



International
Labour
Organization

► Evaluation Office

► *i*-eval THINK Piece, No. 19

Quality assessments of ILO decentralized evaluations: Key results of the quality appraisal 2019 and way forward for the integration of gender equality and empowerment of women considerations into evaluation

Luc Franche, Juan-David Gonzales and
Sophie Pénicaud

December 2020

Sophie



Quality assessments of ILO
decentralized evaluations: Key results
of the quality appraisal 2019 and way
forward for the integration of gender
equality and empowerment of
women considerations into evaluation

Think Piece No. 19

Luc Franche, Juan-David Gonzales, and Sophie Pénicaud

Acronyms

DEFP	Departmental Evaluation Focal Point
EC	Evaluation Consultant
EM	Evaluation Manager
EPI	Evaluation Performance Indicator
EVAL	ILO's Evaluation Office
GEEW	Gender Equality and Empowerment of Women
ILO	International Labour Organization
ILS	International Labour Standards
MOPAN	Multilateral Organisation Performance Assessment Network
OECD-DAC	Organisation for Economic Co-operation and Development's Development Assistance Committee
QA	Quality Appraisal
REO	Regional Evaluation Officer
TOR	Terms of Reference
UMG	Universal Management Group
UN-SWAP	United Nations System-wide Action Plan
UNEG	United Nations Evaluation Group

Contents

Introduction.....	1
1 ILO's Quality Appraisal System and Tool	2
2 Key Results of the Quality Appraisal 2019.....	4
3 Presentation of UN-SWAP-GEEW and Associated Evaluation Performance Indicators	7
4 Key Results on the EPIs from the Quality Appraisal 2019	9
5 Key Areas of Improvement and Way Forward.....	12

Introduction

The objective of this Think Piece is to present and reflect on the quality of ILO's decentralized independent evaluations appraised by Universalialia in the context of the Quality Appraisal (QA) exercise conducted in 2019-20 with a specific focus on the results from the UN System-wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP-GEEW). The focus on the UN-SWAP-GEEW will be informative in light of the upcoming Independent High-Level Evaluation of ILO's Gender Equality and Mainstreaming (2016-2021) that ILO's Evaluation Office (EVAL) will launch early 2021. This focus is also relevant considering that the Quality Appraisal (QA) exercise conducted in 2019-20 is the first one to implement the recent 2018 changes made to the UN-SWAP-GEEW scorecard.

The present document is organized around five main sections:

- **ILO's Quality Appraisal system and tool:** The first section presents the QA approach and tool and describes changes to the tool introduced in 2020.
- **Key results of the Quality Appraisal 2019:** The second section provides a summary of the latest Quality Appraisal Summary Report (produced in July 2020).
- **Presentation of UN-SWAP-GEEW:** The third section presents an overview of the UN-SWAP-GEEW Evaluation Performance Indicators (EPIs) and their use in the UN System and in the ILO as well as an overview of the most recent changes made to the UN-SWAP-GEEW scorecard.
- **Key results on the EPIs of the Quality Appraisal 2019:** The fourth section provides a summary of the latest ILO UN-SWAP-GEEW Summary Report (produced by Universalialia in 2020).
- **Main Areas of Improvements and Ways Forward:** This final section identifies the main areas of improvements regarding the integration of gender considerations in ILO's decentralized evaluation reports. This section also includes possible solutions to address the identified areas of improvements.

Box 1: Overview of ILO's decentralized evaluation system

ILO's evaluation function operates as a completely separate office of evaluation, reporting directly to the Director-General's office, thus guaranteeing the independence of the function. EVAL is the central body responsible for independent evaluations of ILO's strategies, policies and programmes. Regional Evaluation Officers (REOs), Departmental Evaluation Focal Points (DEFPs) and evaluation managers (EMs) report to EVAL regarding the implementation of independent evaluations. Final approval of each decentralized independent evaluation remains the responsibility of EVAL. This structure lays the foundation of EVAL's hybrid decentralized system.

1 ILO's Quality Appraisal System and Tool

A previous [EVAL Think Piece](#) prepared by J.D Gonzales and S. Pénicaud (2019) described ILO's quality appraisal tool. This section offers a short overview of the tool and describes the most recent changes to the tool.

Overview and Implementation of the Quality Appraisal Tool

ILO's quality appraisal tool looks at four different dimensions structured in four sections, allowing the reviewers to collect quantitative data on the quality of ILO's evaluation reports. First, the tool captures descriptive data on demographic variables of each evaluation report, such as the region, department and year. Collected data can consequently be analyzed through the aggregation and identification of trends across these independent variables. The second section of the QA tool requires the reviewers to rate the quality of the content of the evaluation reports according to 58 different items (or criteria) grouped across the 10 standard sections that should structure an evaluation report. Third, the comprehensiveness section of the tool ensures that data is collected on the presence or absence of key components that must be included in the report using a two-point scale (absent-present). Finally, the UN-SWAP assesses four different items, in alignment with the Guidance on Integrating Human Rights and Gender Equality in Evaluation (2014). This is described in-length in section 3 of the document.

