualitative data.

Areas for improvement include greater depth of evaluative insights and use of gender analysis and a more systematic use of data and evidence to support findings.

PARAMETERS 5 & 6: CONCLUSION, LESSONS LEARNED AND RECOMMENDATIONS

Overall, conclusions and lessons sections received the lowest quality with 46 per cent of reports assessed as "satisfactory". For nearly half of all reports (44 per cent), conclusions included a summary of findings without additional insights or identified solutions to problems; however a significant improvement in the quality of conclusions was evident for reports completed in 2016. The need to structure conclusions around the evaluation criteria and to more explicitly marshal the supporting evidence from the evaluation findings was also highlighted in the GERAAS review for a number of reports. Additionally, many reports need to address programme strengths and weakness in a more balanced matter as some reports tended to over-focus on strengths.

FIGURE 19: RATINGS PER PARAMETERS 5& 6 (2013-2016 REPORTS)



A key weakness was that 71 per cent of reports did not include a general lessons learned section with harvested lessons that are universal with potential for broader application (even though in most instances, this was required within the terms of reference). The omission of lessons learned is a significant constraint in ensuring that evaluation reports support organizational learning.

Whilst most reports contained a manageable number recommendations emanating from the findings and conclusions, a general shortfall was that these were not always actionable with a timeframe or specific target group indicated. Where UN Women was listed as responsible, it was not always specified which level of staff. Only a few reports described the process followed in developing the recommendations, including the level and type of consultation with stakeholders.

BOX 5: FEATURED GOOD PRACTICE: EVALUATION CONCLUSIONS, LESSONS LEARNED & RECOMMENDATIONS

FINAL EVALUATION "EMPOWERING WOMEN IN RURAL AND INFORMAL SETTINGS THROUGH CAPACITY DEVELOPMENT", UN WOMEN INDIA MCO (2014)

The report conclusions take the analysis of the findings deeper and are well linked to the evidence. They answer all of the main evaluation questions and correspond with the criteria.

END OF TERM EVALUATION OF THE PROJECT "TOWARDS GENDER EQUALITY: WOMEN'S ECONOMIC EMPOWERMENT HOME-BASED WORKERS, PHASE II, PAKISTAN (2016)

The report is particularly strong at presenting lessons learned that contribute to general knowledge, and at presenting them according to different themes (i.e.: sensitizing and working with men, private sector engagement and broader organizational lessons) which increases its usefulness in different contexts. Recommendations are also presented thematically, identify the target group and reflect an understanding of the commissioning organization as well as potential constraints. The process followed in developing the recommendations is described well, including the level and type of consultation with stakeholders.

FINAL EVALUATION "STRENGTHENING IMPLEMENTATION OF WOMEN, PEACE AND SECURITY AGENDA, NEPAL (2016)

The recommendations are included in a table for greater clarity and clearly identify the target group for each action; reflect an understanding of the commissioning organization and potential constraints; and a timeframe is provided for the actions to be taken in each case.

FORMATIVE EVALUATION OF THE PACIFIC ENDING VAW FACILITY FUND, UN WOMEN FIJI MCO (2014)

The process for developing the recommendations showed a high level of participation designed to support strong actionability and ownership of the recommendations.

PARAMETER 7: GENDER AND HUMAN RIGHTS

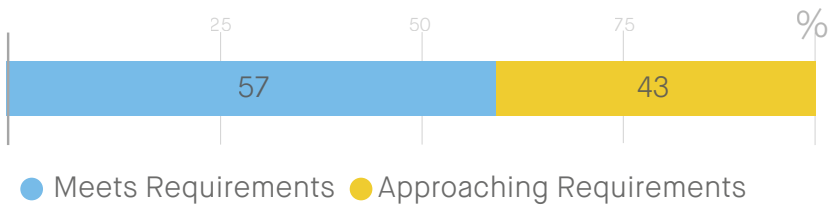
Assessment of integration of GEEW and human rights in reports is based on the four-point criteria set out under the UN-SWAP Evaluation Performance Indicator (EPI).

TABLE 1: UN-SWAP EPI CRITERIA

CRITERIA 1	GEEW is integrated in the evaluation scope of analysis and evaluation indicators are designed in a way that ensures GEEW-related data will be collected
CRITERIA 2	GEEW is integrated in evaluation criteria and evaluation questions specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved.
CRITERIA 3	A gender-responsive methodology, methods and tools and data analysis techniques are selected.
CRITERIA 4	The evaluation findings, conclusions and recommendations reflect a gender analysis.

According to UN-SWAP EPI ratings for 2014-2016 evaluation reports, **more than half of the reports (57 per cent) were assessed as "meeting requirements" for integration of GEEW across the four criteria areas. 43 per cent of reports were rated as "approaching requirements" with no reports "exceeding requirements"** thus, highlighting a need for continued investment in supporting gender-responsive evaluation capacity in the region.

FIGURE 20: PARAMETER 7 RATINGS - UN-SWAP EPI (2014-2016 REPORTS)

