



UNEG
United Nations Evaluation Group

Reference
Document

Lessons Learned from 2016 Peer Learning Exchange

UN-SWAP Evaluation Performance Indicator

This document is issued by the UNEG Working Group on Gender Equality and Human Rights and was drafted by Ekaterina Sediakina-Rivi, Associate Evaluation Specialist, UNESCO with the support of Sabas Monroy, OHCHR and Elisa Calcaterra, UNEP

March 2017

UN-SWAP Evaluation Performance Indicator 2016 Reporting Cycle - Peer Learning Exchange

The second edition of the Peer Learning Exchange (PLE) was launched by the UNEG Working Group on Gender Equality & Human Rights in the fall of 2016 with an open call for participation to all UNEG entities. Evaluation offices without the funds for external reviews were encouraged to participate in the PLE, which had been found to be just as reliable as an external assessment by an independent reviewer.

Nine evaluation entities participated in the second edition of the PLE: IOM, OHCHR, UNHABITAT, UNHCR, ESCAP, UNIDO, UNCTAD, UNCDF, and UNRWA. Entities were paired by size (number of evaluation outputs) and mandate to the extent possible. A sub-group was set up to facilitate the PLE by UNWOMEN, OHCHR, UNEP and UNESCO. These agencies had piloted the PLE in 2015 and were able to share their experience with the newly participating entities. They paired the entities and provided guidance throughout the PLE process. [UNEG Guidance](#) for the PLE was also made available.

The PLE had two main objectives:

- to provide for an independent external assessment of each agency's evaluation reports, at no extra cost (besides staff time); and,
- to enable exchange on good practices of integration of gender equality and human rights into the evaluation process.

The methodology of the PLE consisted of swapping evaluation reports between entities, scoring them according to the UN-SWAP Evaluation Performance Indicator (EPI) scorecard, sharing the results of the scores, and exchanging on the rationale behind the scoring, as well as on good practices. The paired entities then drafted short reports on the PLE experience, highlighting lessons learned. These reports are annexed to this summary report.

In following the PLE Guidance, paired entities agreed on a feasible and equivalent number of reports to be exchanged. Consequently, a selected number of reports (ranging from 1 to 6) were scored by partner agencies, while the remainder were scored internally. As a result, the participating entities did not rely solely on the PLE for their reporting to the UN-SWAP. Reasons behind this are two-fold. First, as the evaluation entities produced unequal numbers of evaluation reports, it was considered unfair to exchange all reports, thereby leaving entities with larger numbers of reports to score the remainder on their own. Second, it appears that agencies preferred to do some of the scoring themselves, all while benefitting from an external assessment for certain reports. The table below provides figures on the number of reports assessed during the PLE, as well as the scores resulting from the PLE and those that were reported by each agency for the EPI to UN-WOMEN. It shows that participants in the PLE mainly included entities that produced less than ten evaluations in 2016, with the exception of IOM and UNIDO. The PLE therefore attracted mainly smaller entities that did not hire external assessors.

Table 1: Peer Learning Exchange Participating Entities

	IOM	UNHABITAT	UNCTAD	UNCDF	UNHCR	UNRWA	UNIDO	ESCAP	OHCHR
Total reports produced in 2016	20	6	4	2	2	5	30	8	1
Reports rated by agency for EPI	18	6	2	1	2	5	28	8	1

Reports reviewed for PLE	3	3	2	1	2	2	6	6	1
EPI Rating reported to UN-WOMEN	Approaches	Meets	Meets	Exceeds	Not submitted	Approaches	Meets	Meets	Exceeds
EPI Rating resulting from PLE					Meets	Approaches	Meets	Meets	Exceeds

All participating entities reported to be very satisfied with the PLE and learned a lot from exchanging with other evaluation offices. The PLE further highlighted the importance of integrating gender equality and human rights throughout the evaluation process and provided for some unique insights on practices adopted by some entities. For example, UN-HABITAT has developed specific requirements for evaluation Terms of Reference (ToR) that include gender equality issues within the scope and methodology, as well as detailed evaluation questions. This ensures that gender equality is systematically included in every evaluation. UNCDF and OHCHR have developed a thorough methodology for integrating gender equality into the entire evaluation process, from the design of the Request for Proposals, to the inception phase and the evaluation reports. UNCTAD makes sure that gender equality is listed as a required skill for Team Leader and Team Members of all evaluations.

Nearly all participating entities stated that the scoring process was a challenging one for a number of reasons. First, not all evaluation reports were presented with supporting documentation such as inception reports, annexes, Terms of Reference, interview guides, questionnaires and a description of the methodology used, expertise/profiles of evaluators, etc. While presenting this information was not mandatory for the purpose of the scoring, it would have facilitated the assessments. This resulted in uneven information on each evaluation and thereby undermined the uniformity of the overall process. Second, each evaluation entity has its own interpretation of the EPI criteria and use of the scorecard. Third, a lack of prior knowledge of each agency's corporate processes and evaluation culture further challenged the assessment of evaluation reports.

The main conclusions of the PLE are:

- The PLE provided for a useful opportunity to test internal assessments against impartial external ones. The PLE is not, however, a substitute for a full external assessment of entities' reports, as its scope did not cover all evaluation reports produced by the larger entities. Some of the scoring of the remaining reports still had to be done by the entities themselves. The table above also shows, however, that agencies reported the rating resulting from the PLE in their final reports on the EPI to UN-WOMEN.
- The EPI reporting is a very subjective exercise in nature. Much of the assessment depends on each agency's or external assessor's interpretation of the EPI scorecard criteria. This is partly due to a lack of guidance on the criteria and partly due to individual understanding.
- The EPI's focus on the evaluation report is rather limiting, as it overlooks important steps in integrating gender equality and human rights into the overall evaluation process or life cycle.
- The provision of supporting documents in addition to evaluation reports facilitated entities' scoring of the latter and provided for a more comprehensive picture of the evaluation process as a whole.
- The inclusion of gender equality and human rights into an evaluation's ToR does not always ensure that they are fully integrated into the findings, with reports lacking a gender analysis.
- Discussions on methodologies among UN colleagues were very useful. A number of agencies shared good practices that should inspire future practices. See annexes.