In 2020, the ILO's Evaluation Office commissioned Universalia to conduct the quality appraisal of 39 decentralized independent evaluation reports produced worldwide in 2019. The 39 evaluations conducted by ILO's offices included project, thematic, sector and clustered evaluations. The sample included 25 final evaluations and 14 mid-term evaluations from seven departments and covered all regions as well as inter-regional evaluations. The main purpose of the QA was to provide a cumulative analysis of the evaluations submitted in 2019 and assess trends and comparisons with previous quality appraisals. The Quality Appraisal Summary Report informed the ILO's latest Annual Evaluation Report for 2019-2020, which was released in October 2020.

The process was implemented by two reviewers that appraised every single evaluation report to ensure inter-observer consistency. Once all reports of a given reporting period were appraised, quantitative ratings and qualitative information justifying the rating were aggregated in an excel sheet and overall scores were calculated for the quality, comprehensiveness and UN-SWAP dimensions. Aggregated scores and individual ratings were analyzed using quantitative and qualitative methods. Finally, an online survey was disseminated to a sample of evaluation managers.

Changes introduced in 2020

2020 was the first year during which EVAL was able to move forward with the implementation of a rolling quality assurance process. Universalia has been appraising a consistent number of evaluations on a bi-weekly basis to allow EVAL to identify and act upon emerging quality issues rapidly and in a more targeted manner.

EVAL has introduced a few modifications to the QA tool to improve the quality and reliability of the data generated through the QA process. In October 2020, EVAL published an evaluation protocol providing guidance on the inclusion of critical questions in evaluations to measure the mitigation strategies implemented to respond to COVID-19 in interventions evaluated. The protocol emphasises that the Terms of Reference (ToR) of all mandatory project and programme evaluations are expected to include COVID-19 related questions to yield relevant results and actionable recommendation. The QA tool was thus amended by adding specific items to assess the extent to which evaluation reports include evaluation questions that

measure ILO's delivery in mitigating the effects of the COVID-19 pandemic and the extent to which reports identify limitations and constraints associated with conducting evaluation in the context of the pandemic as well as adopted mitigation measures.

In line with the first change identified above, the QA tool was also updated to capture the year in which the evaluation reports were produced to include the month of completion as well. This small addition will allow to identify evaluations that are good in covering Covid-19 and therefore feed into various synthesis and the comprehensive Covid19 evaluation.

EVAl has updated the relevant checklists, which are meant to assist evaluation managers in [preparing ToR for evaluations](#) and rating the quality of evaluation reports. These updated checklists call for evaluation manager to put more emphasis on the way evaluations covered considerations for gender equality and non-discrimination and the SDGs in the evaluation reports. In line with these recent changes, an update was done of the quality items that measure the inclusion of considerations for persons with disabilities and the SDGs in the evaluation criteria and questions. Changes were also made to all the items that measure the inclusion of ILO's cross-cutting policy drivers,¹ clarifying that evaluation reports are expected to include these themes in the evaluation questions.

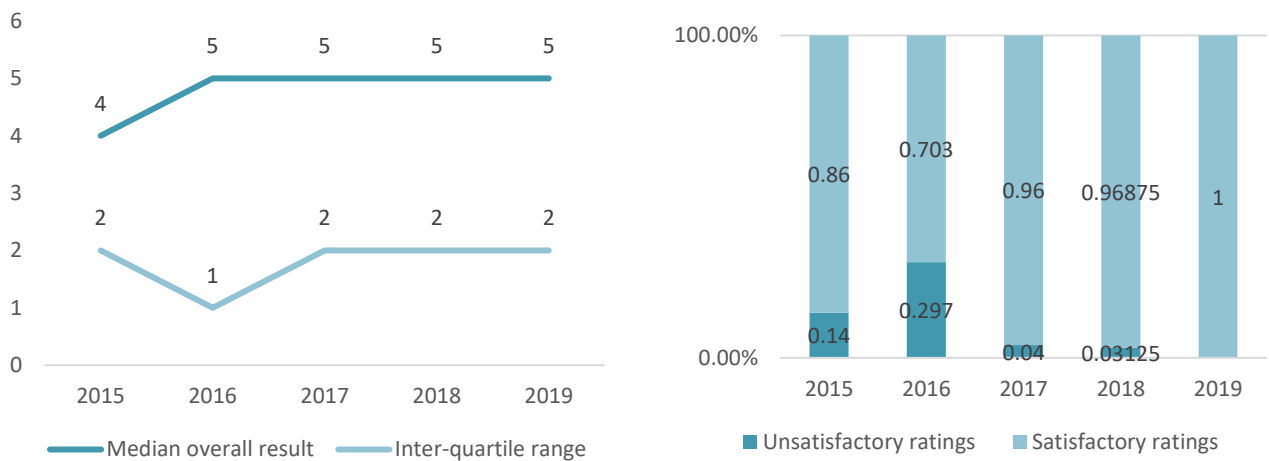
The results from these recent changes will be reflected in the Quality Appraisal (QA) Summary Report to be produced in 2021 on the rolling appraisal of evaluations from 2020.

