



Review of UN SWAP Evaluation Performance Indicator Reporting

SYNTHESIS REPORT

Final Version (12 December 2016)

Prepared for Review Management Group, UNEG Working Group on Gender Equality and Human Rights

Laurence Reichel (UNICEF); Ekaterina Sediakina-Rivière (UNESCO); Barbara Torggler (UNESCO); Sabas Monroy (OHCHR); Elisa Calcaterra (UNEP); Daniel Chen (UNCTAD);

Submitted to UN Women

Sabrina Evangelista, Task Manager

Prepared by ImpactReady LLP

Joseph Barnes
Jo-Anne Bishop

www.ImpactReady.org

Professional Partnership for Social Value
Registered in England and Wales: OC370678

List of Acronyms

CEDAW	Convention on the Elimination of Discrimination Against Women
DSS	Department for Safety and Security
ECLAC	Economic Commission for Latin America and the Caribbean
EPI	Evaluation Performance Indicator
ESCAP	Economic and Social Commission for Asia and the Pacific
ESCWA	Economic and Social Commission for Western Asia
GE	Gender Equality
GEF	Global Environment Facility
GEEW	Gender Equality and the Empowerment of Women
GERAAS	UN Women Global Evaluation Reports Assessment and Analysis System
GEROS	UNICEF Global Evaluation Reports Oversight System
HQ	Headquarters
HR	Human Rights
HRBA	Human Rights-Based Approach
IFAD	International Fund for Agricultural Development
ILO	International Labour Organisation
IOM	International Organisation for Migration
ITC	International Trade Centre
MLLE	Multiple Lines and Levels of Evidence
OECD-DAC	Organisation for Economic Co-operation and Development's Development Assistance Committee
OHCA	Office for the Coordination of Humanitarian Affairs
OHCHR	Office of the High Commissioner for Human Rights
PLE	Peer Learning Exchange
SWAP	System-Wide Action Plan
TOR	Terms of Reference
UN	United Nations
UNCDF	United Nations Capital Development Fund
UNCTAD	United Nations Conference on Trade and Development
UNDP	United Nations Development Programme
UNFAO	United Nations Food and Agriculture Organisation
UNFPA	United Nations Population Fund
UNEG	United Nations Evaluation Group
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
UNV	United Nations Volunteer Programme
UN WOMEN	United Nations Entity for Gender Equality and the Empowerment of Women
WFP	World Food Programme
WHO	World Health Organisation
WIPO	World Intellectual Property Organisation

Table of Contents

List of Acronyms..... 2

1. Introduction 4

 1.1 Background..... 4

 1.2 Review Purpose and Objectives..... 6

 1.3 Review Process, Design and Methods 7

 1.4 Sampling 8

2. Analysis 10

 2.1 Trends in UN-SWAP EPI ratings by criteria 10

 2.2 Trends in UN-SWAP EPI Ratings by Type of Review..... 18

 2.3 Trends in UN-SWAP EPI Ratings by Type of Entity and Evaluation..... 20

 2.4 Challenges and Good Practices in Meeting the UN-SWAP EPI 23

3. Recommended Actions 23

Annexes

Annex A: Terms of Reference 27

Annex B: List of Evaluation Reports Assessed 32

Annex C: Stakeholder Consultation List 34

Annex D: Documents Consulted 35

Annex E: Biographies of Evaluation Team Members 36

Annex F: Addendum-Plan of Action UNEG-SO3 37

1. Introduction

1.1 Background

5 UN System Wide Action Plan on Gender Equality and the Empowerment of Women

- The UN System Wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP) constitutes the first accountability framework for gender mainstreaming in the UN system.¹ The UN-SWAP is composed of 15 performance indicators for tracking six main elements on gender mainstreaming: accountability, results based management, oversight, human and financial resources, capacity, and knowledge exchange and networking. The UN-SWAP framework is accompanied by a set of Technical Notes for each Performance Indicator that provide guidance on how to complete the assessment for each of the 15 Performance Indicators.
- 10
- 15 All UN entities are to self-assess and report on progress. UN entities are expected to meet all UN-SWAP-performance standards by 2017.² Reporting on the UN-SWAP commenced in 2013 and entities are expected to report on a yearly basis through the Report of the Secretary-General to ECOSOC on “Mainstreaming a gender perspective into all policies and programmes in the UN system”³. In 2015, UN Women initiated the process for revising and updating the UN-SWAP framework, as originally envisioned, to be rolled out in 2018.⁴
- 20

UN-SWAP Evaluation Performance Indicator (EPI)

- 25 The oversight element of the UN-SWAP is composed of three performance indicators, including one dedicated to evaluation that is linked to meeting the gender-related UNEG Norms⁵ & Standards⁶ and demonstrating effective use of the UNEG guidance on integrating human rights and gender equality in evaluation⁷.

¹ UN-SWAP Framework was developed by UN Women in 2011/2012 in response to the CEB endorsed UN system-wide Policy on Gender Equality and the Empowerment of Women (CEB/2006/2), which was established based on the ECOSOC Agreed Conclusions 1997/2. On 13 April 2012, the CEB endorsed the UN-SWAP for application throughout the UN system. In its resolution E/RES/2012/24 of 27 July 2012, the ECOSOC welcomed the UN-SWAP and called upon the UN system to actively engage in its roll-out and report on the implementation of the resolution at its substantive session in 2013.

² There is an extended timeframe to 2019 for those entities with a mainly technical focus

³ For example see, United Nations, “Mainstreaming a gender perspective into all policies and programmes in the UN system,” April 2014; accessible online at: http://www.un.org/ga/search/view_doc.asp?symbol=E/2014/63&Lang=E

⁴ Plans for the roll-out of UN-SWAP/2 are still under consideration.

⁵ United Nations Evaluation Group, Norms for Evaluation in the UN system, 2005; accessible online: <http://unevaluation.org/document/detail/21>

⁶ United Nations Evaluation Group, Standards for Evaluation in the UN system, 2005; accessible online: <http://unevaluation.org/document/detail/22>

⁷ United Nations Evaluation Group, Guidance on Integrating Human Rights and Gender Equality in Evaluation, 2014; accessible online: <http://unevaluation.org/document/detail/1616>

The reporting categories for the Evaluation Performance Indicator are as follows:

Not Applicable	Missing	Approaches	Meets	Exceeds
5a. Performance indicator is not relevant to a UN entity	5b. None of the UNEG gender-related norms and standards are met	5c. Meets some of the UNEG gender-related norms and standards	5d. Meets the UNEG gender-related norms and standards	5ei. Meets the UNEG gender-related norms and standards and 5eii. Demonstrates effective use of the UNEG guidance on evaluating from a human rights and gender equality perspective

30

The UNEG Working Group on Gender Equality and Human Rights developed the Technical Note and Scorecard, which aims to support more systematic and harmonized reporting through the use of a common tool that allows for improved comparability across UN entities. The unit of analysis selected as most feasible was the evaluation report. Thus the Evaluation Performance Indicator should be solely based on an assessment of evaluation reports completed in the reporting year. The Technical Note specifies the below criteria for the assessment of integration of gender equality in the evaluation reports (see Table 1):

35

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.

40

Under the Technical Note, it is recommended that evaluation units conduct an external review, however, as this requires financial resources, the UNEG Working Group on Gender Equality and Human Rights piloted a Peer Learning Exchange (PLE) process in 2015, which proved successful. At a minimum, UN entities are highly encouraged to submit the Scorecard to allow for comparability.

45

1.2 Review Purpose and Objectives

50 Several factors led to the UNEG Working Group on Gender Equality and Human Rights
commissioning this review. Following the conclusion of two UN-SWAP EPI reporting cycles using
the same scorecard approach, there is now a wealth of information from the evaluation reports
scored by UN entities. With only one year to go until the 2017 UN-SWAP deadline, the UN-
SWAP EPI 2015 reporting cycle results identified the need for a deeper analysis of UN-SWAP
55 EPI reporting as a means for better targeting support and facilitating a shared understanding
of what it means to meet UNEG Norms and Standards on gender equality.⁸ There is also a
possibility to revise the Evaluation Performance Indicator for the next cycle of UN-SWAP and
thus the corresponding criteria (as outlined in the Scorecard) used to assess performance.⁹

60 With these factors in mind, and in accordance with the Terms of Reference (TOR), the overall
purpose of the review is:

- 65 ➔ To contribute to a common understanding of what it means to ‘meet’ and ‘exceed’
requirements for the UN-SWAP Evaluation Performance Indicator and thus “Meet the
UNEG gender-related norms and standards for evaluation”;
- 70 ➔ To identify where targeted support for UNEG members is required in order to “meet
requirements” for the UN-SWAP Evaluation Performance Indicator, and;
- 75 ➔ To identify whether it is necessary to revise the current Scorecard criteria and
corresponding Technical Note.¹⁰

The objectives of the review, as set out in the review Terms of Reference, are to:

- 70 ➔ Identify trends in reporting against the UN-SWAP EPI scorecard criteria;
- 75 ➔ Identify good practice in integrating gender equality in evaluation reports;
- 80 ➔ Identify challenges to integrating gender equality in evaluation reports;
- 85 ➔ Suggest actions for supporting UNEG members to meet UN-SWAP EPI requirements;
- 90 ➔ Suggest possible revisions to the Scorecard criteria.

⁸ The report can be accessed on the UNEG website: <http://www.uneval.org/document/download/2536>

⁹ Terms of Reference, p. 3.

¹⁰ Terms of Reference, Review of UN-SWAP Evaluation Performance Indicator Reporting, Commissioned by the UNEG Working Group on Gender Equality and Human Rights, p. 3.

1.3 Review Process, Design and Methods

80 The review process was designed to be consultative and inclusive with active participation of members of the UN Working Group on Gender Equality and Human Rights in the design of the TOR and in reviewing and commenting on the review deliverables. UN-SWAP EPI Focal Points were able to learn about and discuss the emerging review findings during a UNEG Webinar held on 25 October 2016.

85 Although evaluative methods were applied, the process was a review and not an evaluation (as defined by UNEG standards). The review applied a multiple lines and levels of evidence (MLLE) design to triangulate evidence and develop answers to the review questions (see below).