Recommendations for the EPI criteria and process:

- Consider revising the EPI criteria to reflect the evaluation process as a whole, rather than just on the final deliverable, which is the evaluation report.
- Provide detailed guidance on each criterion of the EPI to avoid misinterpretation.
- Develop a checklist for the EPI in order to unify the assessment process of each entity.
- Start the EPI reporting process and accompanying PLE earlier in the year to avoid disruption by the end of year holiday break.

Recommendations for improving the integration of gender equality and human rights into evaluation processes:

- Gender equality and human rights need to be integrated into the entire life cycle of an evaluation, starting from the development of the ToR, the launch of the call for proposals, the inception phase, the data collection and analysis, and the reporting and dissemination.
- A process of quality control during the development of the evaluation ToR needs to be put in place to ensure that questions related to gender equality and human rights are included. This subsequently shapes the scope of the evaluation, the questions and the findings.
- Develop a model ToR with the integration of gender equality and human rights in the evaluation objectives, criteria (gender as a cross cutting criterion), questions (including examples), and methodology (selection of consultants, data collection, stakeholders to be interviewed, etc.).
- Evaluation questions on gender equality and human rights should be included for all criteria and not just set aside as a standalone criterion.
- Include the UNEG Guidance on HR and GE in the reference documents for each evaluation ToR.
- Present data disaggregated by gender and type of stakeholder in evaluation report as much as possible. Ensure that this data is used in an analysis of gender equality issues.

Annexes: Reports of the PLE participating entities

Evaluation Performance Indicator Peer Learning Exchange

IOM – UN HABITAT (Summary of the Peer Learning Exercise)

Introduction and Background

IOM conducted and published a total of 20 evaluations and reviews in 2016: 12 internal (including those conducted by the Office of Inspector General/OIG) and 8 external decentralized evaluations, not managed by the OIG and conducted by external consultants. A total of 6 external evaluation reports were rated for the UNSWAP Evaluation Performance Indicator (EPI). The assessment of reports was done internally by the Oversight Officer from the Evaluation Unit.

In UN-Habitat, the Evaluation Unit is charged with responsibility for the evaluation function. The Evaluation Unit manages centralized evaluations; these are most often carried out by external consultants, and the Unit provides technical support to decentralized evaluations, which are managed by project managers. As of September 2016, all new projects in UN-Habitat are reviewed and scored against a 'gender marker' in order to be approved. None of the evaluated projects that were assessed of the 2016 batch had undergone the 'gender marker' scoring. UN-Habitat completed 6 evaluation reports in 2016, which were rated for the UNSWAP EPI. The UN-Habitat Evaluation Reports were assessed against the EPI internally by a professional staff of the Evaluation Unit.

Methodology of the Peer Learning Exchange

IOM and UN-Habitat exchanged a selection of 3 evaluation reports prepared by each agency during the year 2016. The UN-Habitat evaluation reports contained both the final reports and ToRs, while one IOM evaluation report contained ToR. IOM reports included annexes with interview guides and one report included methods used. For UN-Habitat reports, methods and approach used were described in the TORs and in a section of the report. However, only one report had included in annexes the guide and questionnaire templates used. Each agency scored the reports of the other and provided comments to justify the score. A teleconference to discuss the process, results and lessons learned was held on 11th January 2016. Following the call, the agencies exchanged the findings from the peer learning exchange+.

Key Observations

Both agencies agreed that rating each other's reports and exchanging experiences was a very useful and interesting exercise. Assessing the reports of another agency presented the following challenges:

- The evaluation reports shared were considered to be of good quality and following general norms and standards for evaluations in the system without considering the inclusion of GEEW.
- Scoring of the reports among peers produced different interpretation of the way how to rate the evaluation reports. Overall, IOM scored UN-Habitat reports higher than UN-Habitat itself. UN-Habitat scored two of IOM reports lower and one higher.
- Scoring the reports without any prior knowledge of the processes and methods used, but also without having full ToRs or Inception Reports presented a challenge for both agencies. Also information about the programming context of the projects and lack of access to the full logical frameworks of projects was a source of uncertainty. Most of the time, gender specific information on the methodology used, the data collection methods, interview guides, the expertise of the evaluators, etc., are not presented in the final report or even in the ToRs. For a more precise assessment of the evaluation process further supporting documentation is very much needed.

- The major difference between the two agencies' reports is that there exists a system of centralized quality control of ToRs within UN Habitat that does not exist in IOM and which gives the possibility of a better coverage of gender analysis already during the ToR drafting process. In IOM only when ToRs are shared for technical guidance, then such focus can be proposed. The result is that the ToR requirements for UN-Habitat evaluations already include detailed gender issues with the scope and methodology detailing gender questions. In case of IOM reports, gender issues are included in the ToRs only in case the project itself is dealing with women empowerment issues or gender based violence. So in most of the cases ToRs list limited gender questions for IOM reports, which affects the scope of the evaluation questions used in the evaluation process and ultimately the findings in the reports produced.