¹ ILO's cross-cutting policy drivers are International Labour Standards, social dialogue, gender equality and non-discrimination, a just transition to environmental sustainability.

2 Key Results of the Quality Appraisal 2019

The quality of appraised reports has reached satisfactory ratings over the last five years. As illustrated in Figure 1 below, the median score for reports undertaken in a given year remained at 5 since 2016.² The inter-quartile range, which measures the dispersion of results between evaluation reports for a given year, remained stable between 2015 and 2019. Overall, the dispersion of ratings remained low, suggesting a certain homogeneity in the quality of reports over the years. Figure 1 also shows that the proportion of reports that obtained ratings equal or above “somewhat satisfactory” have increased steadily since 2015. While 96% and 97% of reports pertained to that category in 2017 and 2018 respectively, all reports in 2019 received a rating equal or above to “somewhat satisfactory”. In 2015, this result is much lower, with 86% of reports obtaining equal or above “somewhat satisfactory” ratings.

Figure 1. Overall ratings and evolution per year, and proportion of evaluation reports receiving unsatisfactory vs satisfactory ratings per year

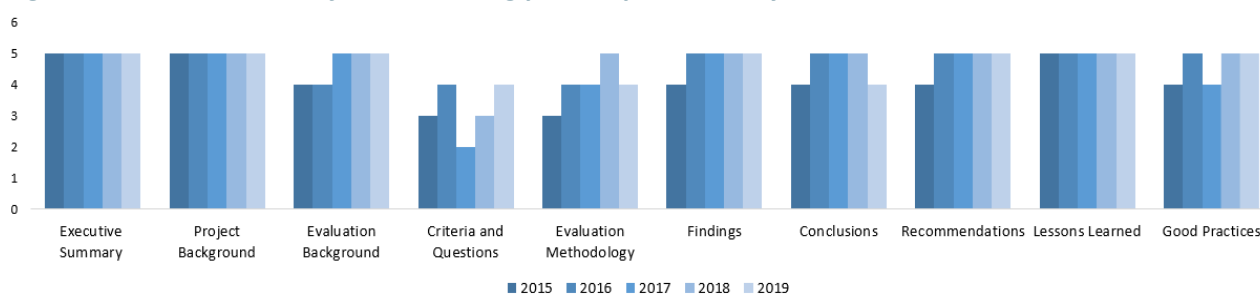


When comparing aggregated results per regions and departments, the reviewers found that evaluation reports obtained a median score equivalent to at least a “somewhat satisfactory” rating regardless of where the evaluations were conducted. Evaluations conducted in the Americas, Arab States, Asia Pacific and Europe, as well as inter-regional offices received a median score considered “satisfactory”; evaluations in Africa received a median score considered “somewhat satisfactory”.

As illustrated in the Figure 2 below, the strongest sections of evaluation reports have been the executive summary, the description of the project background, and the lessons learned since 2015. The findings sections were also positively reviewed and addressed all evaluation criteria in most reports.

² The overall scores are calculated by aggregating the ratings obtained for all items pertaining to the “quality” dimension of the QA, thus excluding the comprehensiveness and UN-SWAP dimensions. The results of the UN-SWAP assessment are presented separately.

Figure 2 Evolution of median rating per component and year



By contrast, the sections that included the evaluation criteria and questions were the only ones for which the median rating per year did not reach the “satisfactory” score since 2015. The inclusion of considerations for ILO’s policy drivers³ remained uneven across evaluation reports, in particular regarding the inclusion of questions on International Labour Standards and the environmental sustainability of the interventions, with 35.9% and 81.1% of evaluations not including evaluation questions in those regards respectively.

Figure 2 also illustrates that the quality of evaluation reports produced in 2019 decreased compared with evaluation reports produced in 2018 with regard to the description of the evaluation methodology and the conclusions. There is room for improvement in evaluation reports through the inclusion of a clear explanation for the rationale for the selection of key stakeholders, other than the ILO constituents which is a prerequisite for all ILO evaluations, with 43.6% of evaluation reports failing to provide any such rationale. The last main area for improvement in reviewed evaluation reports concerned the inclusion of a discussion of how the validity and reliability of the conclusions were determined, with 48.7% of evaluation reports failing to do so.

Observations and Lessons learned

There are a few factors that could explain the overall progression in the quality of evaluation reports since the last IQA (2017-2019). First, **evaluation managers reported via the survey that a few factors could be preventing evaluators from improving their methodology sections, including time constraints, the number of guidelines, the lack of evaluation knowledge and strong methodological skills.** Inception phases are often constrained by tight timeframes, which have consequences on the quality of inception reports’ content. The lack of responsiveness and/or involvement of project staff in the inception phase of independent evaluations was also identified as a common challenge by evaluation managers in this year’s survey. These challenges could explain why there was still room for improvements under several items of the evaluation methodology.