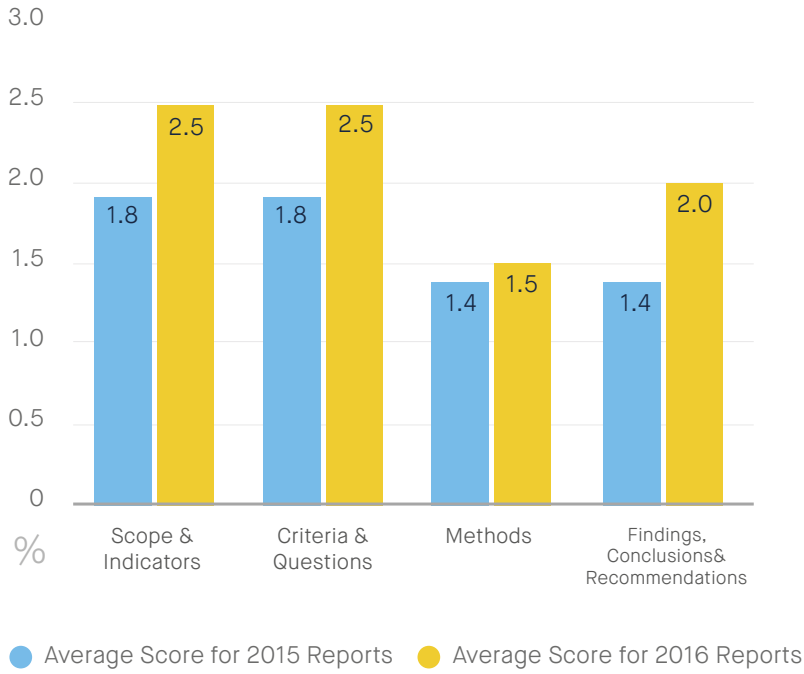


As noted previously, the Asia-Pacific region ranked fourth in terms of its average score and although the region had the highest number of staff completing the UN Women e-learning course on gender responsive evaluation, the completion rate of M&E officers and focal points was 0 per cent. As the main managers of evaluation, it is critical that M&E staff develop a deep understanding of gender responsive evaluation so that they can provide the necessary guidance to programme staff managing evaluations and assume an effective quality assurance role. It is also important for senior managers, including Country Representatives and Deputy Representatives who also supervise and advise on evaluations to have a solid grasp of the key elements of gender-responsive evaluation. To-date, only one senior manager (Country Representative) has undertaken the e-learning course. No other senior programme staff (at the level of P4 and up) have successfully completed it.

When examining quantitative scorecard data from independent assessments of the UN SWAP EPI for 2015 and 2016, **a significant improvement is evident in the integration of GEEW with the average score increasing from 6.6 (approaching requirements) in 2015 to 8.4 (meeting requirements) in 2016**.

Across the criteria areas, the greatest improvements were in integrating GEEW into the scope and indicators of the evaluation and across the evaluation criteria and questions (see Figure 21)

FIGURE 21: COMPARISON OF UN SWP EPI ASSESSMENT SCORES FOR 2015 AND 2016 EVALUATION REPORTS FOR THE ASIA-PACIFIC REGION



An increased number of evaluations included discussions of the relevance of interventions to women's needs and human rights frameworks, the effectiveness of interventions in advancing gender equality, and in the extent to which interventions addressed underlying causes of gender equality; however the inclusion of comprehensive gender analysis in the contextual overview and in the stakeholder analysis was limited.

GEEW and human rights were often addressed under the criterion of effectiveness; but integration of GEEW and human rights across other UNEG criteria was limited especially for the efficiency and sustainability criteria. Only one report contained a stand-alone criterion related to GEEW (on inclusiveness) which enabled it to maintain a strong gender analysis across the findings, conclusions and recommendations.

Triangulation of scores for criteria 1 and 2 with a review of evaluation TOR provides evidence of a strong correlation between TOR that integrated GEEW across evaluation questions and evaluations that received high UN SWAP EPI ratings.

BOX 6: FEATURED GOOD PRACTICE: INTEGRATION OF GEEW IN EVALUATION TOR

CAMBODIA COUNTRY OFFICE

The CO included the following requirement within the TOR for the Final Evaluation of the EVAW Programme in 2016 (which was rated as "meeting" UN SWAP EPI requirements):

"The evaluation will be based on gender and human rights principles, as defined in the UNEG Guidance. The evaluation methodology will employ mixed methods for data collection and HR/GE well integrated across criteria and TOR."

BANGLADESH COUNTRY OFFICE

The TOR for the Final Evaluation of the Gender and Climate Change project in 2016 required the inclusion of a gender expert within the team (2016). The report was rated as "meeting" SWAP EPI requirements).

NEPAL COUNTRY OFFICE

The CO included the following requirement within the TOR for the Final Evaluation of the Strengthening Implementation of Women, Peace and Security Agenda" Programme in 2016 (which was rated as "meeting" UN SWAP EPI requirements):

"The evaluation is expected to take a gender-responsive approach. Gender-responsive evaluations use a systematic approach to examining factors related to gender that assesses and promotes gender equality issues and provides an analysis of the structures of political and social control that create gender equality."

Continuing challenges were seen in designing and applying gender-responsive methods and the lowest ratings were for integration of GEEW and human rights into the evaluation conduct. Whilst many evaluation reports included statements that GEEW and human rights principles and approaches were applied, and an increasing number referenced the UNEG Guidance and Handbook for Integrating Human Rights and Gender, the majority of evaluations did not meet the specific requirements under criterion 3. For example, most reports did not adequately discuss whether and how methods chosen were appropriate for analyzing GEEW and human rights. Reports generally cited the use of participatory approaches to maximize stakeholder inclusion but most did not sufficiently discuss potential barriers to participation and measures taken to address these in the design of data collection tools and methods.

During 2016, improved use of gender-responsive methodologies was evident. A contributing factor to this was the requirement for gender-responsive methods in an increased number of terms of reference across the region (see Box 6 for a summary of good practice examples).

With the exception of a few reports, the integration of GEEW and human rights was largely limited to the scope or the conduct but most evaluations were not fully gender-responsive in both areas which accounts for the fact that no evaluations were rated as "exceeding requirements".

BOX 7: FEATURED GOOD PRACTICE: GENDER RESPONSIVE EVALUATION

END OF TERM EVALUATION OF THE PROJECT "TOWARDS GENDER EQUALITY: WOMEN'S ECONOMIC EMPOWERMENT HOME-BASED WORKERS, PHASE II, PAKISTAN (2016)

The evaluation report explicitly discusses the extent to which methods chosen are appropriate for analyzing gender and human rights considerations. This includes a discussion of how principles of inclusion and participation were applied and assured. There is also an explanation about how gender equality and human rights are addressed within the objectives and scope of the evaluation.