Lessons Learned and Recommendations for the Improvement of the EPI

- It is relatively straight forward to assess and rate scoring criteria 1 and 2 of the UN SWAP EPI, which focus on the planning part of the evaluation, but more difficult to rate scoring criteria 3 and 4 as these assess the evaluation process and its output.
- Supporting documentation attached to the each evaluation report that is assessed remains a prerequisite to ensure future ratings are more aligned, including information about the methodology used, data collection methods, interview guides, the expertise of the evaluators, etc.
- ToR requirements need to include detailed gender issues with the scope and methodology detailing gender questions. It should be considered if projects *without* a gender specific expected accomplishment should be rated in the UN SWAP against the same scoring criteria or scale as projects *with* a gender focus or gender specific expected accomplishment. The sophistication of the methods used and data analysis techniques is heavily affected by the extent to which gender is the focus in the project, which again affects the findings, conclusions and recommendations of the report.
- Evaluators need to have the expertise in not only the evaluation subject matter but also in gender equality and human rights-related competencies. Evidence of such expertise is not readily available in evaluation reports, but might have been considered during the selection and recruiting of the evaluation consultants.

(UNCTAD – UNCDF) Summary of the Peer Learning Exercise

Introduction and Background

UNCTAD and UNCDF joined the Peer Learning Exchange (PLE) at the end of 2016 SWAP Evaluation Performance Indicator (EPI) reporting process on a voluntary basis. Evaluation and Monitoring Unit (EMU) of UNCTAD produced 4 evaluation reports in 2016, out of which two reports were rated using the scorecard for the SWAP (EPI) by UNCDF.¹ UNCDF completed 2 programme evaluation reports in 2016 (one in French and one in English) and UNCTAD selected the one in English to provide the rating using the EPI scorecard.

Methodology and process of the Peer Learning Exchange

The coordinators of the UNEG Working Group on Gender Equality and Human Rights sub-group on peer learning exchange paired UNCTAD and UNCDF on the basis of the rationale that they have similar size of operations as well as some common areas of work. The PLE began on 14 December 2016 and was completed on 16 January 2017.

To initiate the PLE, UNCTAD and UNCDF focal points for the exercise shared background information of the evaluations conducted in 2016, including the ToRs, inception reports and evaluation reports. This allowed both agencies to get a full understanding of how gender equality and human rights were integrated throughout the whole process leading to an evaluation report responding to gender equality requirements.

After UNCDF made the final selection decision, UNCTAD further shared the two selected evaluations' inception reports. UNCTAD chose to assess one of the two evaluation reports (in English) due to language limitations and received the Request for Proposal, inception report, and the final report for that evaluation.

Each agency scored the report(s) of the other and provided detailed comments to explain the rationale for the rating. A teleconference was held on 16 January 2017, between the UNCTAD and UNCDF EPI focal persons to discuss the process, results, lessons learned and also exchanged experiences in applying SWAP EPI guide in managing evaluations.

Following the teleconference, each organization compiled the rating and comments into the final SWAP EPI scorecard report to UN Women. UNCTAD revised the comments made by UNCDF and lowered the rating for one criterion, out of the 4 criteria for one of the evaluation report assessed by UNCDF. The decrease reflects UNCTAD's view that the evaluator wasn't strong enough in that particular criterion. UNCDF submitted the rating done by UNCTAD as such.

Key Observations and Reflection on the peer learning exercise

Overall, both agencies agreed that rating each other's reports and exchanging experiences was a very useful exercise. It was especially beneficial to UNCTAD as UNCDF's programme evaluation has a very thorough methodology to integrating gender equality into evaluations, as reflected in the RFP design, the inception phase, and the evaluation reports. UNCTAD would like to participate in the EPI on a regular basis to ensure that it continues the learning process by partnering with different agencies. The results from this exercise were also shared with senior management and ensured a greater visibility of gender mainstreaming efforts by EMU.

¹ The assessment of remaining two reports was done internally by the Evaluation and Monitoring Unit (EMU) of UNCTAD.

UNCDF also found the external assessment extremely useful, particularly the detailed external analysis and validation provided by UNCTAD. UNCTAD highlighted key elements in our processes that were particularly efficient in integrating gender equality in our evaluations, such as fully integrating gender in the scope and methodological requirements of our evaluations or having GE as required skills for Team Leader and Team Member of the evaluation.

In addition, particularly useful for UNCDF was the discussions held with UNCTAD: on the possibility of using different approaches in mainstreaming gender equality in our evaluations, such as the benefits of having evaluations questions as a separate category in the evaluation matrix vs integrating them throughout the OECD/DAC criteria; how to best further improve our evaluations regarding gender equality.

There are however still some challenges in the peer learning exercise:

- Overall speaking, the EPI scorecard's criteria need to be further improved to facilitate a standard and uniform assessment to avoid different interpretation of requirements/ standards.
- It would be useful to integrate Terms of References and Inception reports as part of the scorecard, as the evaluation reports gender responsiveness is to a large extent influenced by the quality of those two documents.
- Furthermore, although the current EPI scorecard process focuses only on evaluation reports, UNCTAD argues that scoring the reports without assessing the processes and methods used by the agency and/or the evaluator is a major challenge. Just like in evaluations, the assessment cannot simply be based on the desk review of documents and different data collection methods should be applied to help with the triangulation and validation of data and information.

Lessons Learned

Lessons learned by UNCTAD

- UNCTAD has room to improve its evaluation design by stressing more on the requirement to assess programmes/ projects' outcome and impact on gender equality, although this is largely also linked to the lack of integration of gender equality into project design and implementation. As evidenced by UNCDF's Request for Proposal, there could be the assessment of outcome and impact level achievements in terms of a programme's integration of gender equality into its planning and implementation.
- Evaluations conducted in UNCTAD usually list gender equality as a separate criterion in the Terms of Reference (ToR) for evaluations. In certain cases, it was incorporated under relevance. However, efforts can be made to integrate it into most criteria to fully assess the gender impact and constraints of programmes/ projects.