Second, while half of the reports failed to include considerations for International Labour Standards (ILS) in the evaluation criteria and questions, data collected shows an important 16 percentage point increase of satisfactory ratings compared to 2018. **In 2019, EVAL published a new guidance note on ILO’s tripartism and labour standards, designated as “cross-cutting policy drivers” ([Guidance Note 3.2: Adapting Evaluation Methods to the ILO’s Normative and Tripartite Mandate](#)).** **This document might have contributed to improving evaluators’ knowledge on ILS as, according to this year’s survey results⁴, 95% of evaluation managers considered this guidance note to be helpful or extremely helpful.** However, the review team observed there is still room for improvement to better incorporate cross-cutting issues in evaluation matrices and methodologies, especially environmental sustainability, for which 85% of

³ ILO (2020). [Guidance Note 3.2: Adapting Evaluation Methods to the ILO’s Normative and Tripartite Mandate](#)

⁴ Survey respondents were asked to rate from 0 (not helpful at all) to 100 (extremely helpful) the helpfulness of this checklist/guideline (N=21). 57% of respondents (N=12) provided a rating between 50 and 74, and 38% (N=8) provided a rating above 75.

evaluation reports obtained “unsatisfactory” or “highly unsatisfactory” ratings. Only 52% of evaluation managers surveyed considered this guideline to be relevant, 43% considered it clear and 48% considered it applicable, which are relatively low results compared to other guidelines.

Third, the review team observed that in several reports, conclusions appear as a simple summary of findings, which could explain why the conclusions are often merged with the summary of findings in the executive summary. This trend is still valid with 2019 reports, which tends to reduce the quality and value-added of the conclusions.

Finally, the review team noted that clear links between the findings, conclusions and recommendations of the reports are crucial to ensure not only credible recommendations but also realistic and operational recommendations. For example, evaluators can include a table in an appendix presenting these links. Another option is to number all findings and then include these numbers next to each corresponding recommendation.

3 Presentation of UN-SWAP-GEEW and Associated Evaluation Performance Indicators

UN System-wide Action Plan Evaluation Performance Indicators

The EPIs are one of the performance indicators developed as part of the accountability framework of the UN-SWAP to implement the Chief Executive Board for Coordination Policy on gender equality and the empowerment of women. The UN-SWAP EPIs are linked to meeting the gender-related United Nations Evaluation Group (UNEG) Norms and Standards. In particular, the UNEG Norm calls on evaluators and evaluation managers to ensure that human rights and gender equality considerations are respected, addressed and promoted in evaluations. The UN-SWAP EPI serves both as a reporting tool and a benchmark to help UN entities integrate GEEW considerations into evaluations. To guide UN entities, the UN-SWAP framework is accompanied by a set of Technical Notes on each Performance Indicator that provides information on the Performance Indicator as well as guidance on how to complete the rating.

Application of the UN-SWAP EPIs at the ILO

The ILO is one of 69 organizations that are mandated to report against the UN-SWAP-GEEW. Using the UNEG-endorsed Technical Note and related scorecard, the ILO reports against the Evaluation Performance Indicators on a yearly basis on the extent to which GEEW has been mainstreamed in evaluation reports produced by the ILO. In order to do so, the ILO ensures QA consultants rate the EPIs contained in the scorecard in compliance with the instructions found in the UN-SWAP-GEEW Technical Notes.

Since 2015, the ILO has been commissioning external evaluation consultants to review its decentralized evaluation reports against the EPIs, along with the QA tool described above. Teams of evaluation consultants have reviewed a total of 176 evaluation reports.⁵ In 2019, the review team used the UN-SWAP 2.0 scorecard presented in the UN-SWAP Evaluation Performance Indicator Technical Note (April 2018) to generate quantitative and qualitative data on each evaluation report appraised.

The UN-SWAP 2.0 scorecard

UN-SWAP 2.0 scorecard presented in the UN-SWAP Evaluation Performance Indicator Technical Note (April 2018) is structured around the following criteria:

- G.1** GEEW is integrated in the evaluation scope of analysis and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.
- G.2** A gender-responsive methodology, methods and tools, and data analysis techniques are selected.
- G.3** The evaluation findings, conclusions and recommendations reflect a gender analysis.

Using the above-mentioned criteria, the review team rated the extent to which each report integrated gender according to a four-point scale (0-3) that corresponds to the definitions detailed in Box 2.

⁵ More precisely, 43 evaluation reports were reviewed in 2015, 37 in 2016, 25 in 2017, 32 in 2018, and 39 in 2019.

Box 2: Individual rating scale:

0 = Not at all integrated. Applies when none of the elements under a criterion are met.

1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.

2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.