Table 2: List of UN-SWAP EPI Review Questions

Unit of Analysis	Question	Means of analysis
Scorecard	What is the common understanding of how to apply the EPI scorecard criteria? What are the common factors (trends) by scorecard criteria?	Based on the sample of entities and analysis of the comments section of the scorecard to identify thematic trends within each overall reporting category and by scorecard criteria.
Evaluation reports	What are the common characteristics (trends) by scorecard criteria for those evaluation reports that obtained an overall “exceeds” or “meets”?	Based on the sample of entities and identification of thematic trends within each overall entity rating category and by scorecard criteria.
Evaluation reports	What are the good practices within each scorecard criteria from those evaluation reports that scored overall “meets” or “exceeds”?	Same as above.
Evaluation reports	What are the common challenges within each scorecard criteria from those evaluation reports that scored overall “meets” or “exceeds”?	Same as above.
Evaluation reports and Scorecard	To what extent does the external reviewer’s score align with the original score?	Based on a blind sample from each overall reporting category, score evaluation reports.

90 As part of the review, a sample of 46 evaluation reports included in the UN-SWAP EPI were independently assessed and categorised according to approaches to gender responsiveness used in evaluation reports (see Section 1.4 on Sampling for more detail). Each evaluation report was assessed as part of the review based on a consistent application of the UN SWAP indicators, with additional meta data and qualitative and quantitative data captured.

95 Qualitative and quantitative data from Independent Review was then contrasted with secondary documentary evidence from the UN SWAP scorecards and supplementary primary evidence based on key informant interviews with five UN SWAP Focal Points and survey data from 38 UN SWAP Focal Point respondents. Information and data was triangulated based on frequency and comparative analysis of report ratings and averages. Qualitative analysis and meta-synthesis was used to draw out overarching trends and identify common challenges and good practice. Although not part of the TOR, where possible, the review team also reviewed data sources outside the scope of the review including evaluation terms of reference and evaluation and gender equality policies and strategies. In analysing the data, the review team applied

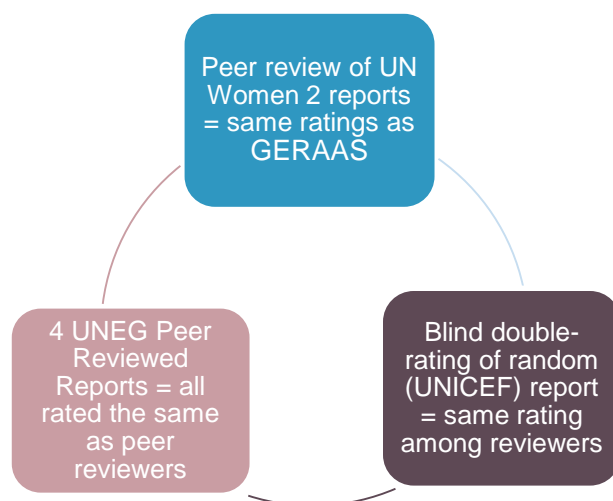
100

105 coding based on the review questions and entity characteristics, with inductive coding of major issues.

Quality assurance of analysis was achieved through the following measures:

- 110
- ➔ In the review of evaluation reports, the review team applied the UNICEF Global Evaluation Reports Oversight System (GEROS) and UN Women Global Evaluation Reports Assessment and Analysis System (GERAAS) system standards.
 - ➔ Quality of and consistency in review ratings, was assured by the fact that: a) the review team’s blind rating of two UN Women reports aligned with the GERAAS rating; b) four of the review team’s ratings aligned with those of the Peer Learning Exchange; c) in a blind double-rating of random UNICEF reports, the review team members arrived at the same rating.
- 115

Figure 1: Triangulating a common standard



120 1.4 Sampling

In order to narrow the universe of evaluation reports to be sampled (approximately 378 were scored in 2015), the universe was limited to those reports that were scored “meets” or “exceeds” requirements in 2015 for a total universe of 161 reports for 25 UN entities including the Secretariat (N=16), Funds and Programmes (N=22) and Specialised Agencies (N=8).

125

Figure 2: Entities included in the review

Secretariat	Funds and Programmes	Technical and Specialised Agencies
<ul style="list-style-type: none"> • ECLAC • ESCAP • ESCWA • DSS • OCHA • OHCHR • OIOS • UNCTAD • UN-HABITAT 	<ul style="list-style-type: none"> • GEF • IFAD • ITC • IOM • UN WOMEN • UNCDF • UNDP • UNFPA • UNICEF • UNV • WFP 	<ul style="list-style-type: none"> • FAO • ILO • UNESCO • WHO • WIPO

* Bolded entities indicate those where only one report was reviewed

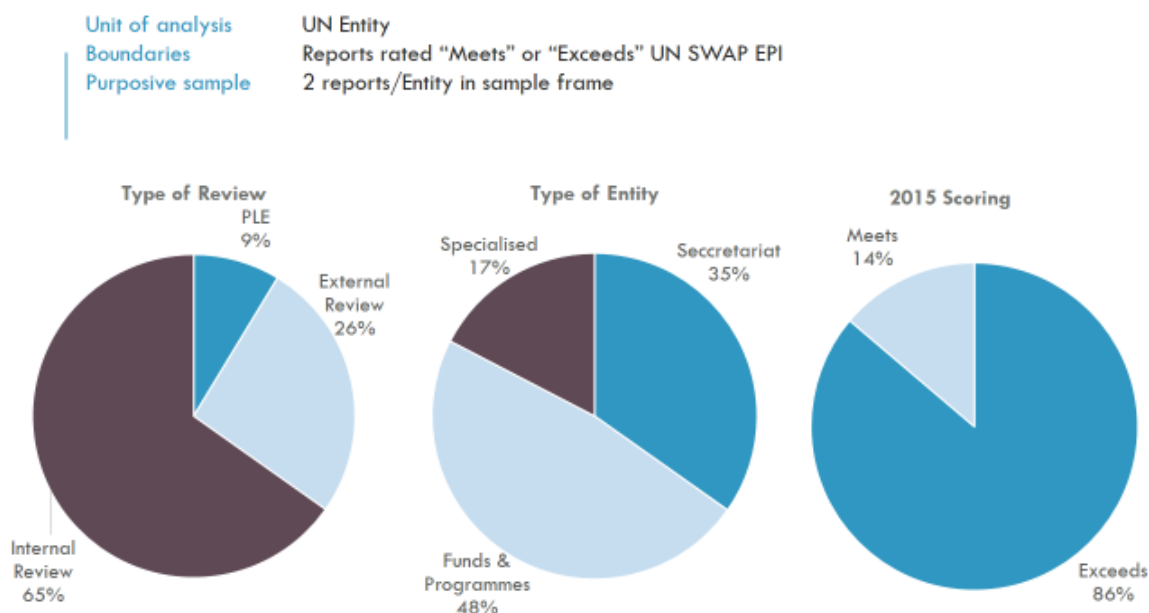
To further reduce the sample size, two reports were included for each entity and where an entity did not have two reports available, only one report was assessed which resulted in a final sample size of 46 reports.¹¹

130

The final sample of 46 entity evaluation reports included a mix of reports scored as ‘meets’ (N=26) and ‘exceeds’ (N=20) as well as a range of reports externally reviewed by outside companies or independent consultants (N=12), reports internally reviewed (N=30) and reports assessed through the Peer Learning Exchange of the UNEG Working Group on Gender Equality and Human Rights (N=4). The selection of reports aimed broadly to maximise coverage of thematic areas across the whole sample.

135

Figure 3: Breakdown of Ratings by Type of Review, Type of Entity and 2015 Scoring



140

The final sample of entities was stratified by type of entity (see table in Annex B) and overall entity rating (exceeds, meets) to identify trends for these categories. 23 entities were included.

¹¹ For the following entities, only one report was reviewed: DSS, GEF, UN Habitat and UNV.

2. Analysis

145

2.1 Trends in UN-SWAP EPI ratings by criteria

This section looks at the questions:

- 150 ➔ What is the common understanding of how to apply the EPI scorecard criteria?
- 150 ➔ What are the common characteristics (trends) by scorecard criteria for those evaluation reports that obtained an overall “exceeds” or “meets”?
- ➔ What are the good practices within each scorecard criteria from evaluation reports that scored overall “meets” or “exceeds”?

155 Within each criterion, the average rating for the 46 reports assessed by the independent review team is included. The rating is based on the EPI indicators using a four point scale (0-3) based on the corresponding numbers: 0 = Missing; 1 = Approaches requirements; 2 = Meets requirements; and 3 = Exceeds requirements.

160 **Criterion 1: Scope and Indicators (Independent Rating = 2.0)**

“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”

165 **Finding 1: On average, the whole body of “Meets” and “Exceeds” evaluations only just meets the required standard for evaluation scope and indicators and there is an inconsistent understanding about how to integrate GEEW in evaluation scope of analysis.**

170 For criterion 1, there is a very diverse understanding among external evaluators and UN agencies about what it means to meet the criterion and significant inconsistencies in application of this criterion were evident from the review of reports. Whilst some evaluations examined gender equality and the empowerment of women (GEEW) from the perspective of women’s participation in the intervention (as beneficiaries), a much more limited number of evaluations examined how GEEW was addressed and analysed during the design and implementation of interventions.

180 Although a number of evaluations that rated as ‘exceeds’ were able to examine the design phase of interventions, very few reports included an analysis of the extent to which GEEW was addressed in the project results framework or theory of change. Reports also did not consistently assess the extent to which gender issues and underlying causes of inequality, were included in situational/contextual analyses. A good practice example was the UNCDF Youthstart evaluation which contained a specific section examining the extent to which a gender focus was part of the project design in terms of result targets and indicators, the extent to which gender analysis was included in market studies commissioned during the intervention and the existence of a strategy to reach out to female clients. The evaluation also assessed the extent to which the programme contributed to increasing the capacity of Financial Service Provider’s to meet the specific needs of girls and young women.

190 In terms of assessing GEEW implications, most reports considered outputs or outcomes only where gender was explicitly addressed. Generally, for interventions where GEEW was not an explicit focus and sex-disaggregated data was not readily available, evaluations tended not to draw

upon established theory to discuss the likely gender considerations or how GEEW could have been better integrated. Whilst the object of the evaluation was gender related, one good practice example of applying theory to undertake gender analysis was UNCTAD's 'External Evaluation of the Development Account Project 1011Q Enhancing Capacities of Developing Countries to Mainstream Gender in Trade Policy'. This evaluation included comprehensive analysis of how the project outputs resulted in changes in capacities and policies and supported integration of GEEW in trade issues.