Lessons learned by UNCDF

- UNCDF will be working to improve its evaluation design by increasing focus on gender analysis throughout the report and in the conclusions and recommendations sections. From a process point of view, we have most of the elements in place to have gender equality fully integrated in our evaluations, however gender equality analysis still has some room for improvement. UNCDF will work on improving its gender analysis, by looking at how best to use different gender analysis frameworks (for example, the Longwe Equality and Empowerment Framework or the Harvard Analytical Framework) in our evaluation when they are applicable to the nature of UNCDF's intervention. Furthermore, UNCDF will continue their strategy of integrating gender criteria in the OECD/DAC criteria and work towards making its recommendations more gender specific.

Recommendations for the Improvement of the EPI

- UNEG needs to further develop the EPI by using a practical checklist to standardize and unify the assessment process. The revised EPI should provide more guidance and standards on evidence-based assessments.
- The EPI process could go beyond individual exercises and also include some previous evaluation reports to help agencies review progress over time and identify systemic gaps. This has the potential drawback of entailing a larger time burden for those agencies in participating in EPI.
- The EPI assessment criteria should look beyond what is stated in the evaluation report and examine the whole "life cycle" of an evaluation, including the evaluation process management, which may not necessarily be reflected in the evaluation reports.

Challenges

- There are challenges in finding evaluators that have a good combination of the knowledge of the subject matter, evaluation expertise, as well as gender equality and human rights-related competencies in the non-traditional development work that agencies like UNCTAD and UNCDF undertake.

(OHCHR - UNRWA) Summary of the Peer Learning Exercise

(Reviewer: Robert Stryk, Chief Evaluation Division)

Background

The Office of the United Nations High Commissioner for Human Rights (OHCHR) represents the world's commitment to universal ideals of human dignity. We have a unique mandate from the international community to promote and protect all human rights. OHCHR's thematic priorities are strengthening international human rights mechanisms; enhancing equality and countering discrimination; combating impunity and strengthening accountability and the rule of law; integrating human rights in development and in the economic sphere; widening the democratic space; and early warning and protection of human rights in situations of conflict, violence and insecurity.

OHCHR is committed to being a fully results-based organization, clearly defining results in terms of the changes it plans to achieve and reviewing all aspects of its work in light of how it contributes to bring about those changes. Following the rules and regulations, as well as United Nations Norms and Standards for Evaluation as developed by the United Nations Evaluation Group (UNEG) and with the aim of becoming a fully results-based organization, OHCHR has therefore established an evaluation function within the Department, and adopts and regularly updates its vision and policy for evaluation.

The United Nations System Wide Action Plan (UN SWAP) includes a systematic self-assessment on how UN system agencies work on gender equality and the empowerment of women at the corporate level. The UN SWAP has a specific performance indicator for evaluation that falls under the oversight category. The United Nations Evaluation Group (UNEG) developed and endorsed the UN SWAP Evaluation Performance Indicator (EPI) Technical Guidance and Scorecard to assess the evaluation reports conducted by entities in a given year.

The UNEG Working Group on Gender Equality and Human Rights is promoting the peer learning exchange as a means for facilitating and strengthening the conduct of UN SWAP EPI. As part of the second year of the exercise, the OHCHR volunteered to be peer reviewed by the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA). This report expands on the short documentation that has been sent to UN Women on the EPI and aims to provide a more detailed feedback to OHCHR.

Methodology

UNRWA applied the EPI Scorecard tool developed by UNEG to the evaluation of OHCHR conducted in 2016: "Evaluation of the Justice Component of the Maya Programme for the full exercise of the right of the indigenous peoples of Guatemala". The documents reviewed during this exercise included the final report of the evaluation, the inception report for the evaluation, the terms of reference for the evaluation, the Evaluation Vision and Policy and the Model ToR of Technical Cooperation Projects.

Observations

The meta-evaluation score assigned to OHCHR for 2016 is 11 out of 12, which equals to "Exceeds requirements". These are the main issues addressed during the assessment under each review criteria.

Evaluation Scope of Analysis (3 out of 3 = fully integrated).

- The OHCHR evaluation vision and policy includes very strong statements for gender as well as human rights. Gender is also specifically mentioned in under human rights.

- The OHCHR procedures very well lay out how gender should be integrated and the rationale for the integration.
- The terms of reference explicitly includes gender references
- The inception report reiterates the gender references.

Evaluation Criteria and Questions (3 out of 3 = fully integrated).

- The terms of reference introduce a specific gender integration criteria.
- The criteria provides good guidance on how the evaluation should take into account gender issues.
- The inception report reiterates the criteria for gender integration into the evaluation.

Evaluation Methodology (3 out of 3 = fully integrated).

- The methodology is described in most detail in the inception report but reiterated in the evaluation report.
- The methodology refers to gender issues throughout. It includes a good description of the gender issues in the project.
- The methodology refers to the gender issues and that the relevant information should be gathered as a priority as well as including a specific indigenous women perspective section.

Evaluation Findings, Conclusions and Recommendations (2 out of 3 = Satisfactorily integrated).

- The evaluation findings cover the parts of the programme that are concerned with indigenous women rights.
- The involvement of women organizations and service to women organizations is mentioned specifically.
- The advocacy component relating to the role of women is analysed.
- The overall conclusions and recommendations however could have been more specific on the mainstreaming aspect of the programme. While this is mentioned in the ToR, and the inception report it is not specifically analysed in the results. A stronger follow through on the agreed questions and methodology could have further improved the results.

Lessons learned and good practice

Systematic integrating of gender issues in the guidance

- The approach of OHCHR to include gender throughout the process and very prominently in the evaluation vision and policy is very good. This enables the logic to flow from the vision, to the policy, to the actual tools for the evaluation.

Follow through

- While the guidance and evaluation preparation documents raise very high expectations regarding the results being gender sensitive and include gender mainstreaming the analysis in the final report seems to have not fully committed to the provisions in the Terms of Reference and Inception Report.