FINAL EVALUATION OF STRENGTHENING IMPLEMENTATION OF WOMEN, PEACE AND SECURITY AGENDA IN NEPAL PROJECT (2016)

The evaluation report effectively describes how gender equality and human rights were integrated into each phase of the evaluation. Within its scope, the evaluation considered if issues related to gender equality and social inclusion were addressed during various stages of project execution and assessed the availability of gender and ethnicity-based disaggregated data. In terms of applying gender-responsive and human rights-based methods during the conduct of the evaluation, the evaluation team ensured the participation of and consultation with all relevant stakeholders, endeavoured to be gender and culturally sensitive and to respect the confidentiality, protection of source and dignity of those interviewed. The evaluation team was gender balanced and field mobilization was also planned to include both male and female members in the team. Likewise, the team also made efforts to capture voices of men and women alike to the extent possible.

FINAL EVALUATION OF UN WOMEN'S ANTI-HUMAN TRAFFICKING PROGRAMME, UN WOMEN INDIA MCO (2014)

Throughout the report, there are reference to how a gender approach was mainstreamed throughout the whole evaluation process (including the evaluation design, data collection, data analysis and elaboration of conclusions and recommendations). The inclusion of a gender expert helped to foster a GEEW approach throughout the entire evaluation process.

FORMATIVE EVALUATION OF THE PACIFIC EVAW FACILITY FUND, UN WOMEN FIJI MCO (2014)

The use of an evaluation criterion on inclusion makes a strong contribution to the gender responsiveness of this evaluation as within it, GEEW and human rights are explicitly mainstreamed and highlighted. Other good practices include the use of participatory workshops to develop conclusions and strong references to CEDAW and statistical data as well as a discussion on the experience and implications of VAW. GEEW and human rights are mainstreamed throughout the findings and boxes are used to highlight issues from CEDAW. The report also discusses implications of findings for different groups (including men).

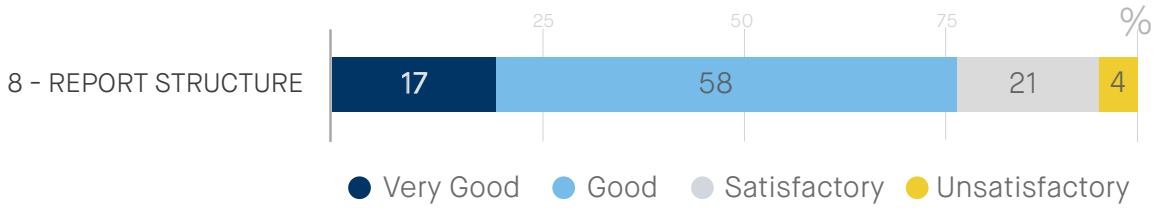
FINAL EVALUATION OF WOMEN LEADERSHIP AND SOCIAL RECONSTRUCTION PROGRAMME, PAKISTAN (2016)

The evaluation adopted the SOAR framework (Strengths, Opportunities, Aspirations, Results) as part of an empowerment approach. SOAR is an innovative strength-based approach to strategic planning that integrates Appreciative Inquiry with a strategic planning framework. The SOAR framework for this particular evaluation was effective as it allowed stakeholders to identify and tap into the project's existing strengths and assets, and then envision potential opportunities for achieving future results and impact.

PARAMETER 8: REPORT STRUCTURE

This parameter area was assessed highly with 75 per cent of reports assessed as "very good" or "good". The majority of reports were structured according to UNEG requirements and in a manner that was clear and logical. Most reports included executive summaries that presented all of the key information required by UNEG standards and were sufficient to stand alone from the report and be used by decision makers although in a few instances, were too lengthy. The use of summary tables in the executive summary and findings by some reports helped to provide key information in an abbreviated and user-friendly manner.

FIGURE 22: RATINGS FOR PARAMETER 8 (2013-2016 REPORTS)



Weaknesses of report structures included omission of a number of key documents from the annexes such as the evaluation matrix, data collection tools and protocols and project logical frameworks. Terms of reference were absent from 38 per cent of reports rendering it difficult for GERAAS reviews and this meta-evaluation to analyse some of the quality parameters such as the extent to which the evaluation managers provided clear requirements to support a gender-responsive evaluation focus and approach. Another area for improvement is to better structure and number findings so that they can be clearly identified and referenced by evaluation users.

BOX 8: FEATURED GOOD PRACTICE: EVALUATION REPORT STRUCTURE

MID-TERM EVALUATION, "STRENGTHENING IMPLEMENTATION OF WOMEN, PEACE AND SECURITY AGENDA IN NEPAL" (2015)

The Executive Summary includes a table clearly and succinctly summarizing the key findings, recommendations and responsible agencies.

FINAL EVALUATION OF EVAW PROGRAMME IN CAMBODIA (2016)

The findings are organized based on the evaluation questions which are embedded throughout the text for ease of reference.

END OF TERM EVALUATION OF THE PROJECT "TOWARDS GENDER EQUALITY: WOMEN'S ECONOMIC EMPOWERMENT HOME-BASED WORKERS, PHASE II, PAKISTAN (2016)

The report annexes contain an exceedingly complete set of information (including case studies, photographic evidence of women's living and working conditions and outcome-specific findings) that increase the credibility of the report.