Across the reports, there was limited discussion of gender equality and human rights standards and the extent to which the design of interventions supported implementation of relevant standards such as CEDAW and national policies and strategies related to GEEW. One good practice example of this was the UN Women ICT evaluation in Jordan which included analysis of interventions for specific CEDAW articles. A very limited number of evaluations included an analysis of relevant organizational or UN system-wide objectives or strategies on GEEW and their relevance to the object of the evaluation despite this being a low-hanging fruit, including for entities whose mandate is not gender-specific. An example of an evaluation where this was done effectively was the WFP mid-term Operational Evaluation "Iran PRRO 200310 Food Assistance and Education Incentives for Afghan and Iraqi Refugees (2013-2015)" where the report analysed gender equality objectives set out in the project document and the extent to which they were consistent with the 2009 WFP Gender Policy.

Although the inclusion of sex-disaggregated data was missing from most evaluation reports, in instances where such data existed, evaluations often did not go beyond number counting to discuss the extent of women and girl's participation and how interventions addressed (or not) their needs and underlying causes of gender inequalities.

In terms of inclusion of GEEW indicators, how these were used and applied was also inconsistent across the reports. Most of the GEEW-related indicators included in the Evaluation Matrices were specific to gender-related questions and there was inadequate mainstreaming of GEEW across other indicator areas in most reports. Good practice exceptions to this were the UNCDF evaluation of the MicroLead programme where gender-related indicators were mainstreamed across all questions including those unrelated to gender. Another noteworthy good practice was the WFP mid-term Operational Evaluation of Food Assistance and Education Incentives for Afghan and Iraqi Refugees which included a mix of qualitative and quantitative evaluation indicators for GEEW.

Where GEEW was not an explicit focus, some evaluations looked more broadly at how empowerment approaches were applied, how human rights was mainstreamed or how interventions contributed to greater accountability towards specific populations (for example humanitarian response-related interventions) or how they addressed issues of social cohesion. A good practice of an evaluation addressing the transversal issues of gender, equity and human rights was the WHO Evaluation of DFATD-funded Project Accelerating Nutrition Improvements in Sub-Saharan Africa.

Finding 2: Criterion 1 in its current formulation provides room for subjective interpretation by reviewers and resulted in the greatest discrepancies in ratings with 33 per cent of reports being re-classified as 'approaching' or 'missing' requirements.

With the exception of examining the extent to which gender indicators address GEEW, the review provided evidence of significant discrepancies in how this criterion was interpreted.

Although the criterion requires the evaluation to analyse how GEEW objectives and GEEW mainstreaming principles were included in the intervention design and how GEEW results were achieved, one third of evaluations rated by the independent review did not meet this criterion.

245

33 per cent of the internally-reviewed reports that were rated as ‘exceeding’ or ‘meeting’ requirements (N=15) were re-classified by the independent review as either only ‘approaching’ or ‘missing’ requirements for Criterion 1.¹²

250

One of the reasons for this discrepancy relates to the lack of clear guidance in interpreting this criterion which meant that for a number of evaluations, inclusion of GEEW in the evaluation scope was rated highly even though the reports failed to provide any detailed analysis explaining how GEEW was integrated in the intervention design and results. In this regard, the scoring for Criterion 1 appeared to be based on subjective assessments and further guidance is therefore

255

needed to provide evaluators with the precise elements that need to be addressed in this criterion.

Common characteristics of reports that exceeded Criteria 1:

- ➡ A comprehensive gender analysis was included in the context section
- ➡ The evaluation analysed GEEW in relation to the intervention design, implementation and results
- ➡ The evaluation considered the gender mainstreaming approach of project, whether it was guided by organizational/system-wide objectives on GEEW and what measures were taken to ensure participation of women and the most marginalized and discriminated against groups
- ➡ Gender-responsive indicators were included in evaluation matrix

260

265

Criterion 2: Criteria and Questions (Independent Rating = 2.0)

270

“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.”

275

Finding 3: There is a general understanding about how to integrate GEEW in evaluation criteria and questions although approaches in applying the criterion are somewhat inconsistent and uneven; 26 per cent of evaluation reports were re-classified as ‘approaching’ or ‘missing’ requirements.

280

Most UN agencies with evaluation reports rated ‘exceeds’ or ‘meets’ were able to explicitly integrate GEEW across at least one or two criteria and/or questions although discrepancies were evident based on the review type. For reports internally assessed, reports were often rated as ‘exceeds’ in instances where GEEW was included in one question and not any of the criterion; whereas for the independently reviewed reports, a rating of ‘exceeds’ was only given in instances where GEEW was integrated in multiple criterion and questions.

285

26 per cent of the reports that were originally rated as ‘exceeding’ or ‘meeting’ requirements (N=12) were re-classified by the independent review as either only ‘approaching’ or ‘missing’ requirements for Criterion 2.¹³

¹² Of the 15 reports re-classified, 13 were rated as ‘approaches requirements’ and 2 as ‘missing requirements’.
¹³ Of the 12 reports re-classified, 10 were rated as ‘approaches requirements’ and 2 as ‘missing requirements’.

290 The most common criteria for integrating gender was ‘effectiveness’ followed by ‘relevance’. A
 very limited number of evaluation reports were able to include GEEW within the criteria of
 efficiency with the exception of the UN Women Evaluation of the Joint Programme in Uganda
 which included questions and analysis on the extent to which project outputs and input were
 equitably distributed across groups of beneficiaries. Another notable good practice was the
 295 IFAD Nigeria evaluation which endeavoured to estimate costs per beneficiaries within its
 efficiency criterion.

**Finding 4: Including GEEW as a separate stand-alone criterion tended to correlate with a
 lower level of mainstreaming across other criteria**

300 The review also found that including GEEW as a stand-alone criterion generally
 ensured that it was addressed in the report but also resulted in a lower level of
 mainstreaming across the other criteria. Very few reports achieved systematic
 integration of GEEW across all criteria. A good practice exception to this was the UN
 305 Women Evaluation of the Joint Programme in Uganda which integrated GEEW across
 the criteria and also within a stand-alone criterion. Given the challenges of most
 evaluations in addressing GEEW under the ‘efficiency’ and ‘sustainability’ criteria, the
 evaluation stood out in its ability to effectively integrate also under these areas. In
 terms of the former, the evaluation posed the question: “To what degree are partners
 310 changing their policies or practices to improve human rights and gender equality
 fulfillment (e.g. new services, greater responsiveness, resource re-allocation,
 improved quality etc.)?” and regarding the latter, the included questions was “To what
 degree are partners changing their policies or practices to improve human rights and
 gender equality fulfillment (e.g. new services, greater responsiveness, resource re-
 315 allocation, improved quality etc.)? Another good practice example is the UNCDF
 evaluation of the MicroLead programme where gender was included as a cross-cutting
 area of analysis.

In terms of inclusion of GEEW-related questions and sub-questions, there were significant
 320 variations in the percentage of questions included for each evaluation. The review found that
 where GEEW-related questions were included in the TOR, generally, they were reflected in the
 evaluation report although this was not always the case and ratings did not always account for
 discrepancies between the TOR and final report.

325 Common characteristics of reports that exceeded Criteria 2:

- ➔ GEEW was integrated across at least two criteria and within multiple questions/sub-questions
- ➔ GEEW included as a cross-cutting theme across the criteria or as a sixth stand-alone OECD-DAC criterion
- 330 ➔ The best reports contained a mainstreaming of GEEW across criteria and questions as well as a sixth GEEW-specific criterion.

Criterion 3: Methods (Independent Rating = 1.5)

335 “A gender-responsive methodology, methods and tools and data analysis techniques are selected.”

Finding 5: Gender responsive methods are the weakest aspect of the overall body of reports, and are insufficient to meet the required UN SWAP standard. More than half of the reports independently assessed failed to meet or exceed requirements.

340

Gender responsiveness of evaluation methodology, methods and tools, and data analysis techniques received the lowest rating across the criteria with 28% of independently assessed reports not meeting the requirements. This aligns with the finding from the UN-SWAP EPI 2015 Reporting Cycle Results which found that evaluation reports scored lower regarding use of gender-responsive evaluation methodology/analytical techniques.¹⁴ This finding was also validated during interviews with members of the UNEG Working Group on Gender Equality and Human Rights when asked which criterion they thought was the most difficult to assess.

345

350

Criterion 3 was also the most overrated with 54 per cent (N=25) of reports rated as ‘exceeding’ requirements under the original assessments compared with 15 per cent (N=7) of reports reviewed under the independent assessment.

355

52 per cent of the reports that were originally rated as ‘exceeding’ or ‘meeting’ requirements (N=24) were re-classified by the independent review as either only ‘approaching’ or ‘missing’ requirements for Criterion 3.¹⁵

Finding 6: Most reports contained limited detail about how GEEW was included in data collection tools and methods of analysis

360

Many reports included a general statement that gender-responsive methods were used but most reports lacked sufficient detail about how gender equality was incorporated into the evaluation design and approach and, more specifically, how analytical methods were used to draw out gender implications.

365

Most reports described the use of triangulation and mixed methods and to a lesser extent, the integration of gender within data collection methods which tended to focus on the use of disaggregated data and the inclusion of women in focus groups discussions and interviews. Some evaluations provided concrete examples of how data collection tools were modified and adapted to maximise participation of women and girls in the evaluation process (see good practice examples below). For the majority of reports, there was limited detail on how gender was integrated within data collection tools and analysis methods.

370

375

Evaluation reports rated as ‘meeting’ or ‘exceeding’ requirements often mentioned the different UNEG standards in the methods sections, but a very limited number made explicit reference to the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation.

380

Whilst mention was often made of participatory approaches and the inclusion of women and/or girls within evaluation processes, most reports did not include comprehensive stakeholder analysis whereby barriers to participation, including potential unequal power relations, were identified and mitigation measures to maximize inclusion were discussed.


It is also worth noting that in a number of instances where internally-reviewed evaluation reports were rated as ‘meets’ or ‘exceeding’, there were examples where basic UNEG standards were not fully met (i.e. there was no description of methods, no conclusions section or no indicators


¹⁴ UNEG UN-SWAP Evaluation Performance Indicator, 2015 Reporting Cycle Results.


¹⁵ Of the 24 reports re-classified, 20 were rated as ‘approaches requirements’ and 4 as ‘missing requirements’.


385 included), rendering it difficult to assess integration of the specific UNEG guidance on gender equality and human rights. In other cases, although the TOR required integration of GEEW in the methods, there were some cases where the evaluation reports did not deliver on the requirements of the TOR.