(UNHCR - OHCHR) Summary of the Peer Learning Exercise

(Reviewer: Sabas Monroy, Evaluation Officer)

Background

The Office of the United Nations High Commissioner for Refugees (UNHCR) is mandated to lead and co-ordinate international action to protect refugees and resolve refugee problems worldwide. It strives to ensure that everyone can exercise the right to seek asylum and find safe refuge in another State, with the option to return home voluntarily, integrate locally or resettle in a third country. It also has a mandate to protect stateless people and has been authorized to work with internally displaced populations and former refugees in their countries of origin.

UNHCR's Policy Development and Evaluation Service (PDES) is committed to the systematic analysis and assessment of UNHCR policies, programmes, projects, practices and partnerships. PDES is housed in the High Commissioner's Executive Office and its work is carried out in accordance with the UNHCR Evaluation Policy adopted in 2010 which commits UNHCR to the UNEG Norms and Standards for Evaluation in the UN System.

The United Nations System Wide Action Plan (UN SWAP) includes a systematic self-assessment on how UN system agencies work on gender equality and the empowerment of women at the corporate level. The UN SWAP has a specific performance indicator for evaluation that falls under the oversight category. The United Nations Evaluation Group (UNEG) developed and endorsed the UN SWAP Evaluation Performance Indicator (EPI) Technical Guidance and Scorecard to assess the evaluation reports conducted by entities in a given year.

The UNEG Working Group on Gender Equality and Human Rights is promoting the peer learning exchange as a means for facilitating and strengthening the conduct of UN SWAP EPI. This report presents the results of the peer review conducted by OHCHR on UNHCR evaluations conducted in 2016.

Methodology

The Evaluation Officer of OHCHR applied the EPI Scorecard tool developed by UNEG to two evaluations conducted by UNHCR in 2016: "Evaluation of Emergency Transit Centres in Romania and the Slovak Republic" and "Evaluation of UNHCR's Response to the South Sudan Refugee Crisis in Uganda and Ethiopia". The documents reviewed during this exercise included the Terms of Reference and the Final Reports of the evaluations assessed.

Observations

The meta-evaluation score assigned to UNHCR evaluations conducted in 2016 is 8 out of 12, which equals to "Meets requirements". These are the main issues addressed during the assessment under each review criteria.

Evaluation Scope of Analysis (1.5 out of 3 = Partially/Satisfactorily integrated).

- The background sections of the evaluations ToR addresses gender considerations of the intervention being evaluated
- The interventions being evaluated have indicators related to gender issues (health).

- The Terms of Reference state that the evaluations will be participatory, but without specific mention of women.
- The assessment of gender equality issues is not specifically listed among the evaluation objectives, purpose or scope.

Evaluation Criteria and Questions (1.5 out of 3 = Partially/Satisfactorily integrated).

- The evaluation criteria in the Terms of Reference of one of the evaluations mention the assessment of gender, age and diversity perspectives.
- In one case, the evaluation questions in the ToR include the assessment of gender issues under some of the evaluation criteria, including the identification of gaps in the benefits received by women.

Evaluation Methodology (2.5 out of 3 = Satisfactorily/Fully integrated).

- The evaluation methodology requires consultations ensuring that diverse groups of stakeholders are included, including men, women, boys and girls.
- The evaluation used a gender perspective during data collection and analysis phases. Focus group discussions were held, including gender-segregated groups.
- The report emphasizes on the use of the organization's gender approach and policy during the assessment.

Evaluation Findings, Conclusions and Recommendations (2.5 out of 3 = Satisfactorily/Fully integrated).

- The report findings and conclusions address several issues related to gender equality.
- There are recommendations related to gender issues in one of the evaluations assessed during the exercise.

Lessons learned and good practice

Evaluation Scope of Analysis.

- The evaluation scope should mention the assessment of gender integration among the objectives to ensure that the analysis will cover these issues. The background sections of the Terms of Reference highlight the importance of gender considerations in the interventions evaluated, but it would be useful if the evaluation purpose, objectives and scope could be more specific about the gender issues to be addressed during the evaluations.

Evaluation Criteria and Questions.

- Evaluation questions related to gender equality should be included under all the evaluation criteria in order to guide the evaluator’s analysis. For example, questions assessing the participation of women in the design of the intervention, requesting to analyse if the services offered in the centres benefit equally men and women, and addressing ways to propose different approaches in these services that could promote gender equality (the case of vocational training).

Evaluation Methodology.

- UNEG provides guidance on evaluation standards and methodologies, including the integration of gender in evaluations. The guide on integrating human rights and gender should be included in the Terms of Reference as one of the main documents to be consulted by the evaluators in terms of methodology.
- The reports present data disaggregated by gender and the methodology reflects an effort to integrate women and girls in the data collection tools, which could help to enrich the evaluation results. It would be very useful if this data and methodological efforts are used in the analysis of gender equality issues.

Evaluation Findings, Conclusions and Recommendations.

- Gender issues are addressed in the evaluation findings, conclusions and recommendations, reflecting the organizational approach in this area. The evaluations could look further into gender equality issues: What changes could be made in those centres with more women than men? Was the gender disaggregated data collected in the centres used to take strategic decisions? Do men and women have equal access to the services provided in the centres? What changes could be made in these services to promote gender equality?

General

- In addition to the organizational approach and policy on gender equality, the evaluation ToRs would benefit of the inclusion of a request to follow the UNEG Guidance “Integrating Human Rights and Gender Equality in Evaluations”. The “UN SWAP Evaluation Performance Indicator Scorecard” could be also suggested to the evaluators as a methodology for a self-assessment.
- The development of a model of Terms of Reference to be consulted during the planning of every evaluation conducted in UNHCR would be useful, particularly if this is based in good practices like the case of the evaluation of the response to South Sudan’s crisis. This model should address the integration of gender in the evaluation objectives, criteria (gender as a cross cutting criterion), questions (including examples), and methodology (selection of consultants, data collection, stakeholders to be interviewed, etc.).