CONSTRUCTIVE IDEAS TO STRENGTHEN FUTURE EVALUATIONS

1. ENSURE THAT UN WOMEN'S EVALUATIONS RECEIVE HIGH MARKS FOR GENDER-RESPONSIVENESS

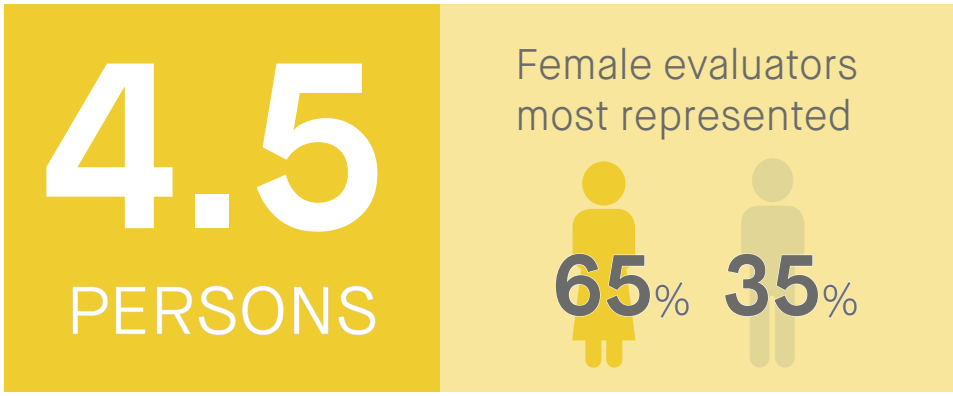
- ☑ Provide good practice reports as an example for evaluators (including those featured in this report).
- ☑ Use evaluation terms of reference to mandate gender-responsive evaluation methods and to require integration of GEEW across all evaluation criteria in addition to an (optional) standalone criterion.
- ☑ Support all M&E Officers and Focal Points, as well as managers and programme staff to successfully complete the e-learning course on gender responsive evaluation.

2. STRENGTHEN EXISTING QUALITY ASSURANCE SYSTEMS

- ☑ Require the Regional Evaluation Specialist to sign off on all inception reports in order to provide an extra layer of quality assurance at the front end, rather than after the report is submitted. If inception reports are strong and aligned with UNEG standards, contain sound and gender-responsive methods and integrate GEEW and human rights across the criteria and questions, they can provide an early indicator about whether quality benchmarks are likely to be met.
- ☑ Ensure the selection of "tried and tested" consultants and companies from a regional roster. Include only consultants and companies on the roster who have conducted evaluations rated "good" or above and ensure that they have experience with gender-responsive evaluation.
- ☑ Develop an example-based checklist for evaluation managers and consultants to use to ensure that the UNEG quality parameters are met, including requirements for gender-responsive evaluation are met (see Annex B for a sample checklist for gender responsive evaluation). Annex the checklist to inception and draft reports and provide quality assurance by requiring review and sign-off by the evaluation manager and consultant when draft deliverables are submitted. As an annex to the checklist, include report templates (based on UN Women and UNEG standards) and a samples of best practice evaluation reports.
- ☑ Within the checklist, give special attention to ensure that:
  - TORs, evaluation matrix and data protocols and tools are included within the report annexes;
  - The evaluation purpose provides deeper insights about why the evaluation was needed and how it will be used;
  - A discussion of ethics is included and that ethical safeguards and measures are implemented to ensure that the evaluation process observes ethical standards;
  - Findings are clearly highlighted and numbered;
  - Conclusions offer deeper insights into the findings;
  - All evaluation reports elaborate the underlying Theory of Change alongside M&E frameworks;
  - Rigorous methods for data collection and analysis are in place that are based on mixed methods and enable triangulation of multiple data sources;
  - A general lessons learned section is included with harvested lessons that are universal with potential for broader application;
  - Template and sample of best practice evaluation report (optional)

3. EXPAND TOOLS AND LEARNING OPPORTUNITIES

- ☑ Support greater in-depth learning exchange by compiling and sharing evaluation good practices on a bi-annual basis. Consider inter-regional exchange opportunities in order to broaden and expand the knowledge pool.
- ☑ Require the inclusion of national evaluation experts in all evaluation teams in order to support increased evaluation capacity in the region.
- ☑ Develop tools to help build skills and knowledge for areas assessed to be weak including:
  - Guidelines on stakeholder mapping and analysis of stakeholder roles and responsibilities;
  - Reconstruction of theories of change and theory-based evaluation approaches;
  - A practical overview of methods to support collection and analysis of quantitative data.