390 Among the reports that were rated as ‘exceeds’ criteria, featured good practice examples in data collection methods and tools included the following:

395  The WFP mid-term Operational Evaluation “Iran PRRO 200310 Food Assistance and Education Incentive for Afghan and Iraqi Refugees (2013-2015) included sites for field visits with women’s committees, developed gender-sensitive questionnaires and included specific questions on gender roles and relations within the data collection tools.

 As part of the evaluation of GEF’s Small Grant Programme, the country visit project performance review template contained a specific section dedicated to rating GEEW.


400  The evaluation team conducting the OCHA evaluation of the Inter-Agency Humanitarian Evaluation of the Response to the Crisis in South Sudan included a specific team member responsible for community consultations who travelled with other members of the evaluation team to allow for triangulation and for communities’ voices to act as a cross check to agencies’ perceptions. Having one person dedicated to community consultation and capturing the voices of a wide range of community members allowed for the views of affected people on various aspects of the response to be recorded.


405  The evaluation of UNCDF’s Youthstart programme included a separate section in the report explaining how gender was considered as part of the evaluation approach. A lengthy list of measures is provided including use of outreach indicators with a gender breakdown, convening all female focus group discussions where needed and use of gender-sensitive language.


410  Innovative approaches and mechanisms were used during evaluations to capture the voice of key beneficiaries. During the evaluation of the UN Women project ‘Achieving E-Quality in the ICT Sector’ in Jordan Most Significant Stories of Change were collected and used to reflect the voice of women impacted by the project. During the WFP evaluation in Iran (PRRO 200310), voice of refugees were also reflected in disaggregated fashion.


415


Common characteristics of reports that exceeded Criteria 3:

 Stakeholder analysis was conducted and methods were designed to reflect and/or address stakeholder diversity and needs

420  Detail was provided about how data collection methods and tools were designed to be gender-responsive and to maximize inclusion by addressing potential participation barriers (surveys/interview protocols) address GEEW

 Use of data disaggregated by sex

425  The evaluation team included members with specific gender and/or human rights expertise (in instances where it was feasible and possible)

 GEEW was applied to purposive sampling framework to ensure that potential issues identified during context analysis could be brought out.

430 **Criterion 4: Findings, Conclusions and Recommendations (Independent Rating = 2.1)**

“The evaluation findings, conclusions and recommendations reflect a gender analysis”

435 **Finding 7: Approaches to mainstreaming GEEW across the findings, conclusions and recommendations varied and a systematic approach to weaving gender more comprehensively across all three was often missing. The independent review rated 24 per cent of reports as ‘approaching’ or ‘missing’ requirements.**

440 Reports that effectively integrated GEEW into the evaluation scope of analysis and across multiple criteria and questions, consistently mainstreamed GEEW across findings, conclusions and recommendations. Where reports included a stand-alone section on GEEW, this usually resulted in a specific gender-related finding but not always a reflection of GEEW in the conclusions and recommendations.

445 **24 per cent of the reports that were originally rated as ‘exceeding’ or ‘meeting’ requirements (N=11) were re-classified by the independent review as either only ‘approaching’ or ‘missing’ requirements for Criterion 4.¹⁶**

450 A common trend among the findings sections, was for evaluations to reference gaps in monitoring data related to GEEW and to conclude that gender was not adequately addressed in the intervention design and implementation. Evaluation reports also often concluded that where data was limited, there was insufficient information to conduct a gender analysis, thus passing responsibility back to the entity. Whilst most evaluation reports rely on primary monitoring data, without this, evaluators can still look at the wider body of theory and sources of secondary data in assessing how a particular intervention was designed (or not designed) to address gender.

455 For example, in the case of a protection programme being designed without consideration to GEEW, an evaluator can draw on evidence and theory about effective gender-responsive protection models and refer to this within the analysis and findings.

460 Among the reports that ‘exceeded’ requirements, two reports went beyond the requirements of UN-SWAP EPI by identifying lessons learnt, challenges and recommendations for conducting gender-responsive evaluations based on the experience of the particular evaluation. These reports were the UN Women End of Programme Evaluation of the UN Joint Programme on Gender Equality in Uganda and the Mid-Term UNDAF Review for Kyrgyzstan commissioned by UNDP.

465 **Common characteristics of reports that exceeded Criteria 4:**

- ➔ Gender analysis was reflected in the report findings
- ➔ Conclusions and/or recommendations addressed GEEW
- ➔ Gender-related lessons learnt were included

¹⁶ Of the 11 reports re-classified, 10 were rated as ‘approaches requirements’ and 1 as ‘missing requirements’.

470 Summing up: What does it mean to meet the UN-SWAP EPI?

Based on the review of reports and the identified areas of convergence where entities were assessed as having met the UN-SWAP EPI requirements, the following table below summarised the key requirements of each rating area:

475

Table 3: Checklist of requirements for meeting UN-SWAP EPI

Misses	Approaches	Meets	Exceeds
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> May refer to UNEG Standards for gender	<input type="checkbox"/> "Misses" plus: <input type="checkbox"/> May state that the object did not take account of gender <input type="checkbox"/> May state that the M&E system did or did not collect disaggregated data <input type="checkbox"/> May be participatory in methods <input type="checkbox"/> May have a separate paragraph or subsection discussing gender in isolation	<input type="checkbox"/> "Approaches" plus: <input type="checkbox"/> Will include at least one gender-specific question <input type="checkbox"/> Will include gender-disaggregated indicators <input type="checkbox"/> Will consult with rights holders <input type="checkbox"/> Includes gender analysis in the background context section <input type="checkbox"/> Will include discussion of gender implications throughout the discussion of effectiveness <input type="checkbox"/> May include standalone conclusions and recommendations of gender	<input type="checkbox"/> "Meets" plus: <input type="checkbox"/> Will include gender questions and indicators under all criteria <input type="checkbox"/> Will disaggregate the gender of evaluation informants <input type="checkbox"/> May include rights holders in participatory analysis phase <input type="checkbox"/> May discuss the implications of different standpoints and power of informants <input type="checkbox"/> Will use theory and evidence to discuss gender under all criteria <input type="checkbox"/> Will highlight gender dimensions under multiple conclusions <input type="checkbox"/> Will include a recommendation on how to strengthen gender responsiveness

480 2.2 Trends in UN-SWAP EPI Ratings by Type of Review

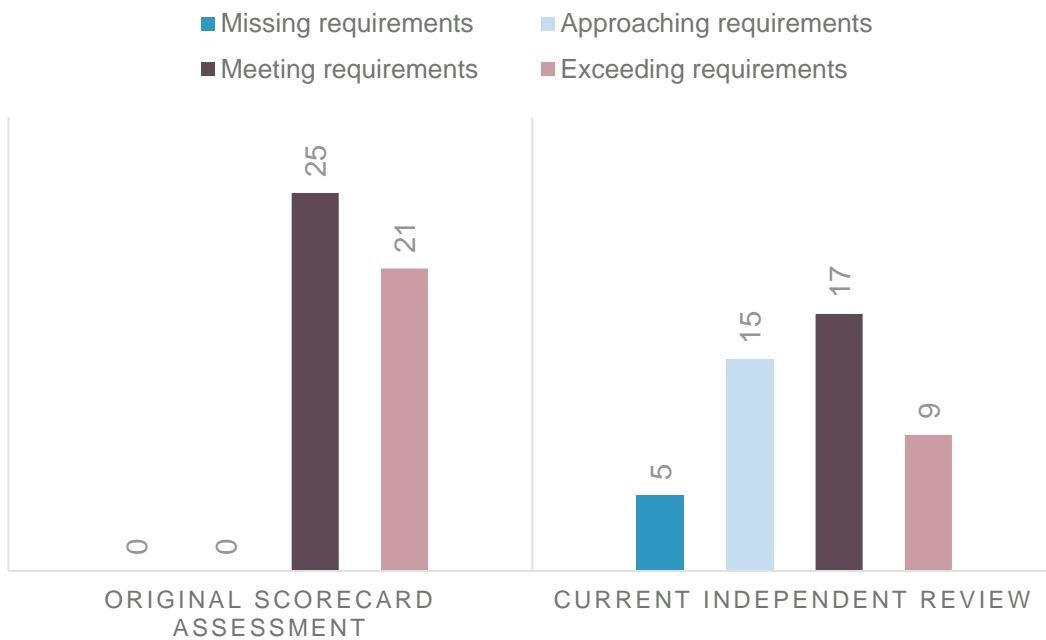
Trends by Type of Review

485 **Finding 8: There were significant discrepancies between review ratings included in the original scorecard and the independent assessment with 81 per cent of the exceeds categories re-classified as ‘meets’, ‘approaches’ or ‘misses’.**

490 Differences between the original UN-SWAP EPI scorecard assessments (conducted by UN entities, external companies/consultants and the PLE) and the independent assessment reflect inconsistent understandings of what it means to ‘exceed’ requirements. These discrepancies also demonstrate how difficult it is to compare and assess adherence to UN-SWAP criteria across UN entities based on the self-assessment scores.