Lessons learned from the peer learning exchange process

- The documents to be reviewed during the exercise should include the evaluation ToRs, not only the final report (ideally, the ToRs should be in the annexes of the report). This is the best way

to assess the integration of gender issues in the purpose, objectives, criteria, questions and methodology of the evaluation.

- The peer review exercise timing should consider the effect of the Christmas and New Year's holidays leave. If the deadline for the UN SWAP reporting is in January, the documents to be reviewed should be shared not later than the first half of December.

(UNRWA - OHCHR) Summary of the Peer Learning Exercise

(Reviewer: Sabas Monroy, Evaluation Officer)

Background

The United Nations Relief and Works Agency for Palestine Refugees (UNRWA) was established by the United Nations General Assembly resolution 302 (IV) of 8 December 1949 after the 1948 Arab-Israeli conflict to carry out direct relief and works programmes for Palestine refugees. In the absence of a just and lasting solution to the Palestine refugee problem, the General Assembly has repeatedly renewed UNRWA's mandate, most recently extending it until 30 June 2017. UNRWA has contributed to the welfare and human development of four generations of Palestine refugees. Today, more than 5 million Palestine refugees are eligible for UNRWA services. UNRWA has field offices in Lebanon, Gaza, West Bank, Jordan, and Syria and has a staff of approximately 30,000.

The UNRWA Evaluation Division assumes functional leadership and oversight over UNRWA's evaluation system, which consists of the central evaluation system in the Department for Internal Oversight Services and decentralized evaluation functions. Evaluations aim to systematically and impartially determine the relevance, effectiveness, efficiency, impact and sustainability of ongoing and completed Agency activities. These evaluations are conducted for strategies, policies, programmes and projects, as well as emergency interventions. They examine expected and achieved accomplishments, result chains, processes, contextual factors and causality. Evaluation findings, recommendations, and lessons learned are then incorporated into decision-making processes in order to strengthen learning and accountability in the Agency with the aim of improving operations and results. UNRWA's evaluation function was established in 2009.

The United Nations System Wide Action Plan (UN SWAP) includes a systematic self-assessment on how UN system agencies work on gender equality and the empowerment of women at the corporate level. The UN SWAP has a specific performance indicator for evaluation that falls under the oversight category. The United Nations Evaluation Group (UNEG) developed and endorsed the UN SWAP Evaluation Performance Indicator (EPI) Technical Guidance and Scorecard to assess the evaluation reports conducted by entities in a given year.

The UNEG Working Group on Gender Equality and Human Rights is promoting the peer learning exchange as a means for facilitating and strengthening the conduct of UN SWAP EPI. This report presents the results of the peer review conducted by OHCHR on UNRWA evaluations conducted in 2016.

Methodology

The Evaluation Officer of OHCHR applied the EPI Scorecard tool developed by UNEG to one evaluation conducted by UNRWA in 2016: "Evaluation of the e-health project". The final report of the evaluation was reviewed during this exercise.

Observations

The meta-evaluation score assigned to UNRWA evaluations conducted in 2016 is 10 out of 12, which equals to "Meets requirements". These are the main issues addressed during the assessment under each review criteria.

Evaluation Scope of Analysis (2 out of 3 = Satisfactorily integrated).

- The evaluation report highlights the importance of the intervention being evaluated in terms of gender issues (maternal health).
- The evaluation reflects the organizational gender awareness.

Evaluation Criteria and Questions (3 out of 3 = Fully integrated).

- The evaluation addresses gender issues under the criteria of relevance and impact.
- There are evaluation questions related to gender issues.

Evaluation Methodology (2 out of 3 = Satisfactorily integrated).

- The evaluation was guided by UNEG norms and standards.
- Gender issues are analysed in the survey conducted as part of the evaluation methodology.

Evaluation Findings, Conclusions and Recommendations (3 out of 3 = Fully integrated).

- There are several findings related to gender issues.

Lessons learned and good practice

Evaluation Scope of Analysis.

- The evaluation highlights the importance of gender considerations in the intervention evaluated, but it would be useful if the evaluation purpose, objectives and scope could be more specific about the gender issues to be addressed during the evaluation.

Evaluation Criteria and Questions.

- Evaluation questions related to gender equality should be included under all the evaluation criteria in order to guide the evaluator's analysis. For example, questions assessing the participation of women in the design of the system, or requesting to analyse if the benefits are received equally by men and women.

Evaluation Methodology.

- UNEG provides guidance on evaluation standards and methodologies, including the integration of gender in evaluations. The guide on integrating human rights and gender should be included in the Terms of Reference as one of the main documents to be consulted by the evaluators in terms of methodology, in addition to UNEG norms and standards.

- A survey was conducted as part of the evaluation, and gender issues are addressed in the analysis of the survey. It would be useful to include in the report the proportion of women respondents or tables disaggregated by gender that could have used to identify other lines of analysis.

Evaluation Findings, Conclusions and Recommendations.

- Gender issues are addressed in the evaluation findings and conclusions, reflecting the organizational approach in this area. The evaluation could look further into gender equality issues, for example the participation of women beneficiaries during consultations for the design of the system.

General

- In addition to UNEG norms and standards, the evaluation ToRs would benefit of the inclusion of a request to follow the UNEG Guidance “Integrating Human Rights and Gender Equality in Evaluations”. The “UN SWAP Evaluation Performance Indicator Scorecard” could be also suggested to the evaluators as a methodology for a self-assessment.
- The development of a model of Terms of Reference to be consulted during the planning of every evaluation conducted in UNRWA would be useful, particularly if this is based in good practices like the case of the evaluation of the e-health project. This model should address the integration of gender in the evaluation objectives, criteria (gender as a cross cutting criterion), questions (including examples), and methodology (selection of consultants, data collection, stakeholders to be interviewed, etc.).