4.5
PERSONS
Female evaluators
most represented
65%
35%

ANNEX A: EVALUATION REPORTS INCLUDED IN META-EVALUATION

No.	Completion Date	Evaluation Title	Evaluation Type	Thematic Area	Country	GERAAS Ratings
1	2011	Final Evaluation of UN Joint Programme to Facilitate the implementation of the CEDAW concluding comments (UNJP-CEDAW) in the Philippines	Final Programme Evaluation	DRF 6 – Global Norms and Standards	Philippines	Very weak/ unsatisfactory
2	2011	Mid-Term Evaluation of the Integrated Programme for Women in Politics and Decision-Making (IPWPDM)	Mid-Term Programme Evaluation	DRF 1 – Political Participation	Timor-Leste	Very weak/ unsatisfactory
3	2011	Programme Evaluation of SAARC Gender Info Base of SAARC-UN Women South Asia Regional Office	Thematic Final Evaluation	Cross-cutting	South Asia	Very Good
4	2012	Evaluation of the Gender Equality and Political Governance Project (GEPG)	Final Programme Evaluation	DRF 5 – Governance and National Planning	Pacific	Excellent/ Very Good
5	2012	Evaluation of UN Women's Work on Gender Responsive Budgeting in India	Thematic Evaluation	DRF 5 – Governance and National Planning	India	Excellent/ Very Good
6	2012	Evaluation of UN Women Support To PWN+ On HIV and AIDS Projects in India	Final Programme Evaluation	Cross-cutting	India	Excellent/ Very Good
7	2012	Evaluation of Supporting Gender Equality and Women's Rights Programme in Timor-Leste	Final Programme Evaluation	Cross-cutting	Timor-Leste	* No rating
8	2012	Evaluation of China Gender Facility for Research and Advocacy	Mid-Term Programme Evaluation	Cross-cutting	China	Excellent/ Very Good
9	2012	Evaluation of the Partnership for Equality and Capacity Enhancement (PEACE): Towards Implementation of UNSCRs 1325 and 1820"Project	Final Programme Evaluation	DRF 4 – Women, Peace & Security	Nepal	Good
10	2013	Final Evaluation of the project "Sustaining the gains of foreign labour migration through the protection of migrant workers' rights in 2012	Final Evaluationl Country-leve	DRF 2 – Economic Empowerment	Nepal	Good
11	2013	Mid-Term Evaluation Of UN Women's Anti-Human Trafficking Programme	Mid-Term Programme Evaluation	DRF 3 - EVAW	India	Unsatisfactory
12	2013	Final Evaluation of the Making Politics Work with Women (MP3W) project	Final Programme Evaluation	DRF 1 – Political Participation	Nepal	Good
13	2013	External Evaluation Of The EVAW Commission Project	Final Programme	DRF 3 - EVAW	Afghanistan	Good
14	2014	Formative Evaluation of the Pacific Ending Violence Against Women (EVAW) Facility Fund	Programme Evaluation	DRF 3 - EVAW	Pacific	Very good
15	2014	Final evaluation of the project on regional mechanisms to protect the human rights of women and girls in South East Asia (ASEAN)	Final Evaluation-Programme Evaluation	DRF 6 – Global Norms & Standards	Regional Office for Asia and the Pacific	Satisfactory
16	2014	Mid-Term Review of Regional Programme on Improving Women's Human Rights in South East Asia, CEDAW SEAP	Mid-Term Programme Evaluation	DRF 6 - Global Norms & Standards	Regional Office for Asia and the Pacific	Satisfactory
17	2014	'Empowering Women in Rural and Informal Settings through Capacity Development' Final Evaluation Report	Final Programme Evaluation	DRF 2 – Economic Empowerment	India	Very Good
18	2014	Final Evaluation UN Women's Anti Human Trafficking Programme	Final Programme Evaluation	DRF 3 – EVAW	India	Very Good
19	2015	Formative Evaluation of Elimination of Violence Against Women Special Fund	Mid-Term Programme Evaluation	DRF 3 – EVAW	Afghanistan	Good
20	2015	Mid-tem evaluation of Strengthening Implementation of Women, Peace and Security Agenda in Nepal (SIWPSAN) project	Mid-Term Evaluation Country-level	DRF 4 – Women, Peace & Security	Nepal	Good
21	2015	Final evaluation of Gender Responsive Recovery for Sustainable Peace (GRRSP) project	Final Programme Evaluation	DRF 4 – Women, Peace & Security	Nepal	Good
22	2015	Final evaluation of the project on "Empowerment of widows and their coalitions programme"	Final Programme Evaluation	DRF 2- Economic Empowerment	South Asia	Good
23	2015	Final Evaluation of the Ford funded Programme on "Partnership with civil society and the Comptroller and Auditor General (CAG) of India to promote gender responsive planning, budgeting and auditing in select States"	Final Programme Evaluation	DRF 5 – Governance and National Planning	South Asia	Satisfactory
24	2016	Final evaluation of Gender and Climate Change Project	Final Project Evaluation	DRF 1, 3 & 5	Bangladesh	Satisfactory
25	2016	Final Evaluation of EVAW Programme	Final Programme Evaluation	DRF 3 – EVAW	Cambodia	Very Good
26	2016	Mid-term Evaluation of the China Gender Fund	Mid-Term Programme Evaluation	DRF 1, 2, 3 & 5	China	Good
27	2016	Final Evaluation of the Regional Programme on improving Women's Human Rights in South East Asia – CEDAW-SEAP	Final Evaluation-Programme Evaluation	DRF 6 – Global Norms & Standards	Regional Office for Asia and the Pacific	Good

No.	Completion Date	Evaluation Title	Evaluation Type	Thematic Area	Country	GERAAS Ratings
28	2016	Final Evaluation of SDC-funded part of Regional Programme on Empowering Women Migrant Workers in Asia (Phase III EWMWA (2012-2015)	Final Evaluation-Programme Evaluation	DRF 2- Economic Empowerment	Regional Office for Asia and the Pacific	Good
29	2016	Endline Evaluation Report Leveraging Technical Tools, Evidence and Community Engagement to Advance the Implementation of Laws and Provision of Services to Women Experiencing Violence in South-East Asia	Final Evaluation-Programme Evaluation	DRF 3 – EVAW	Regional Office for Asia and the Pacific	Good
30	2016	End of Term Evaluation of the Project "Towards Gender Equality Women's Economic Empowerment Home-based Workers, Phase II" (2012-2015)	Meta- Evaluation/ Programme Evaluation	DRF 2 – Economic Empowerment	Pakistan	Very Good
31	2016	Mid-term Evaluation of the Pacific Regional EVAW Facility Fund	Mid Term Evaluation-Thematic Evaluation	DRF 3 – EVAW	Multi-Country Office for the Pacific (Fiji)	Satisfactory
32	2016	Final Evaluation of Strengthening Implementation of Women, Peace and Security Agenda in Nepal Project (SIWPSAIN)	Final Evaluation Country-level	DRF 4 – Women Peace and Security	Nepal	Very Good
33	2016	Final Evaluation of the "Localizing Women, Peace and Security Agenda in Central Terai Districts of Nepal" (GPI-CT)	Joint Country-level Evaluation	DRF 4 – Women Peace and Security	Nepal	Satisfactory
34	2016	Women Leadership and Social Reconstruction Programme (July 2014 – June 2016)	Final Programme Evaluation	DRF 1 & Humanitarian	Pakistan	Good

ANNEX B: PROPOSED ELEMENTS OF A CHECKLIST FOR UN-SWAP EPI CRITERIA

*SOURCE: Barnes Joseph and Bishop, Jo-Anne. Review of UN-SWAP Evaluation Performance Indicator Reporting, Synthesis Report, Final Version (December 2016)