495 Of the 21 reports that were rated as ‘exceeding’ based on the original scorecard ratings, 81 per cent of these (N=17) were re-classified by the independent assessment as ‘meets’, ‘approaches’ or ‘misses’ requirements.¹⁷

Figure 4: Comparison of Original Scorecard and Independent Review Ratings



500

505 **Finding 9: Internal reviews were the least reliable type of review with the highest rate of inconsistency**

Whilst internal reviews are still the most common review type and comprised 65 per cent of the overall reports originally assessed, significant discrepancies between the original scorecard assessment and the independent review rating were evident. Of the 21 reports that were rated as ‘exceeding’ based on the original scorecard assessment, 71 per cent of these were internally

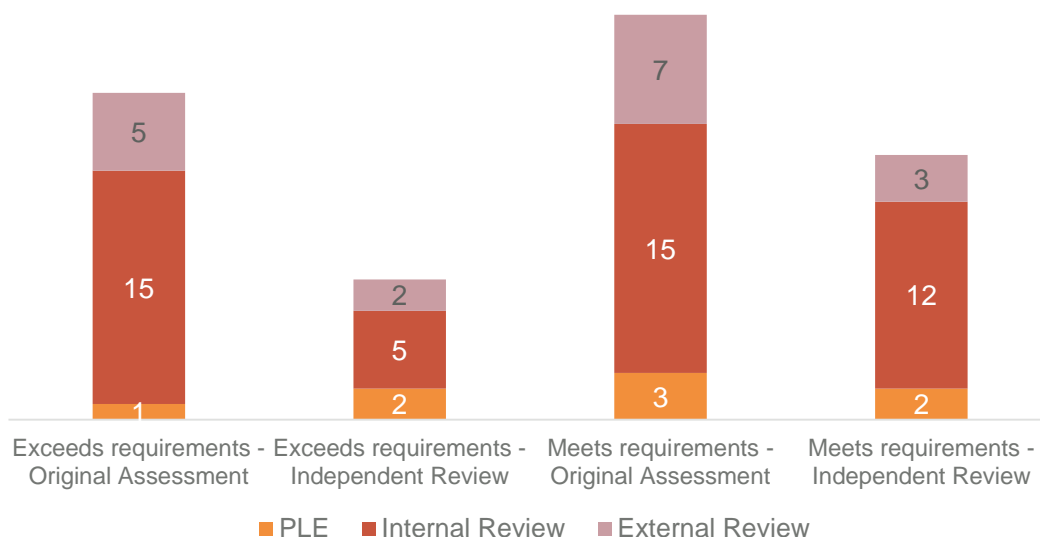
¹⁷ Out of the 17 reports that were re-classified by the independent assessment, 9 were re-classified as ‘meeting requirements’, 5 as ‘approaching requirements’ and 3 as ‘missing requirements’.

510 reviewed reports and out of the 17 reports that were re-classified at a lower level by the independent assessment 76 per cent of these were internally reviewed reports.

This aligns with the finding from the UN-SWAP EPI 2015 Reporting Cycle Results which found that out of the 31 entities that used the UNEG scorecard, those with internal reviews were about
 515 four times more likely than those with an external perspective to score “exceeds” or “meeting” requirements.¹⁸

Another striking discrepancy was that four reports rated as ‘exceeding’ requirements by the original scorecard assessment, were rated as ‘missing’ requirements by the independent review
 520 and did not even meet the general UNEG evaluation standards.

Figure 5: Comparison of Ratings by Review Type¹⁹



525 **Finding 10: External and PLE, were more systematic in applying UNEG Guidance on HR/GE.**

Out of the 17 scorecard reports that were re-classified at a lower level by the independent assessment, only 29 per cent of these were external or PLE reviews. A key reason for this was
 530 that external and peer reviews were more systematic in applying the UNEG Guidance on Gender Equality and Human Rights.

Although part of a pilot phase, the PLE review process rated as the most reliance with the least infrequencies. As shown above in Figure 5, there was very strong alignment between the PLE
 535 review scores and those of the independent review

The results of reports reviewed through the UNEG Peer Learning Exchange were consistently the most objective and therefore allow for the greatest comparability across entities.

¹⁸ UNEG UN-SWAP Evaluation Performance Indicator, 2015 Reporting Cycle Results.
¹⁹ The assessment did not include any reports originally scored as ‘missing’ or ‘approaching requirements’ so the original assessments included in the graph are only for reports rated as ‘exceeding’ or ‘meeting’ requirements. Of the 15 reports re-classified as ‘approaching requirements’, 10 of these were Internal Reviews and 5 External Reviews. For the 5 reports re-classified as ‘missing requirements’, 3 were Internal Reviews and 2 were External Reviews.

540 2.3 Trends in UN-SWAP EPI Ratings by Type of Entity and Evaluation

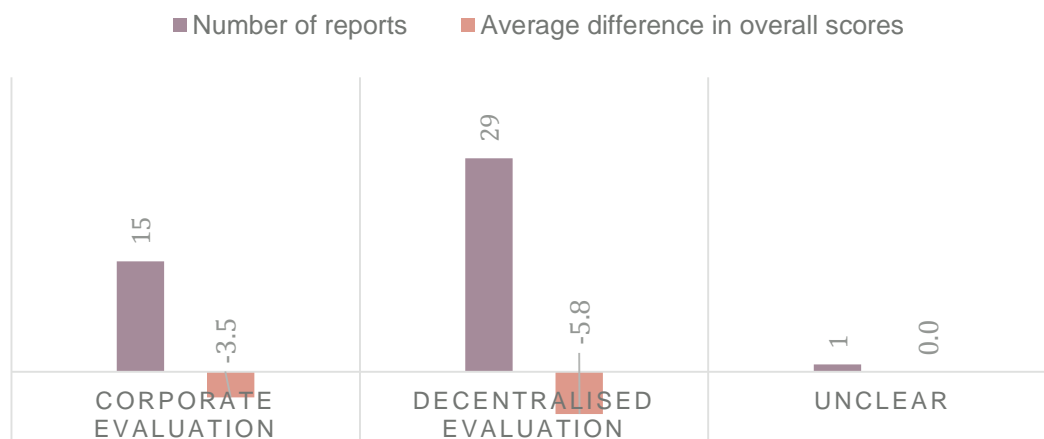
Finding 11: UN-SWAP EPI ratings were generally lower for decentralized evaluations

545 UN-SWAP EPI ratings were generally higher for corporate evaluations and lower for decentralized evaluations. Some of the possible reasons for this might be related to challenges some entities face in ensuring quality assurance at the decentralized level where there are more limited human resources for managing evaluations and where the pool of evaluation experts, including those with specific expertise in gender-responsive evaluation, may be more limited.

550 Deeper analysis of the review data on evaluation type showed that decentralized evaluations for Funds and Programmes entities scored higher than the average for decentralized evaluations overall. Possible reasons for this might be due to the fact that many of these entities have independent quality management systems and either evaluation or M&E specialists in place at the regional level. Many of the Funds and Programmes entities also have strong guidance and tools to support decentralized evaluations and are generally better supported by their entity's central evaluation unit.

555

Figure 6: Trends in UN-SWAP EPI Ratings by Evaluation Type



560

2.4 Challenges and Good Practices in Meeting the UN-SWAP EPI

565 Challenges in Applying the UN-SWAP EPI Technical Note and Scorecard

570 Although the Technical Note and Scorecard is being used widely by UN entities to review their evaluation reports, 54% of UN-SWAP Focal Points surveyed felt that further changes were needed to improve the application and use. Feedback harvested through the survey of UN-SWAP Focal Points and key informant interviews with members of the UNEG Working Group on Gender Equality and Human Rights identified the following challenges in relation to the use and application of the Technical Note and Scorecard:

- 575 ➔ The Technical Note is too long and not practical enough and a more example-based approach is needed, especially since the UNEG guidelines and guidance on gender equality and human rights were viewed by some Focal Points as heavy in content and difficult for non-gender and non-human rights experts to navigate through and apply during the UN-SWAP EPI reporting.
- 580 ➔ The Technical Note does not define what is meant by demonstrating “effective use” of the UNEG Guidance.
- 585 ➔ The Technical Note and scorecard are not mandatory which compromises comparability of reporting.
- 585 ➔ The Technical Note does not include review of TOR and inception reports as a unit of analysis (unless they happen to be in the annex of the report) which limits the ability of reviewers to pull out more information and examine the potential influence of prescriptive TOR that include GEEW specific questions and require the use of gender responsive methods.
- 590 ➔ Institutional efforts to put systems and processes in place to strengthen gender-responsive evaluation are not taken into account in the UN-SWAP EPI scorecard. A number of the UN-SWAP Focal Points felt that the Scorecard should take into account measures to review evaluation and organizational policies for gender-responsiveness and that this should factor into overall performance of entities in relation to the UN-SWAP EPI.

595 **Institutional Challenges in Ensuring Gender Responsive Evaluation**

600 Although the UN-SWAP Technical Note and Scorecard and the UNEG guidelines on integration of HR/GE provide important guidance for UN entities, work on institutional evaluation and gender policies, strategies and guidelines is equally critical and such efforts will take time to result in improved gender-responsive evaluations.

605 A large number of UN-SWAP Focal Points, both in the survey responses and interviews, noted that without integration and analysis of GEEW into programme design, evaluations will be limited to assessing *whether* gender was considered rather than *how effectively*. UN-SWAP Focal Points widely agreed that if GEEW constitutes an element of future programmes and interventions, it will be easier to assess GEEW in evaluations. Focal Points felt that many programme managers still have limited knowledge about gender mainstreaming and that there is a continued need to provide training and to strengthen organisational policies and programme management guidance to support and require the integration of gender into programmatic work. Evaluation can therefore serve as a way to incentivize and hold organisations accountable for integrating human rights and gender quality in programme.

610

615 Some UN-SWAP Focal Points noted that evaluations are often guided more by evaluation TOR and the evaluation policy of the entity than the UNEG HR/GE guidance which is why it is important that both evaluation policies and TOR reference the Guidance. Without TORs that require gender to be included within the scope of an evaluation, many of the Focal Points felt that evaluation reports are less likely to be gender responsive and more likely receive a lower UN-SWAP score.

620 The fact that many entities have very limited human resources for supporting evaluation was also raised by a number of Focal Points and that it is difficult to adequately support gender

625 responsive evaluations within these limitations. The challenges of finding evaluation consultants with a background in GEEW and a strong knowledge of the required technical field was also highlighted widely. Whilst some Focal Points were able to engage specific gender specialists within evaluation teams, most of the Focal Points noted that resources are often too limited for this. In such cases, an important approach for tapping into existing internal resources is to support the active engagement of gender focal points in each entity.

630 A number of Focal Points underscored the need for greater UNEG support in developing and implementing gender-responsive evaluation methodologies, and in particular the need for tools to support and reflect the results of interaction with affected populations/rights holders. One Focal Point identified the need to move beyond sex-disaggregation of data to evaluation of changes in gender dynamics, roles and relations and the requirement for greater UNEG support in this regard.

635

Good Practice Examples of Evaluation TOR Requiring Integration of GEEW

640 The review of reports and discussions with UN-SWAP EPI Focal Points underlined the importance of strong and prescriptive TOR requiring GEEW integration that:

645 **➔ Require full integration of GEEW across all criteria and as a stand-alone criterion.** An example of this was the TOR for the End of Programme Evaluation of the Joint Programme on Gender Equality in Uganda which included specific GEEW-related questions across all criteria and the UNCTAD Evaluation “Enhancing Capacities of Developing Countries to Mainstream Gender in Trade Policy” which included a stand-alone criterion on gender.