Lessons learned from the peer learning exchange process

- The documents to be reviewed during the exercise should include the evaluation ToRs, not only the final report (ideally, the ToRs should be in the annexes of the report). This is the best way to assess the integration of gender issues in the purpose, objectives, criteria, questions and methodology of the evaluation.
- The peer review exercise timing should consider the effect of the Christmas and New Year’s holidays leave. If the deadline for the UN SWAP reporting is in January, the documents to be reviewed should be shared not later than the first half of December.

(UNRWA - OHCHR) Summary of the Peer Learning Exercise (Reviewer: Helen Morris, Senoir Evaluation Officer)

Background

The United Nations Relief and Works Agency for Palestinian Refugees in the Near East has a mandate to provide protection, relief and human development assistance to Palestine refugees, defined as “persons whose normal place of residence was Palestine during the period 1 June 1946 to 15 May 1948, and who lost both home and means of livelihood as a result of the 1948 conflict.” The descendants of Palestine refugee males, including legally adopted children, are also eligible for registration. Currently UNRWA’s population of concern numbers approximately 5.2 million. UNRWA operates in Lebanon, Jordan, Syria, Gaza and the West Bank.

The evaluation division in UNRWA is within the Department of Internal Oversight Services.

The United Nations System Wide Action Plan (UN SWAP) includes a systematic self-assessment on how UN system agencies work on gender equality and the empowerment of women at the corporate level. The UN SWAP has a specific performance indicator for evaluation that falls under the oversight category. The United Nations Evaluation Group (UNEG) developed and endorsed the UN SWAP Evaluation Performance Indicator (EPI) Technical Guidance and Scorecard to assess the evaluation reports conducted by entities in a given year.

The UNEG Working Group on Gender Equality and Human Rights is promoting the peer learning exchange as a means for facilitating and strengthening the conduct of UN SWAP EPI. As part of a pilot exercise, UNRWA volunteered to have its evaluation reports for 2015/16 reviewed by UNHCR and OHCHR. This report present the results of the peer review conducted by UNHCR on an evaluation conducted in 2015.

Methodology

The Senior Evaluation Officer of UNHCR applied the EPI Scorecard tool developed by UNEG to the evaluation conducted by an external consultant commissioned by UNRWA in 2015: “Ensuring the protection of conflict-affected Palestine refugees from Syria”. The documents reviewed during this exercise included the Terms of Reference and the Final Report, including annexes of the evaluation assessed.

Observation

The evaluation score assigned to the evaluation conducted in 2015 is 4 out of 10, which equals “Approaches requirements”.

Evaluation Scope of Analysis (1 out of 3 = Partially integrated)

There is no specific mention of gender in the purpose, scope or objectives of the evaluation as set out in the Terms of Reference. The background section mentions that the regional Syria response include services and "programmes for women and youth," and one of the sub-questions under Impact relates to the "impacts of the project for refugee woman and girls."

Evaluation Criteria and Questions (1 out of 3 = Partially integrated)

There is no specific mention of gender in the evaluation criteria and only one of the sub-questions refers to impact on refugee women.

The intervention being evaluated included a micro-finance component specifically targeting women, and the project results framework includes an objective with regard to women's household credit, and an indicator of number of female-headed households that benefited. The objective relating to emergency shelter also included an indicator of number of female-headed households. Therefore, some more sub-questions specifically referring to gender might have drawn out more gender-specific findings and recommendations.

Evaluation Methodology (1 out of 3 = Partially Integrated)

As noted, one of the sub-questions referred to gender, and the report states that "the cross-cutting themes of gender, protection and disability were integrated in the interview and focus group discussions." One of the interview questions (Annex 10 – data collection instruments) included a question on whether protection impacts were different for female headed households, but there is no data analysis on the answers to this question. The survey sample of staff included 65% female respondents. However, there does not appear to be gender-disaggregated data flowing from this survey.

Of the 26 focus groups conducted only five were separated by gender: one female only group in Syria; and two female only and two male only in Lebanon,. There is gender disaggregated data on the composition of the focus groups, but not the data from the FGDs. Moreover, the fact that the groups were mixed may have impeded communication by women. The report does not note this as a limitation, nor address how it might have affected the data collected. In many contexts, including in the Middle East, it may be important to ensure that women feel freer to voice their opinions by holding gender-disaggregated focus groups. Pages 97-98 of the UNEG handbook on integrating gender and human rights can provide more detail on how to conduct focus groups, and pages 92-97 suggest other techniques such as purposeful sampling and snowball sampling in order to identify respondents from diverse categories.

Evaluation Findings, Conclusions and Recommendations (1 out of 3 = Partially integrated)

The findings make some reference to the fact that the intervention planned to address gender equality, but there is no discussion as to how the project may or may not have impacted differently on women than on men. There is no discussion of different needs. There is a gender-disaggregated analysis of school enrolment, but for example, there is no such analysis for the emergency shelter component of the programme. Only one of the recommendations has a gender focus: "look into ways to extend outreach to women in Syria to address a decline in the proportion of female clients" using microfinance. The other recommendations, for example those on security and cash assistance, could have also considered the different needs of men and women.

Lessons learned and good practice

Evaluation Scope

It could be interesting as an objective to assess if the programme interventions are reaching men and women and boys and girls equally, for example, or if the education and microfinance elements were reaching all sectors of the population.

Evaluation Criteria and Questions

Although there was one evaluation sub-question that included a gender perspective, some of the Key evaluation questions could have for example asked about the relevance of the programme to women, or the best practices and lessons learned with regard to reaching different sectors of the population.

Evaluation Methodology

It would have been useful to include specific requests for analysis on gender issues, and also requests related to the participation of women in the preparation and conduction of the assessment. For example, the ToR could have included a request to ensure the equal participation of women during the planning meetings and in the assessment team.