GEEW INCLUDED IN SCOPE/ANALYSIS

- ☐ Evaluation scope: mentions that GEEW will be addressed
- ☐ Evaluation criteria: GEEW integrated across all criteria (ideal) and/or addressed specifically as a stand-along/cross-cutting criterion
- ☐ Evaluation questions: Specific questions/sub-questions included about the extent to which GEEW was addressed in the intervention design/ implementation; alignment of intervention with needs of women, men, girls and boys; and participation of women, men, girls and boys in the intervention
- ☐ Gender disaggregated indicators included and mainstreamed across the Evaluation Matrix
- ☐ Gender analysis reflected in the evaluation background/contextual overview (including reference to relevant organizational policies/directives on GEEW and international GE commitments/standards/treaty body recommendations)

GEEW INCLUDED IN METHODS

- ☐ Sex-disaggregated breakdown of stakeholders/ respondents
- ☐ Stakeholder analysis conducted which breakdowns categories of stakeholders (RH/DB) and their role in the intervention/evaluation process and indicates possible barriers to participation along with mitigating strategies to maximize inclusion
- ☐ Detail about how GEEW considerations addressed in data collection tools (i.e. interview protocols/surveys) and how analytical methods were used to draw out gender considerations
- ☐ Detail about availability/use of GEEW-related documents and data
- ☐ Detail about how ethical considerations were addressed

GEEW INCLUDED IN FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

- ☐ GEEW analysis included and reflected across findings
- ☐ GEEW reflected in the evaluation conclusions
- ☐ Specific recommendations related to GEEW included
- ☐ Lessons learnt and good practices related to GEEW included

ANNEX C: DOCUMENTS CONSULTED

GERAAS-RELATED DOCUMENTS

1. Women GERAAS 2013, Global Evaluation Reports Assessment and Analysis System, Meta-Evaluation Report, 30 March 2015.
2. UN Women Independent Evaluation Office, Global Evaluation Reports Assessment and Analysis System, January 2015.
3. Women GERAAS 2013, Global Evaluation Reports Assessment and Analysis System, Meta-Evaluation Report, 18 March 2014.

UN WOMEN DOCUMENTS

4. Findings for the 2016 UN SWAP EPI assessment for UN Women-managed evaluations
5. UN Women 2015 Scorecard, SWAP EPI 2015 reporting cycle
6. UN Women Fund for Gender Equality, Meta Evaluation Report 2015.
7. UN Women Planning and Programme Coordination Unit, Meta-Analysis of UN Women 2017 Field Annual Work Plans (AWP).

EVALUATION GUIDANCE DOCUMENTS

8. UN SWAP Evaluation Performance Indicator and related Scorecard
9. UNEG Quality Checklist for Evaluation Terms of Reference and Inception Reports
10. UNEG Standards for Evaluation in the UN System 2005
11. UNEG Norms for Evaluation in the UN System
12. UNEG Code of Conduct for Evaluation in the UN System, UNEG
13. UNEG Ethical Guidelines for Evaluation
14. Integrating Human Rights and Gender Equality in Evaluation - Towards UNEG Guidance
15. DAC Principles for Evaluation of Development Assistance

ANNEX D: TERMS OF REFERENCE

Title: Evaluation Consultant to review Evaluation Oversight System (GEOS), Meta-Evaluation and Meta-Analysis (GERAAS) of UN Women Asia and the Pacific in 2010-2015

Location: Home-Based

Contract Duration: 3 months: 1 October – 31 December 2016

Contract Supervision: Regional Evaluation Specialist

Application Deadline: 23 September 2016

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, in Asia and the Pacific region, 20 evaluations get carried out by UN Women Offices each year. Therefore, UN Women Regional Office for Asia and the Pacific is giving increased emphasis to strengthening support for decentralized evaluations conducted by Multi-Country Offices (MCOs) and Country Offices (COs) in the region.

To address the organizational demands for enduring good quality and credible evaluations particularly at decentralized level, the UN Women's Independent Evaluation Office (IEO) has designed a Global Evaluation Oversight System (GEOS) and 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 the oversight system on a quarterly basis and corporate and decentralized evaluations on a yearly basis and made available assessment findings to senior managers, programme unites and the UN Women 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 system 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 for the body of evaluation reports that meet UN Women quality requirements.

UN Women Regional Office for Asia and the Pacific is seeking for a consultant to conduct a review on GEOS and a meta-evaluation and meta-analysis of the evaluation reports produced in the region during 2010-2015, using the GERAAS. The selected consultant will review the GEOS and the GERAAS reports, focusing on the evaluation reports produced by the Asia and the Pacific region, as well as reviewing the evaluation reports, to analyze trends, results, contributions, strengths and weaknesses across UN Women Asia and the Pacific region. It should also synthesize the recurrent findings, recommendations, conclusion and, and lessons learnt for all evaluation reports completed in the given years.

UN Women has the Regional Evaluation Strategy for Asia and the Pacific Region for 2014 -2017 and this review will contribute to reviewing the current Evaluation Strategy.

Duties and Responsibilities

- To examine the quality of the evaluation oversight system in Asia and the Pacific by providing a summary of the GEOS reports to the Senior Managers and staff in the region;
- To examine the quality of the past evaluation reports in Asia and the Pacific by providing a summary of the GERAAS reports to the Senior Managers and staff in the region;
- To analyze the impacts of the UN Women Asia and the Pacific's programmes/projects;
- To provide better understandings and insights into UN Women Asia and the Pacific's performance;
- To capture experiences and lessons learnt from the evaluations to enhance organizational learning and knowledge management.

Key questions

The review is guided by the following core questions:

Oversight system (GEOS):
- How is the situation of the human resources in the field of ME in the region?
- How is financial resources related to evaluation in the region? Any changes?
- How is evaluation coverage by country?
- How is the evaluation implementation rate?
- How is the evaluation quality? (this overlaps with the meta-evaluation below)
- Are Management Responses available in the GATE system? How are they managed?
- How are the implementation of the key action points in the GATE system?
- How are the use of the evaluations?
- How many people in the region have undertook the evaluation e-learning course? Is the number increasing?