650 **➔ Require integration of GEEW in scope, findings, conclusions and recommendation and collection of sex-disaggregated data.** A good practice example was the UNCDF Youthstart evaluation (see section 2.1). Following an interview review of its TOR, UNEP has since included included specific questions under each TOR (using the UNEG guidance on gender equality and human rights).

655 **➔ Reference and include the UN-SWAP Technical Note and Scorecard references in evaluation ToRs.** This was done in the Report of the Inter-Agency Humanitarian Evaluation of the Response to the Crisis in South Sudan.

660 **➔ Consult with the internal gender focal point/unit for advise on integrating GEEW into evaluations and, where resources permit, require or recommend inclusion of gender specialists within evaluation teams.** This is a regular practice in IFAD and during the WFP evaluation of Food Assistance and Education Incentive for Afghans and Iraqi Refugees, an evaluation team member was designated as focal point on gender and protection. During the Evaluation of WFP’s Regional Response to the Syrian Crisis, two team members covered gender and the evaluation of UNCDF’s MicroLead Programme, the TOR required experience with gender- and equity-focused evaluations.

665

3. Recommended Actions

670

In order to meet the UN-SWAP EPI, the following actions are recommended:

Recommendation 1:

675

The UNEG Working Group on Gender Equality and Human Rights can develop a practical, example-based checklist for integration of GEEW in evaluation reports.

680

Given the wide discrepancies in how the UN-SWAP EPI criteria are understood and applied, there is a need for a more simplified and user-friendly checklist to assist entities commissioning evaluations and evaluators/reviewers themselves in ensuring that evaluation reports meet the requirements of the UN-SWAP EPI. The checklist should be annexed to TOR and evaluation reports and used by the commissioning entity as a quality assurance tool whereby evaluations need to tick all the annexed to evaluation TOR and also to reports).

Table 4: Proposed elements of a checklist for UN-SWAP EPI criteria

<p>GEEW included in scope/analysis</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Evaluation scope: mentions that GEEW will be addressed <input type="checkbox"/> Evaluation criteria: GEEW integrated across all criteria (ideal) and/or addressed specifically as a stand-alone/cross-cutting criterion <input type="checkbox"/> 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 <input type="checkbox"/> Gender disaggregated indicators included and mainstreamed across the Evaluation Matrix <input type="checkbox"/> 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)
<p>GEEW included in methods</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Sex-disaggregated breakdown of stakeholders/respondents <input type="checkbox"/> 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 <input type="checkbox"/> 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 <input type="checkbox"/> Detail about availability/use of GEEW-related documents and data <input type="checkbox"/> Detail about how ethical considerations were addressed

GEEW included in Findings, Conclusions and Recommendations	<input type="checkbox"/> GEEW analysis included and reflected across findings <input type="checkbox"/> GEEW reflected in the evaluation conclusions <input type="checkbox"/> Specific recommendations related to GEEW included <input type="checkbox"/> Lessons learnt and good practices related to GEEW included
---	---

685

Recommendation 2:
The UNEG Working Group on Gender Equality and Human Rights is recommended to continue to promote and encourage agencies with limited funds for external reviews of their UN-SWAP EPI to utilize the Peer Learning Exchange

690

The Peer Learning Exchange process which was piloted in 2015 by the Working Group on Gender Equality and Human Rights has proven to be a highly effective and reliable mechanism for agencies with limited funds to use when assessing their evaluation reports for gender-responsiveness. It is therefore recommended that the Working Group continue to promote this process to UN-SWAP EPI Focal Points and more widely to UNEG members, especially in the lead up to the 2016 UN-SWAP Reporting Cycle.

695

Recommendation 3:
The UNEG Working Group on Gender Equality and Human Rights is best placed to provide increased opportunities for more in depth learning exchange and peer mentoring.

700

Because SWAP reporting is not only for accountability purposes, but also to help evaluation entities self-assess and see where and how they need to improve, UNEG Working Group members expressed a strong desire for increased learning and exchange opportunities.

705

Feedback from surveys and interviews conducted during the review process highlighted a strong desire for the UNEG Working Group to play a greater role in mentoring, providing continued help desk support and regular opportunities for UN entities to share concrete examples of gender integration in evaluation, especially among entities that do not have GEEW as their focus. A number of Focal Points highlighted the need for more training, tools and webinars related specially to gender-responsive methods in evaluation. Given that gender-responsive methods was identified as the weakest area of evaluation reports, the need for more focused support in this area is quite high.

710

715

Recommendation 4:
The UNEG Working Group on Gender Equality and Human Rights are recommended to develop guidelines for agencies to use when: a) reviewing and evaluating their Gender Policies; and b) aligning their evaluation report quality systems and evaluation policies with UNEG guidance on gender equality and human rights

720

Reponses to the survey of UN-SWAP EPI Focal Points indicated that eight agencies are planning a review of their corporate gender equality policy in the next three to five years. In supporting these efforts, it would be useful for the Working Group to develop guidelines for gender quality policies as well as a sample of good practice policies from different entities.

725

Recommendation 5:

730

The UNEG Working Group on Gender Equality and Human Rights is advised to provide Working Group members with access to a means of identifying evaluators with specific expertise in gender-responsive evaluation

735

During review consultations, members of the UNEG Working Group identified challenges their entities face in finding evaluators with demonstrated gender expertise, especially in technical areas. Members felt that it would be extremely useful to have access to lists of evaluators who have demonstrated competencies in conducting gender-responsive evaluations and whose evaluations were rated by the independent, PLE process or external reviews as ‘meeting’ or ‘exceeding’ UN-SWAP EPI requirements. Ideas tabled to achieve this included: 1) providing UN-wide common access to rosters of gender-responsive evaluators through the Working Group (updated on a quarterly basis with UN-SWAP EPI Focal Points); or 2) establishing a self-guided certificate course on gender responsive evaluation modelled on UNDSS Basic Security in the Field or ILO’s forthcoming evaluators’ course on tripartism.

740

Recommendation 6:

745

To further rationalise the UN-SWAP EPI criteria, the first two criteria can be merged in order to prevent overlap, and specific changes to each criterion can be made based on the table provided.

750

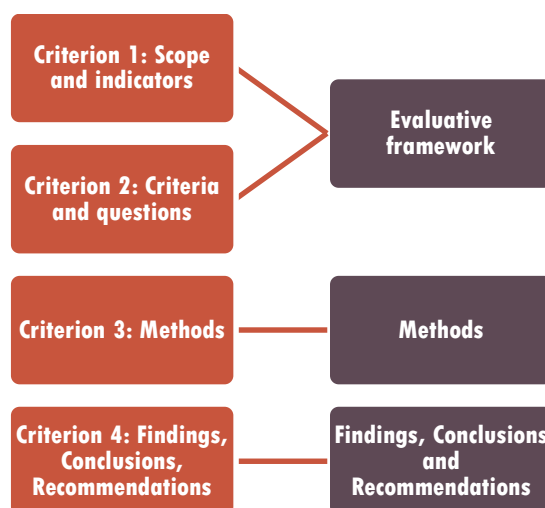
Analysis of the interpretation of EPI criteria reveals that the boundaries between criterion 1 and criterion 2 can be easily confused, since both cover the planning of the evaluation. Accurate analysis requires a clear delineation in the evaluation report of indicators, questions, criteria and scope; which is not always the case in documents. Furthermore, having two indicators relating to planning weights the overall score towards how evaluations were proposed to be gender responsive, rather than how they actually delivered.

755

In order to further clarify the criteria and ensure consistent interpretation it is recommended to combine the existing criteria 1 and 2 into a single criterion that covers the integration of gender within the evaluative framework. This would also help to balance the weight of the overall UN SWAP score equally between the aspects of planning (evaluation framework), process (evaluation methods) and outputs (evaluative analysis).

760

Figure 7: Suggested Actions to Adjust the Criteria



765

770

In addition to the proposed streamlining of the criteria, it is recommended that the definitions used for rating evaluations and evaluation functions are made more systematic, and are based explicitly on the proposed UN SWAP EPI checklist. Although this approach has implications for the level of professional judgement and contextualisation that can be applied to an individual evaluation report, it is considered to be necessary at this stage given the wide variance in how ratings are being interpreted across the different entities. A proposed set of definitions are provided below.

Table 5: Suggested Amendments to the UN-SWAP EPI Scorecard Criteria

Rating	Individual evaluation	Evaluation function
Exceeds	<p>Fully meets UNEG GE-related norms and standards, applies the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation during all phases of evaluation</p> <p>(Meets all requirements within the UN-SWAP EPI Checklist and includes examples of good practice beyond the requirements of the checklist)</p>	<p>On average, reports meet all requirements within the UN-SWAP EPI Checklist</p> <p>Plus</p> <p>Conducts at least one evaluation of corporate gender mainstreaming or evaluation of its GE policy/strategy every 5-8 (or 3-5) years and implements GEEW evaluation recommendations and conducts SWAP assessments through independent or peer-review process</p>
Meets	<p>Fully meets UNEG GE-related norms and standards and applies the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation</p> <p>(Meets all requirements within the UN-SWAP EPI Checklist)</p>	<p>On average, reports meet all requirements within the UN-SWAP EPI Checklist</p>
Approaches	<p>Partially meets UNEG GE-related norms and standards in the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation (based on UN-SWAP EPI Checklist)</p> <p>(Meets more than half of requirements in the UN-SWAP EPI Checklist)</p>	<p>On average, reports meet more than half of requirements within the UN-SWAP EPI Checklist</p>
Misses	<p>Does not meet UNEG GE-related norms and standards in the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluation (based on UN-SWAP EPI Checklist)</p> <p>(Meets less than half of the requirements in the UN-SWAP EPI Checklist)</p>	<p>On average, reports meet less than half of requirements within the UN-SWAP EPI Checklist</p>

775

Annex A: Terms of Reference

780

Review of UN-SWAP Evaluation Performance Indicator Reporting

Commissioned by the UNEG Working Group on Gender Equality and Human Rights

785

Background

The UN System Wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP) constitutes the first accountability framework for gender mainstreaming in the UN system.²⁰ The UN-SWAP is composed of 15 performance indicators for tracking 6 main elements on gender mainstreaming: accountability, results based management, oversight, human and financial resources, capacity, and knowledge exchange and networking. All UN entities are to self-assess and report on progress. UN entities are expected to meet all UN-SWAP-performance standards by 2017.²¹

790

795

Reporting on the UN-SWAP commenced in 2013 and entities are expected to report on a yearly basis through the Report of the Secretary-General to ECOSOC on “Mainstreaming a gender perspective into all policies and programmes in the UN system”²². UN Women’s Coordination Division provides secretariat services and support to the UN system for reporting on the UN-SWAP. The Gender units of all UN entities play an internal coordinating role, ensuring that progress is accurately reported and that plans of action are developed. Accountability rests, as noted in the CEB policy, with senior managers of the different UN-SWAP reporting entities.