3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

Individual scores were calculated for each evaluation report in accordance with the UN-SWAP Evaluation Performance Indicator Technical Note. The aggregated score obtained by each individual report corresponded to the quality levels scale presented in Box 3. A **meta-score is then calculated** by aggregating and averaging the scores of all individual evaluations appraised for a given year. The meta-score obtained corresponded to the quality level scale in Box 3.

It is important to note that previous UN-SWAP summary reports produced between 2015 and 2018 used the previous UN-SWAP 1.0 scorecard. In 2018, the UN-SWAP 2.0 scorecard was revised, with amended EPIs and rating scales. While the 1.0 scorecard appraised reports according to four different criteria, the 2018 2.0 scorecard appraised reports on three non-corresponding criteria. As a result, meta-scores from the reviewed 2019 evaluation reports are not directly comparable with meta-scores of evaluation reports from previous years, thus limiting the possibility to conduct a temporal comparison of numerical meta-scores and of individual scores obtained for specific criteria.

Box 3: Rating scales**Individual score scale:**

0-3 points = Misses requirements

4-6 points = Approaches requirements

7-9 points = Meets requirements

Aggregated meta-score scale

0 - 3.49 points = Misses requirements

3.50 - 6.49 points = Approaches requirements

6.50 - 9 points = Meets requirements

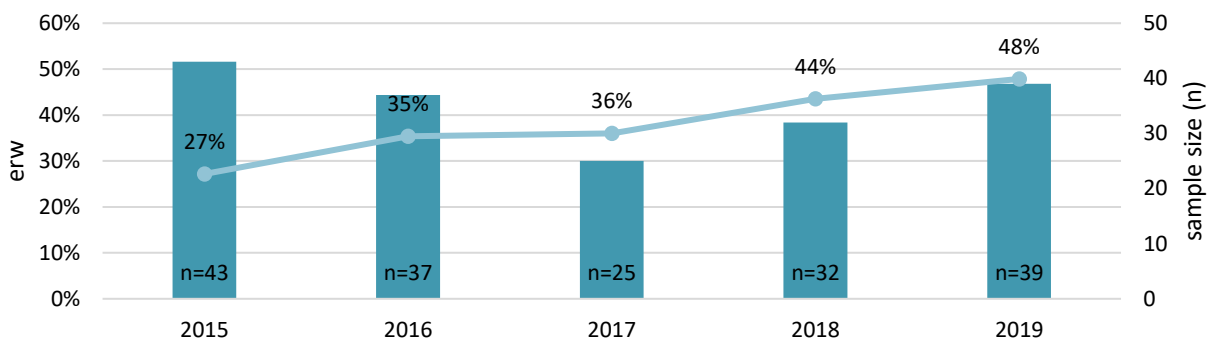
4 Key Results on the GEEW EPIs from the Quality Appraisal 2019

Evolution of ILO’s decentralized evaluation reports’ compliance with the UN-SWAP EPIs

Since the most recent revision to the UN-SWAP scorecard has changed the rating scale to calculate the meta-score from being a 12-points scale to a 9-points scale, and changed the EPIs as well, the adjusted scores are presented below in percentage of the maximum number of points that could be obtained each year. Presenting the meta-scores per year in such way allows a yearly comparison of ILO’s decentralized evaluations’ compliance with the UN-SWAP EPIs.

As seen in Figure 3 below, **ILO’s decentralized evaluations increasingly mainstreamed GEEW considerations in their reports**. The figure shows a 78% increase in the average meta-score obtained between 2015 and 2019 (or a 21-percentage-point increase).

Figure 3 Adjusted meta-scores obtained between 2015 and 2019 (%)



Analysis of ILO’s 2019 decentralized evaluation reports’ compliance with the UN-SWAP EPIs

The meta-score for 2019 evaluations being 4.31 is below the 6.50 threshold needed to “meet requirements”. In other words, in aggregate, evaluation reports produced in 2019 do not quite meet the requirements established by UN-SWAP. **While the majority (24 out of 39) of reports appraised in 2019 “approached requirements”, only 10% (4 out of 39) “met” the UN-SWAP requirements**. The remaining 28% (11 out of 39) “missed requirements” (see Figure 4).

Figure 4 Disaggregated/individual scores obtained in 2019 evaluations

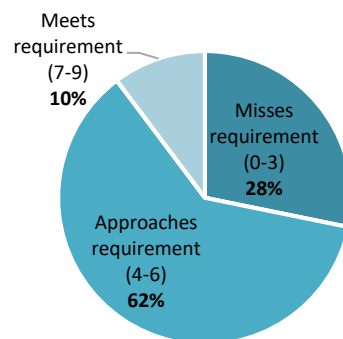


Figure 5 below shows that evaluators performed slightly better in both ensuring evaluation findings, conclusions and recommendations reflect a gender analysis (G3), as well as in

integrating GEEW in the evaluation scope of analysis and evaluation criteria and questions (G1), than they performed in ensuring that gender-responsive methodology, methods and tools, and data analysis techniques are selected (G2).