Meta-evaluation (GERAAS):
- How many evaluations have been conducted by UN Women's Asia and the Pacific in 2010-2015?
- What are their thematic areas, evaluation types, and countries? Any trends?
- is there any missing evaluations in terms of thematic areas, types, and countries?
- How were they rated by the GERAAS? Any improvements in the last 5 years?
- What are the strengths and weaknesses of the evaluation reports?

Meta-analysis (GERAAS):
- What are the strengths that emerge from the evaluations of UN Women Asia and the Pacific's programmes/projects?
- Which types of efforts/strategies being implemented have shown high degrees of relevance, effectiveness, sustainability, efficiency and impact?
- What factors have contributed to this or inhibited success?
- Are there any patterns and lessons to be learnt regarding results produced by UN Women Asia and the Pacific's programmes in general?
- Are there findings and conclusions that point in the same direction?
- What strengths and challenges do the evaluations expose?
- What are lessons learnt?

Methodology

The evaluation will have **three phases/components:**

Phase 1 – Oversight system: Review the GEOS reports since 2013. The purpose of reviewing the oversight system is to analyze how UN Women Asia and the Pacific has been ensuring to produce quality evaluation reports. To produce quality evaluation reports, human and financial resources are inevitable, and this review should examine this. It is already knows that the low evaluation implementation rate is a challenge/weakness in UN Women. The review should summarize the implementation rate in the past and discuss the solutions. It should also examine the GATE system, especially the management responses and their implementations.

Part 2 - Meta-Evaluation - ME: Review the 30 evaluation reports produced from 2010 until 2015 (See Annex 1) that will be part of this exercise and undertake a meta-evaluation of these.

The purpose of the ME is to capture the quality of evaluation reports. 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 UN Women Asia and the Pacific. This phase is mainly designed to strengthen UN Women Asia and the Pacific's evaluation capacity by providing practical recommendations to improve future programmes and projects.
This will be done based on UN Women's past GERAAS reports and UNEG standards.

Phase 3 – Meta-Analysis - MA: Evaluation Reports that are found to be "satisfactory or above" (using GERAAS), will be selected to take part in the Meta-Analysis. The MA aggregates the recurrent findings, conclusions, lessons learned, good practices and recommendations that have

come out of the evaluations. The Meta-Analysis is poised to provide a basis to better understand UN Women Asia and the Pacific's programme interventions around the UNEG criteria (relevance, effectiveness, efficiency, sustainability and impact). It also provides further analysis on the progress made against UN Women's organizational goals and priorities (and strategies).

The synthesis of this information will support the use of evaluation findings by UN Women Asia and the Pacific as well as stakeholders. It will also inform donors and other development partners about the effectiveness of the interventions supported by UN Women Asia and the Pacific in the last 5 years.

Usability will be ensured through different strategies, including tailoring of the deliverables to ensure that the results of the Oversight system, ME and MA are captured in a way that stimulates sharing and understanding of knowledge.

Expected Deliverables
The four main expected deliverables of the consultancy will be[27]:

1- Inception Report		
This report will be completed after initial desk review of program documents. It will be 7 pages maximum in length and will include: • Introduction • Background to the evaluation: objectives and overall approach of both phases of this exercise. • Identification of evaluation scope (see points above regarding reports to be included and excluded). • Description of evaluation methodology/methodological approach. • Work Plan with dates and deliverables.	Expected Duration: 1st week	Management Notes: • This report will be used as an initial point of agreement and understanding between the consultant and the Evaluation Manager (the Regional Evaluation Specialist) & Reference Groups. • Payment: 10% of total on approval of deliverable.
2- Oversight Report		
A 5-10 page Oversight report with findings based on GEOS reports.	Expected Duration: 2nd -6th weeks	Management Notes: • This will be reviewed by the Evaluation Manager (the Regional Evaluation Specialist). • Payment: 10% of total on approval of deliverable.

[27] Please see Annex for additional information on reports. Please note that the UN Women Evaluation checklist for reports will be shared with the selected evaluation experts as will all other tools as per UN Women Evaluation Handbook. http://genderevaluation.unwomen.org/en

3- Meta-Evaluation Report		
An 8-10 page Meta-Evaluation report with findings and assessment based on UNEG and GERAAS standards, all ER will be qualified: [Very Good, Good, Satisfactory, or Unsatisfactory] and key concrete recommendations for FGE on evaluation and RBM will be provided. It will include: • Executive Summary (1 page) • Assessment/systematization of key strengths and weaknesses observed/identified and lessons learned and good practices emerging from the review of the 25 FGE evaluation reports. (5 pages) • A set of concrete recommendations for UN Women FGE to improve evaluation based on RBM programme management. (2-3 pages) • Annex: Scoring of the quality of the 25 ER using GERAAS tool (tailored to this assignment as needed) using the Quality Review Template (GERAAS Annex III) of each of the 25 reports.	Expected Duration: 2th -8th Weeks	Management Notes: • The final Meta-Evaluation report will go through a process of review and approval by the Reference Group. Once approved Phase 2 can begin with selected Evaluation Reports. • Payment: 40% of total on approval of deliverable.

4- Meta-Analysis Report		
This report will analyze and synthesize what are some of the findings, conclusions, recommendations and lessons learned and best practices from selected programme evaluations undertaken worldwide. It will be a 25 page report that will help to paint a perspective of the achievement of UN Women Asia and the Pacific's programmes and projects vis-à-vis its organizational goals. The MA is poised to provide a basis to better understand UN Women Asia and the Pacific's interventions and achievement of its goal and outcomes. The content of the report will be discussed with the consultant in detail prior to the MA phase starting – after the ME is finalized and there is a clearer picture of what programme ER will be part of the MA. However, it is important to note that this report will include 4 knowledge pieces produced by the consultant: this can be, for example, 2-page document on strategies that work per Outcome or an analysis of trends per region, etc.	Expected Duration: 9th -11th Weeks	Management Notes: The final Meta-Analysis report will go through a process of review and approval. Final approval of findings will be done by the Regional Director. Payment: 40% of total on approval of deliverable.