800

The UN-SWAP framework is accompanied by a set of Technical Notes for each Performance Indicator that provide guidance on how to complete the assessment for each of the 15 Performance Indicators. While the UN-SWAP Performance Indicators approved by the CEB are established, the Technical Notes are considered live documents that can be enhanced.

805

In 2015, UN Women initiated a review of the implementation of the UN-SWAP, which will result in a revised UN-SWAP framework to be rolled out in 2018.²³

810

UN-SWAP Evaluation Performance Indicator (EPI)

The oversight element of the UN-SWAP is composed of three performance indicators, including one dedicated to evaluation that is linked to meeting the gender-related UNEG Norms²⁴ & Standards²⁵ and demonstrating effective use of the UNEG guidance on integrating gender equality in evaluation²⁶.

815

The reporting categories for the Evaluation Performance Indicator (PI5) are as follows:

²⁰ UN-SWAP Framework was developed by UN Women in 2011/2012 in response to the CEB endorsed UN system-wide Policy on Gender Equality and the Empowerment of Women (CEB/2006/2), which was established based on the ECOSOC Agreed Conclusions 1997/2. On 13 April 2012, the CEB endorsed the UN-SWAP for application throughout the UN system. In its resolution E/RES/2012/24 of 27 July 2012, the ECOSOC welcomed the UN-SWAP and called upon the UN system to actively engage in its roll-out and report on the implementation of the resolution at its substantive session in 2013.

²¹ There is an extended timeframe to 2019 for those entities with a mainly technical focus

²² For example see, United Nations, “Mainstreaming a gender perspective into all policies and programmes in the UN system,” April 2014; accessible online at: http://www.un.org/ga/search/view_doc.asp?symbol=E/2014/63&Lang=E

²³ Plans for the roll-out of UN-SWAP/2 are still under consideration, as the framework will need to be endorsed by CEB.

²⁴ United Nations Evaluation Group, Norms for Evaluation in the UN system, 2005; accessible online: <http://unevaluation.org/document/detail/21>

²⁵ United Nations Evaluation Group, Standards for Evaluation in the UN system, 2005; accessible online: <http://unevaluation.org/document/detail/22>

²⁶ United Nations Evaluation Group, Guidance on Integrating Human Rights and Gender Equality in Evaluation, 2014; accessible online: <http://unevaluation.org/document/detail/1616>

Not Applicable	Missing	Approaches	Meets	Exceeds
5a. Performance indicator is not relevant to a UN entity	5b. None of the UNEG gender-related norms and standards are met	5c. Meets some of the UNEG gender-related norms and standards	5d. Meets the UNEG gender-related norms and standards	5ei. Meets the UNEG gender-related norms and standards and 5eii. Demonstrates effective use of the UNEG guidance on evaluating from a human rights and gender equality perspective

820 The possible revision of the Evaluation Performance Indicator is currently under discussion by UNEG, which would be proposed as part of the overall process for revision of the UN-SWAP. If any revision were to take place, it would be implemented in the 2018 reporting cycle.

825 The UNEG Working Group on Gender Equality and Human Rights developed the Technical Note and Scorecard, which aims to support more systematic and harmonized reporting through the use of a common tool that allows for improved comparability across UN entities. UNEG Heads endorsed the Technical Note and Scorecard in August 2014 (available on the UNEG website: <http://www.uneval.org/document/detail/1452>). The unit of analysis selected as most feasible was the evaluation report. Thus the Evaluation Performance Indicator should be solely based on

830 an assessment of evaluation reports completed in the reporting year. The Technical Note specifies the below criteria for the assessment of integration of gender equality in the evaluation reports:

- 835 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*
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.*
- 840 3. *A gender-responsive methodology, methods and tools and data analysis techniques are selected.*
4. *The evaluation findings, conclusions and recommendations reflect a gender analysis.*

It is recommended that evaluation units conduct an external review, however, as this requires financial resources, the UNEG Working Group on Gender Equality and Human Rights piloted a Peer Learning Exchange (PLE) process in 2015, which proved successful. At a minimum, UN entities

845 are highly encouraged to submit the Scorecard to allow for comparability. Through the UNEG Working Group on Gender Equality and Human Rights, UN Women Independent Evaluation Office provides training webinars on the UN-SWAP EPI reporting process and Scorecard, and help desk support.

850 **Purpose and Objectives of the Review**

The overall purpose of the review is to contribute to a common understanding of what it means to “meet” or “exceed” requirements for the UN-SWAP Evaluation Performance indicator and thus “Meet the UNEG gender-related norms and standards for evaluation”, and to identify

855 where targeted support for UNEG members is required in order to “meet requirements” for the UN-SWAP Evaluation Performance Indicator, and to identify whether it is necessary to revise the current Scorecard criteria and corresponding Technical Note. Several factors have led to the UNEG Working Group on Gender Equality and Human Rights commissioning this review:

- 860 • Two reporting cycles have been completed using the same scorecard approach²⁷ and thus a wealth of information exists from the evaluation reports scored by UN entities

²⁷ Access the Technical Note and Scorecard here: <http://www.uneval.org/document/download/2148>

- The UN-SWAP EPI 2015 reporting cycle results²⁸ identified the need for reviewing how UN entities are applying the gender norms and standards in evaluation practice
- We only have one year until the 2017 UN-SWAP deadline; and
- There is the possibility to revise the Evaluation Performance Indicator for the next cycle of UN-SWAP and thus the corresponding criteria (as outlined in the Scorecard) used to assess performance.

The objectives of the review are:

1. Identify trends in reporting against the UN-SWAP EPI scorecard criteria
2. Identify good practice in integrating gender equality in evaluation reports
3. Identify challenges to integrating gender equality in evaluation reports
4. Suggest actions for supporting UNEG members to meet UN-SWAP EPI requirements
5. Suggest possible revisions to the Scorecard criteria

875 Methodology

As mentioned above, the UN-SWAP EPI is based on the scoring of evaluation reports against 4 criteria. Thus, both the scorecard (submitted by 31 UN entities in 2015) and the evaluation reports (approximately 378 were scored in 2015) will be used as the main source of data for this review. In order to narrow the universe of evaluation reports to be sampled, the universe will be limited to those reports that were scored “meets” (N=93) or “exceeds” (N=68) requirements in 2015 for a total universe of 161 reports. The sample of entities will be stratified by type of entity (see table below) and overall entity rating (exceeds, meets, approaches, and missing) in order to identify trends for these categories.

Type of entity	Number of entities submitted scorecard	Number of reports	Exceed	Meets	Approaches	Missing
Funds & Programmes	12	203	1	5	6	0
Secretariat	14	90	3	4	5	2
Specialized & Technical	5	85	0	3	2	0
TOTAL	31	378	4	12	13	2

The following questions and means of analysis are proposed:

Unit of Analysis	Question	Means of analysis
Scorecard	What are the common factors (trends) by scorecard criteria for those evaluation reports that obtained an overall “exceeds”, “meets”, “approaches”, and “missing” requirements?	Based on the sample of entities, analyse the comments section of the scorecard to identify thematic trends within each overall reporting category and by scorecard criteria.
Evaluation report	What are the common characteristics (trends) by scorecard criteria for those evaluation reports that obtained an overall “exceeds” or “meets”?	Based on the sample of entities, identify thematic trends within each overall entity rating category and by scorecard criteria.

²⁸ The report can be accessed on the UNEG website: <http://www.uneval.org/document/download/2536>

Evaluation report	What are the good practices within each scorecard criteria from those evaluation reports that scored overall “meets” or “exceeds”?	Same as above.
Evaluation report	What are the common challenges within each scorecard criteria from those evaluation reports that scored overall “meets” or “exceeds”?	Same as above.
Evaluation report and Scorecard	To what extent does the external reviewer’s score align with the original score?	Based on a blind sample from each overall reporting category, score evaluation reports.

890

The UN-SWAP EPI Focal Point at UN Women will provide the reviewer with the entity scorecards, and on this basis, the reviewer will identify the sampled evaluation reports either via email or through online research. Additionally, approximately 5-10 UN-SWAP EPI focal points may be consulted through select interviews in order to gather opinions on the Technical Note/Scorecard and clarify any issues that may arise during the analysis.

895

In addition to the above trends, the report will identify trends according to the type of organization (Funds and Programmes, Secretariat, etc.), type of original UN-SWAP EPI review (internal, external /Peer Learning Exchange or other type of review), and according to the number of reports assessed by the entity. The actions proposed should be aligned with the application of the UNEG Guidance on Integrating Human Rights and Gender Equality in Evaluations²⁹. The final methodology applied will be agreed in consultation with the Management Group.

900

905 **Management Arrangements**

This review is being commissioned by the UNEG Working Group on Gender Equality and Human Rights. The Working Group has established a management group consisting of representatives from: UN Women, UNESCO UNEP, UNESCO, UNCTAD, OHCHR, UNICEF, and UNCDF. UN Women will serve as the focal point for the review, managing the overall process, including hiring and accountability for overall quality assurance.

910

The management group will be consulted at key steps in the review process: Terms of Reference, hiring of Consultant, methodology note, draft report and final report. They will also support communications of the report findings and development of a plan of action to take forward the key actions proposed by the review. One individual consultant will be hired.

915

920 **Deliverables**

The Review will commence in July 2016 and be completed in September 2016; approximately 35 days of work are required.