The focus groups could have benefited from ensuring that women were not only present in the discussions, but that their voices were heard. Approximately 65% of the survey respondents were women – but there is no disaggregation of the data received by gender

Evaluation Findings, Conclusions and Recommendations

The findings speak of gender, youth, disability and protection as cross-cutting issues, and that gender quality was planned to be addressed by (1) ensuring that girls continue to have access to education; (2) raising the proportion of women accessing microfinance loans; (3) aiming that half of the beneficiaries of the cash assistance and shelter programme in Lebanon should be female-headed households. However, the evaluation questions did not ask questions that would provide answers pertinent to these issues.

The findings did respond to the goal of ensuring that girls would have equal access to education, as there is gender-disaggregated data on school enrolment.

With regard to micro-finance, there is discussion in the findings that the proportion of women clients declined, but there is no gender-disaggregated data on the beneficiaries of the cash and shelter programme. Similarly, the conclusions do not focus on gender or human rights considerations. There is one recommendation to “look into ways to extend outreach to women in Syria” with regard to microfinance.

General

The development of a model of Terms of Reference to be consulted during the planning of every evaluation conducted by UNRWA would be useful. This model should address the integration of gender in the evaluation objectives, criteria (gender as a cross cutting criterion), questions (including examples), and methodology (selection of consultants, data collection, stakeholders to be interviewed, etc.). More focus on gender integration at the TOR stage would steer the evaluators towards more systematic inclusion in the methodology and in the eventual findings and recommendations.

(UNIDO - ESCAP) Summary of the Peer Learning Exercise

Introduction and Background

In 2016, UNIDO produced 30 evaluation reports (26 project evaluations, 3 country evaluations and 1 thematic review). For the purpose of the UNSWAP exercise, though, 2 were considered to be not relevant, one being a mid-term review and one being a technical project evaluation that was deemed non-applicable in terms of evaluability of gender equality/empowerment aspects. A total of 28 reports were therefore included in the UNSWAP reporting.

On the other side, ESCAP produced 8 evaluation reports during the same period, 6 being project evaluations and 2 subprogramme/thematic evaluation. ESCAP considered all 8 evaluation reports for the UNSWAP reporting exercise.

Methodology of the Peer Learning Exchange (PLE)

UNESCO's Internal Oversight Service facilitated and guided the entire exercise which consisted of an initial briefing teleconference in 6 December 2016, a desk review of evaluation reports and a final teleconference to share the findings, challenges and lessons learned on 6 January 2017.

UNIDO and ESCAP exchanged a total of 6 evaluation reports prepared by each agency during 2016. The set of UNIDO reports was selected to provide an overview over the total UNIDO portfolio; one thematic evaluation on the topic of job creation, one country evaluation and four project evaluations covering different thematic areas (energy efficiency, renewable energy, medicinal value chain and POPs).

Concerning UNIDO, the remaining 22 reports were rated internally by IEV Evaluation Officers. Each agency scored the reports of the other and provided substantial comments to justify the score. Two *Skype* meetings were held at the beginning and at the end of the exercise, under the supervision and with the facilitation of UNESCO. During the first meeting an evaluation officer from UNESCO explained the guidelines and the methodology for the peer-to-peer exercise, while the second meeting focused on the exchange of views over the ratings provided by the two Agencies.

Finally, the set of UNIDO projects were rated 8.33 by ESCAP, while the meta-evaluation score for ESCAP projects was 7.51 as rated by UNIDO. Both score *meet requirements*, according to UNSWAP criteria.

Key Observations

The UNIDO IEV team members involved in the exercise strongly believe that the exercise was very useful. Thanks to the peer-to-peer review, in fact, UNIDO IEV was able to test its own rating scale against an impartial external assessment. At the same time, then, the exercise shed a closer light on the current level of GEEW mainstreaming within UNIDO projects. In particular, it was pointed out that while significant progress has been made to include GEEW into the scope and evaluation criteria, the findings, lessons learned and recommendations coming out of evaluation reports are still not reflecting gender analysis satisfactorily in a majority of cases.

The operation of assessing the reports of another agency presented the following challenges:

- It is not easy to assess project reports from specialized agencies whose mandates are not specifically related to Human Rights and GEEW;
- It is particularly problematic to score reports dealing with another agency's divisional or corporate level, instead than a project. In fact, the assessment of a report without an in-depth knowledge of the corporate procedures and guidelines is very challenging and can lead to an imperfect rating scale.

From ESCAP's perspective, the exercise was very useful in understanding the assessment methodology and the four criteria being used, and how they were applied in evaluations of other entities. It also revealed the subjective nature of the exercise and therefore the need to have a common understanding among all parties to ensure consistency and objectivity.

Lessons Learned and Recommendations for the Improvement of the EPI

- As far as possible, pair agencies with similar size and project portfolios;
- A more detailed guidance on what is meant by each criterion would be beneficial to achieve a more coherent rating for the set of projects analyzed;
- While the synthesis report provides some more guidance on what qualifies as *exceeding* criteria, it's sometimes difficult to interpret if a report *approaches* or *meets*, especially on reports which are on very specific technical subjects. A further clarification on this issue would be beneficial for the evaluators;
- The first two criteria are very similar and it can be difficult to differentiate the evaluation scope from the questions and evaluation criteria. Questions could be further refined or the criteria combined;
- The 3rd criterion (*A gender-responsive Evaluation Methodology, Methods and tools, and Data Analysis Techniques are selected*) is very difficult to be assessed, in particular for projects whose focus is not mainly on GEEW. It is very rare to find ad-hoc methodology within the ToR or the report itself;
- Overall, the criteria need to be re-formulated in a better way, with more clear distinction from one another;
- Even when they are included in the report, recommendations and lessons learned do not seem to be based on a particularly deep gender analysis.