Work Plan

	Weeks										
	1	2	3	4	5	6	7	8	9	10	11
Inception Report	X										
Oversight System – Desk review		X	X								
ME Phase – Desk Review		X	X	X	X						
Draft ME Report submitted						X					
Validation of ME Report Final ME Report submitted.					X	X					
MA Phase - Desk Review		X	X	X	X	X	X	X	X		
Draft MA Report submitted									X		
Validation of MA Report										X	
Final MA Report submitted											X

Evaluation Management

Who: Actors and Accountability	What: Roles and Responsibilities
Reference Group	• Receive information throughout the entire evaluation process and participating in relevant meetings at strategic points during the evaluation. • Review key evaluation deliverables such as the Inception Report and Draft Final Report • Provide input on these evaluation deliverables as needed • Support dissemination of the findings and recommendations.
UN Women Evaluation Task Manager (Regional Evaluation Specialist)	▪ Ensure the quality of evaluation and management decisions to be made on time. Facilitate selection of the consulting firm ▪ Facilitate communication between the consultant and the reference group ▪ Monitor the process of review and provide guidance to the consultant ▪ Report any significant deviation from the evaluation plan ▪ Facilitate the preparation, conduct and report finalization. ▪ Facilitate a management response to all evaluation recommendations and ensure the implementation of committed actions in the management response ▪ Facilitate dissemination of initial and final evaluation findings to relevant stakeholders

Skills and Competencies

Position	Education	Professional Experience	Skills
Evaluation Consultant	A Masters or higher level degree in International Development or a similar field related to political and economic development, monitoring and evaluation, etc.	A minimum of 10 years relevant experience undertaking evaluations is required including proven practical professional experience in designing and conducting major evaluations. Substantive experience in evaluating similar development projects related to local development and political and economic empowerment of women. Substantive experience in evaluating projects and programmes with a strong gender focus is preferred. Excellent and proven knowledge of evaluation methodologies and approaches. Experience with meta-evaluation and meta-analysis of evaluation reports, preferably with UN agencies, is an asset. Proven experience in producing coherent, clear analytic reports and knowledge pieces is a requirement.	Excellent English writing and communication and analytical skills are required.

Evaluation ethics

Evaluations in the UN will be conducted in accordance with the principles outlined in both UNEG Norms and Standards for Evaluation in the UN System and by the UNEG 'Ethical Guidelines for Evaluation'. These documents will be attached to the contract. Evaluators are required to read the Norms and Standards and the guidelines and ensure a strict adherence to it, including establishing protocols to safeguard confidentiality of information obtained during the evaluation.

- Anonymity and confidentiality. The evaluation must respect the rights of individuals who provide information, ensuring their anonymity and confidentiality.
- Responsibility. The report must mention any dispute or difference of opinion that may have arisen among the consultants or between the consultant and the heads of the Programme in connection with the findings and/or recommendations. The team must corroborate all assertions, or disagreement with them noted.
- Integrity. The evaluator will be responsible for highlighting issues not specifically mentioned in the TOR, if this is needed to obtain a more complete analysis of the intervention.
- Independence. The consultant should ensure his or her independence from the intervention under review, and he or she must not be associated with its management or any element thereof.
- Incidents. If problems arise at any other stage of the evaluation, they must be reported immediately to the manager of the evaluation. If this is not done, the existence of such problems may in no case be used to justify the failure to obtain the results stipulated in these terms of reference.
- Validation of information. The consultant will be responsible for ensuring the accuracy of the information collected while preparing the reports and will be ultimately responsible for the information presented in the evaluation report.
- Intellectual property. In handling information sources, the consultant shall respect the intellectual property rights of the institutions and communities that are under review.
- Delivery of reports. If delivery of the reports is delayed, or in the event that the quality of the reports delivered is clearly lower than what was agreed, the penalties stipulated in these terms of reference will be applicable.

HOW TO APPLY

Interested candidates are requested to submit electronic application to apirada.khachonpan@unwomen.org and hr.bangkok@unwomen.org not later than **[date]**

Submission of Package

- CV
- Letter of Interest containing the statement on candidate's relevant experience
- Financial proposal. The financial proposal shall specify a total lump sum amount breaking down **proposed number of working days** and a daily professional fee.

All applications must include (as an attachment) the CV and the financial proposal. Applications without financial proposal will be treated as incomplete and will not be considered for further assessment.

Please note that only candidates selected for an interview will be contacted.

ANNEX E: BIOGRAPHY OF EVALUATOR

Jo-Anne Bishop is a gender and human rights expert with senior leadership experience in results-based programme management and strategy review and development.

Jo-Anne has 15 years of experience supporting and advising governments, national institutions and intergovernmental organizations in the areas of human rights, gender equality, gender mainstreaming and non-discrimination in a number of countries including Afghanistan, Bosnia and Herzegovina, Liberia and Timor-Leste.

She has held senior positions as Head of Department for the OSCE Office for Democratic Institutions and Human Rights, Director of the Canadian Governance Support Office in Afghanistan, Advisor to the Liberian Governance Commission and Advisor to the Secretary of State for the Promotion of Equality in Timor-Leste. Her experience also includes work with UN Women, UNFPA, UNDP, IOM and the Afghanistan Independent Human Rights Commission to develop strategic plans and lead reviews and evaluations at a meta, global, regional and country-level.

In addition to serving as a team leader for a global evaluation of women's economic empowerment, Jo-Anne has led regional evaluations on gender responsive budgeting and recently conducted a review of the UN System-Wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP) Evaluation Performance Indicator Reporting for the UN Evaluation Group. She also conducted a meta-evaluation and meta-analysis of the UN Women Fund for Gender Equality

UN WOMEN

UN WomenAsia and the Pacific