920

The deliverables are as follows:

- Draft and final (after consultation with Management Group) methodology note
- Draft report (including all data and analytical tools) + 2 rounds of revisions
- Audit Trail matrix of how comments were addressed
- Final report (maximum 15 pages + annexes)
- Communication products (SlideDoc and one-two page brief with info-graphics to communicate key messages of report including good practices identified)

925

930

²⁹ UNEG, Integrating Human Rights and Gender Equality in Evaluations, Aug. 2014: <http://www.uneval.org/document/detail/1616>

Requirements

- 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 responsive 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
- Familiarity with UN-SWAP an asset
- Language: must be fluent in English and either Spanish or French

935

940

945

Annex B: List of Evaluation Reports Assessed

Entity	Evaluation title	2015 rating	2015 score
ECLAC	Strengthening Government and Civil Society Capacity to Incorporate Economic and Social Rights into Macroeconomic Policy	Exceeds Requirements	11
ECLAC	Strengthening National Capacities to Design and Implement Sustainable Energy Policies for the Production and Use of Bio-Fuels in Latin America and the Caribbean	Exceeds Requirements	11
DSS	Report of the Evaluation of the Security Programme in El Salvador	Meets Requirements	10
ESCAP	Capacity building for control authorities and transport operations to improve efficiency of cross-border transport in landlocked and transit developing countries	Meets Requirements	9
ESCAP	Regional financial and monetary architecture in Asia-Pacific	Meets Requirements	9
ESCWA	Strengthening National Capacities in the ESCWA Region on Developing Green Production Sectors	Exceeds Requirements	11
ESCWA	Strengthening Statistical Capacity of Arab Countries (Egypt, Jordan and Palestine) in Producing Energy Statistics and Energy Consumption Surveys	Exceeds Requirements	11
GEF	Evaluation of the Small Grants Programme	Exceeds Requirements	11
IFAD	PPA - Nigeria	Exceeds Requirements	12
IFAD	SS - Indigenous Peoples	Meets Requirements	8
ILO	GLO/11/53/SID	Meets Requirements	8
ILO	GLO/12/50/RUS	Meets Requirements	9
IOM	Mauritania - Assessment Report of the Community Stabilization Project implemented by IOM in Hodh El Chargui	Exceeds Requirements	12
IOM	Enhancing Resilience and Protection of Marginalized Communities Affected by Protracted Conflicts and Effects of Adverse Climatic Conditions in Kenya	Exceeds Requirements	11
ITC	Evaluation of the Trade, Climate Change and Environment Programme	Exceeds Requirements	12
ITC	Midterm Evaluation of Horticulture Productivity and Trade Development Project in Lesotho	Meets Requirements	10
OCHA	Report of the Inter-agency Humanitarian Evaluation (IAHE) of the Response to the Crisis in South Sudan	Meets Requirements	10
OCHA	IAHE Central African Republic	Meets Requirements	8
OHCHR	Tajikistan	Exceeds Requirements	11
OHCHR	Moldova	Meets Requirements	10

OIOS	Evaluation of the implementation and results of protection of civilians mandates in United Nations peacekeeping operations, A/68/787	Exceeds Requirements	11
OIOS	Evaluation of the UN Economic and Social Commission for Asia and the Pacific (ESCAP) IED-14-008	Meets Requirements	10
UN Women	Final Evaluation of the United Nations Joint Programme for Gender Equality in UGANDA	Exceeds Requirements	11
UN Women	Achieving E-Quality in the ICT Sector	Meets Requirements	9
UNCDF	YouthStart Final Evaluation	Meets Requirements	10
UNCDF	MicroLead Expansion programme: mid-term evaluation	Meets Requirements	8
UNCTAD	Enhancing Capacities of Developing Countries to Mainstream Gender in Trade Policy	Exceeds Requirements	12
UNCTAD	External evaluation of UNCTAD subprogramme 3: International trade	Meets Requirements	9
UNDP	Mid-term UNDAF Review	Exceeds Requirements	11
UNDP	Democratic Governance: CPD	Meets Requirements	9
UNESCO	Evaluation of UNESCO's Work on Culture and Sustainable Development	Meets Requirements	9
UNESCO	Final Evaluation UNESCO Project "Empowering Local Radios with ICTs"	Meets Requirements	10
UNFAO	Evaluation of FAO's Contribution to Climate Change Adaptation and Mitigation	Exceeds Requirements	12
UNFAO	EU FAO Forest Law Enforcement, Governance and Trade Support, II GCP /GLO/395/EC	Meets Requirements	9
UNFPA	EVALUACIÓN DEL PROGRAMA DE PAÍS DEL UNFPA EN URUGUAY 2011/2015	Exceeds Requirements	12
UNFPA	China CP7 CPE Report	Meets Requirements	8
UNHabitat	Evaluation of the UN-Habitat Sudan Country Programme 2012-2015, November 2015	Exceeds Requirements	11
UNICEF	Evaluation of Country Programme of Co-operation between Government of Uzbekistan and UNICEF 2010-2014	Exceeds Requirements	12
UNICEF	RKLA3 multi-country evaluation: increasing access and equity in early childhood education – final evaluation report	Meets Requirements	8
UNV	Arab Youth: Volunteering for a Better Future	Meets Requirements	10
WFP	Iran PRRO 200310 Food Assistance and education incentive for Afghan and Iraqi Refugees (2013-2015): A mid-term Operation Evaluation	Exceeds Requirements	11
WFP	An Evaluation of WFP's Regional Response to the Syrian Crisis, 2011-2014	Meets Requirements	8
WHO	Evaluation of DFATD-funded Project Accelerating Nutrition Improvements in Sub-Saharan Africa	Exceeds Requirements	12
WHO	Evaluation of WHO's Presence in Countries	Meets Requirements	8
WIPO	Copyright and Related Rights Evaluation	Meets Requirements	10
WIPO	Small and Medium-sized Enterprises and Innovation	Meets Requirements	8

Annex C: Stakeholder Consultation List

Stakeholders	Number Consulted	Female	Male	Data Method	Collection
UNEG WG on GE and HR	-	-	-	Webinar	
UN-SWAP EPI Focal Points	24	-	-	Survey	
UN-SWAP EPI Focal Points	14	-	-	Survey ³⁰	
Elisa Calcaterra UNEP	1	1	0	Semi-structured interview (SSI)	
Andrew Fyfe UNCDF	1	0	1	SSI	
Dawit Habtemarian WFP	1	0	1	SSI	
Daniel Chen UNCTAD	1	0	1	SSI	
Pascale Reinke-Schreiber UNODC	1	1	0	SSI	
Total Number of Stakeholders Consulted	43				

³⁰ UN-SWAP EPI Revision— Gauging Interest Survey administered by UN Women (1 April 2016)

Annex D: Documents Consulted

1. United Nations Evaluation Group, Norms for Evaluation in the UN system, 2005; accessible online: <http://unevaluation.org/document/detail/21>
2. United Nations Evaluation Group, Standards for Evaluation in the UN system, 2005; accessible online: <http://unevaluation.org/document/detail/22>
3. United Nations Evaluation Group, Guidance on Integrating Human Rights and Gender Equality in Evaluation, 2014; accessible online: <http://unevaluation.org/document/detail/1616>
4. UNEG UN-SWAP Evaluation Performance Indicator, 2015 Reporting Cycle Results.
5. UNEG UN-SWAP Evaluation Performance Indicator, 2015 Reporting Cycle Results.
6. UNEG, Integrating Human Rights and Gender Equality in Evaluations, Aug. 2014: <http://www.uneval.org/document/detail/1616>
7. United Nations, "Mainstreaming a gender perspective into all policies and programmes in the UN system," April 2014; accessible online at: http://www.un.org/ga/search/view_doc.asp?symbol=E/2014/63&Lang=E

Annex E: Biographies of Evaluation Team Members

Joseph Barnes



Joseph is a specialist in gender responsive evaluations, with extensive experience in designing and implementing evaluations of complex objects, including the UN Women Corporate Evaluation of Women’s Economic Empowerment, and other impact and programme-level evaluations for WFP, UNICEF, Irish Aid, and DFID. He has also worked extensively with UNDP, other multi-lateral organisations, trusts and NGOs.

Joseph is team leader of both the UNICEF and UN Women evaluation report quality assessment and analysis systems. In 2011, Joseph founded ImpactReady to help bridge the gap between evaluation, programme design, and social business.

Jo-Anne Bishop



Jo-Anne 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, 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.



Addendum

Independent Review of UN-SWAP EPI Reporting – Plan of action UNEG – SO3 Working Group on Gender Equality and Human Rights December 2016

Recommendation	Management Group response	Timeframe	Responsible
Practical, example-based checklist for integration of GEEW in evaluation reports	Using the proposed checklist and UN Women checklist as a starting point, build upon this to adapt to UNEG needs → using the levels of reporting to indicate examples of how it could be approaching, meeting, exceeding (building on the review)	Q1 2017	UNESCO UN Women
Broaden Peer Learning Exchange and share concrete examples of gender integration	PLE call for applicants was sent – deadline of 30 Nov. Review guidance to see how it can be strengthened to highlight examples of gender integration.	Q1/2 2017	Subgroup working on this will discuss
More in-depth learning exchange	Using the “good practice examples” identified by the review: a) Host webinars on specific topics that were highlighted (i.e. methods) and invite the UN entity to present on this experience – lessons learned and how they would strengthen b) Develop 2-pager briefs on these examples so that they can be stored and accessed by others online c) One-on-one consultations with the agencies that were reviewed as part of the UN-SWAP EPI to identify challenges, good practice and suggestions for moving forward	Q1/2 2017	Subgroup on webinars to discuss/ focal point for UN-SWAP EPI
Guidelines for agencies to use when reviewing/evaluating GE Policies	This is part of the UNEG Working Group plan for 2016-2017 (sub-group was established and an way forward discussed) consultant to be recruited.	Q1 2017	UN Women & GEF (sub-group established)



<p>Align evaluation report quality assessment systems and evaluation policies with the proposed UNEG Guidance</p>	<p>Identify good practices in the annual UN-SWAP EPI trends report that goes to UNEG and/or “good practice note” on the cases where UN-SWAP has been integrated into the quality assessment systems (UNICEF, UN Women, UNDP, etc) to show case the approaches and provide ideas for others.</p>	<p>Q1/2 2017</p>	
<p>Require integration of GEEW across all evaluation criteria in addition to an (optional) standalone criterion</p>	<p>As we cannot “require” integration, we can develop a “good practice note” on the cases where this has been done well, as identified by the review. As noted above.</p>	<p>As noted above</p>	<p>As noted above</p>
<p>Access to a roster of evaluators with specific expertise in gender-responsive evaluation</p>	<p>Work with professionalization working group to advertise consolidated list of existing rosters; including advertising better the UN Women Gender and Evaluation Consultant Database.</p>	<p>TBD</p>	<p>TBD</p>
<p>Revisions to the Technical Note</p>	<p>Propose that this be undertaken as part of the working group activities for 2017-2018 (in time for implementation of the new UN-SWAP, which will come into effect in 2018</p>	<p>May – August 2017</p>	<p>TBD</p>