Figure 5 Disaggregated ratings per criterion in 2019 evaluations⁶

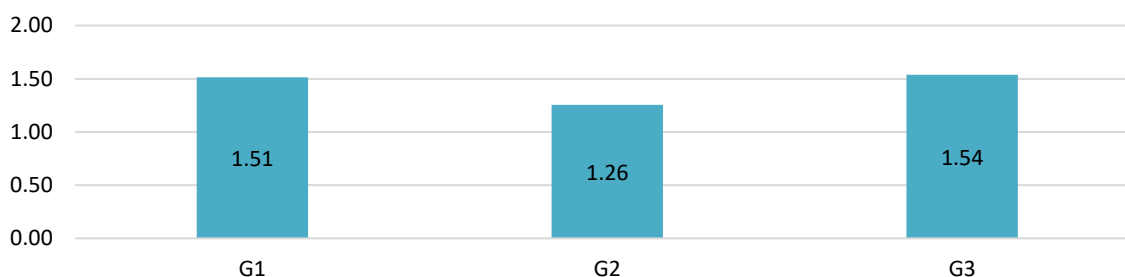


Table 1 below provides a detailed overview of the distribution of ratings attributed across all reports for each criterion. On average, 54% of reports fall on the negative spectrum (0 and 1), while the remaining 46% fall on the positive spectrum (2 and 3). Evaluation reports fall short in integrating gender-responsive methodology, methods, tools and data analysis techniques (G2), as 66% of reports appraised fall on the negative spectrum of the rating scale for that criterion. A slight majority (51%) of reports partially integrated, or did not integrate at all, gender analysis in the findings, conclusions, and recommendations (G3).

Table 1 Percentage of reports' average ratings per UN-SWAP criterion for 2019

UN-SWAP CRITERIA	NOT AT ALL INTEGRATED (0)	PARTIALLY INTEGRATED (1)	SATISFACTORILY INTEGRATED (2)	FULLY INTEGRATED (3)
G.1 GEEW is integrated in the evaluation scope of analysis and evaluation criteria and questions are designed in a way that ensures GEEW related data will be collected	10%	33%	51%	5%
G.2 A gender-responsive methodology, methods and tools, and data analysis techniques are selected	10%	56%	31%	3%
G.3 The evaluation findings, conclusions and recommendations reflect a gender analysis	5%	46%	38%	10%

⁶ Ratings range from "0 – not at all integrated" to "3 – fully integrated"

Figures 6 and 7 below show that in 2019, evaluations managed by a woman evaluation manager (EM) approached or met requirements in 76% of reports. In comparison, 64% of evaluations managed by a man EM approached or met requirements. Regarding evaluation consultants, data shows that no evaluation conducted by a woman met UN-SWAP requirements in 2019, while 24% of evaluations conducted by a man met requirements. Overall, 76% of evaluations conducted by a man approached or met requirements, while 71% of evaluations conducted by a woman approached requirements.

Figure 6 *Proportion of average scores, disaggregated by sex of manager*

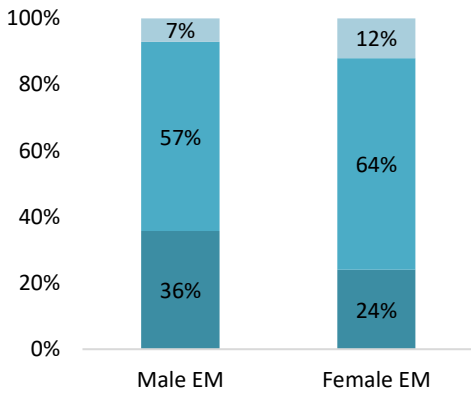
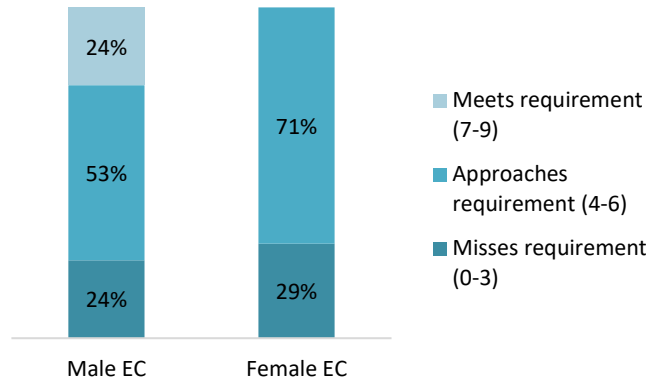


Figure 7 *Proportion of average scores, disaggregated by sex of consultant*



5 Key Areas of Improvement and Way Forward

Based on the observations presented above, and based on the experience acquired by the team of reviewers in conducting the quality appraisal of ILO's decentralized evaluations using the UN-SWAP-GEEW scorecard, following areas of improvements can be identified regarding the integration of gender considerations in ILO's decentralized evaluation reports. This section also includes possible solutions for independent evaluators, evaluation managers and EVAL to address the identified areas of improvements.

5.1 Ensuring that GEEW considerations are addressed comprehensively across evaluation reports

A recurring area for improvements found in evaluation reports is the extent to which **evaluations are not comprehensive when it comes to reporting on gender considerations throughout the entire evaluation report**. This inconsistency in reporting was reflected in two main ways in evaluation reports. First, while most reports included a section describing the evaluation's findings on gender, many reports did so in a separate section rather than addressing gender considerations under all the relevant evaluation criteria. Often, the evaluation findings dealing with gender considerations were found in the effectiveness section of the findings.

Second, the reviewers found **that evaluations that did report well on how gender considerations are integrated in the evaluation scope of analysis, as well as in the evaluation methodology, did not necessarily result in an evaluation report that reflects a gender analysis in the evaluation's findings, conclusions, and recommendations**. The opposite is also true, evaluation reports can include findings, conclusions, and recommendations that reflect a gender analysis, while the evaluation scope and methodology are silent on how gender was considered in the evaluation process.

In order to address these inconsistencies in reporting, four possible solutions can be considered to ensure that evaluation reports comprehensively cover gender considerations across every section of a report:

- **Independent evaluators and evaluation managers should integrate gender considerations in all evaluation criteria.** [The Guidance Note 3.1: Integrating gender equality in monitoring and evaluation](#), produced by EVAL and last updated in June 2020, offers an interesting starting point, since it offers a list of gender-responsive evaluation questions and indicators for each one of the Organisation for Economic Co-operation and Development's Development Assistance Committee (OECD-DAC) criteria.
- In line with the point above, **evaluators should use an evaluation matrix to ensure that all aspects of gender equality are taken into consideration in the evaluation scope.** While gender-responsive evaluation questions can serve to ensure that due consideration is given to gender equality, evaluation indicators can also specify the data that is expected to be collected to ensure proper reporting on gender considerations. In particular, gender-responsive indicators can be used as means of measurement on evaluation questions that are not gender-focussed. On that note, while EVAL's Checklist 3 on writing the inception report does include guidance on completing an evaluation matrix, it does not provide guidance on, nor communicate expectations about, the

inclusion of cross-cutting policy drivers, such as gender, in the evaluation criteria questions, and indicators. EVAL could consider reviewing this checklist accordingly, aligning it with its Guidance Note 3.1 on integrating gender equality in monitoring and evaluation which have been recently updated.

- **Independent evaluators should make sure that they have a good understanding of the gender issues related to the object of the evaluation from the very beginning of the evaluation process.** This can be done by developing an inception report that includes a background section that offers an intersectional analysis of the specific social groups affected by the object of the evaluation or spells out the relevant normative instruments or policies related to gender equality. Such analysis allows the evaluators to enhance their understanding of gender issues related to the evaluation and is likely to increase the quality of reporting on gender considerations in the evaluation findings, conclusions and recommendations.
- When defining the evaluation methodology, **independent evaluators should make sure they purposefully seek the participation of women or other marginalized groups in the evaluation process.** When independent evaluators detail the rationale for stakeholder participation in the evaluation process, they should clearly explain how the sampling addresses the diversity of stakeholders in the intervention, particularly women. They should also develop participatory tools for consultation with stakeholder groups. According to UN Women's Evaluation Handbook on managing gender-responsive evaluations,⁷ a defining characteristic of gender-responsive evaluations is ensuring the inclusion and participation of both women, men and vulnerable groups in the evaluation process. The [UNEG Guidance Document on Integrating Human Rights and Gender Equality in Evaluations](#) also mentions that evidence should be collected with balanced perspectives and fair representation of different points of view. Ensuring the participation of women via participatory tools for consultation could increase the comprehensiveness of evaluation reports in terms of covering gender considerations across all sections of a report. It is also important to note that the duration or time allocated to evaluators to conduct proper data collection with affected populations also has a significant impact on the feasibility of a participatory approach.

5.2 Enhancing the quality of gender analysis in evaluation reports

Another area of improvement identified while reviewing evaluation reports was **the lack of depth and breath of some evaluations' analysis of the gender responsiveness of interventions. An evaluation that reports on equal participation and representation of women and men by simply using gender-disaggregated data is deemed insufficient in terms of reporting gender considerations,** since evaluations are opportunities to assess the extent to which interventions were designed, implemented and reported on in a gender responsive manner.

In order to ensure that evaluation reports assess the extent to which the interventions were informed by an analysis indicating the main gender equality considerations at stake (analytic element) and the extent to which the intervention addressed them (normative element), the following possible lines of analysis

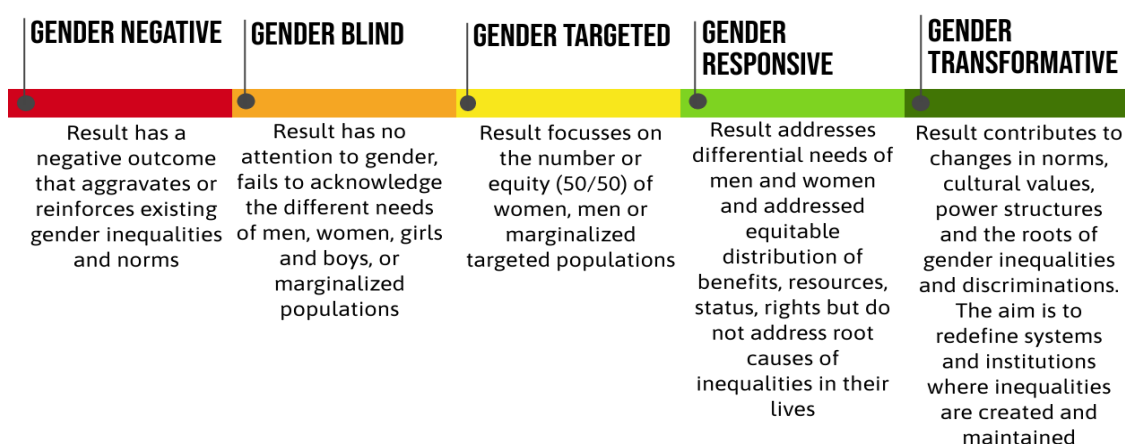
⁷ UN Women, Independent Evaluation Office. 2015. *How to Manage Gender-Responsive Evaluation: Evaluation Handbook*. Available online: <https://genderevaluation.unwomen.org/-/media/files/un%20women/gender%20evaluation/handbook/evaluationhandbook-web-final-30apr2015.pdf?la=en&vs=4246>

could be included by the evaluation managers as part of the evaluation objectives or evaluation questions in the ToR:

- Describe **how gender issues were implemented** as a cross-cutting theme and assess the **extent to which the object of the evaluation prioritized the promotion of gender equality**.
- Assess whether the object of the evaluation paid **attention to the differentiated effects of the interventions on women and other marginalized groups**.
- Include an **analysis of gender and women and marginalized groups** indicating gaps in the capacity of rights holders to claim their rights and of duty bearers to fulfil their obligations. Assess **how the design and implementation of the intervention addressed these gaps**.
- Assess **how the object of the evaluation monitored and analysed results** within a framework (M&E system) that is gender sensitive.

Recognizing that “gender-blind” interventions can have implications on their evaluability under a gender-lens or on the ability of independent evaluators to conduct a proper gender analysis, evaluators can also consider using the Gender Results Effectiveness Scale (GRES) produced by the United Nations Development Programme as criteria for assessing interventions (see Figure 8 below).⁸

Figure 8 Gender Responsive Assessment Scale



Box 4: Definition of Gender Mainstreaming provided in EVAL’s Guidance Note 3.1: Integrating gender equality in monitoring and evaluation

Gender mainstreaming recognizes that gender roles affect and are affected by any project, programme and policy dealing in the social, political or economic realm. Mainstreaming gender has two elements: the analytical, element (understanding power relations between men and women), and the normative element (creating more gender equality so that women and men have equitable access and control of productive resources and benefits). Monitoring and evaluation support these two components by systematically analysing the effects an intervention has on power relations between men and women, and on the goal of creating more gender equality; and by recommending actions to improve the effectiveness of an intervention to address the different needs of women and men and to contribute to greater gender equality.

⁸ UNDP. 2015. Summary: Evaluation of UNDP Contribution to Gender Equality and Women’s Empowerment.

5.3 Enhancing the usefulness of gender-responsive evaluations for future programming

The last area of improvement identified is that **the evaluations' lessons learned and good practices were the two main sections in which evaluation reports tended to report the least on the gender responsiveness of interventions.** While most reports included at least one finding, recommendation or conclusion that reflected gender considerations, only a few evaluation reports included lessons learned and good practices that reflected a gender analysis. This is a significant missed opportunity since lessons learned and good practices are important tools that can contribute to improved project performance and promote organizational learnings. **Gender-responsive lessons learned and good practices could support the definition of strategies and ways forward that can inform the ILO in making sure that future interventions are more gender responsive.**

While EVAL's guidance note provides a background on the definition and usage of lessons learned and good practices for evaluation managers and independent evaluators, the guidance note does not provide indication as to how lessons learned and good practices can be specifically focussed on ILO's cross-cutting policy drivers. EVAL could consider reviewing this guidance note by including possible avenues for lessons learned and good practices regarding gender mainstreaming which are relevant beyond the immediate scope of the project. Revising this Guidance Note 3 could align its content with that of the Guidance Note 3.1 on integrating gender equality in monitoring and evaluation, which does mention that evaluation reports can include lessons learned and good practices regarding gender mainstreaming.

5.4 Strengthening a shared understanding and usages of available guidance documents on preparing gender-responsive evaluations

The possible solutions provided above are in line with one of the principles outlined in the ILO results-based Evaluation Strategy 2018–21, which states that evaluation approaches methods and framework should be participatory and people-centered, inclusive of disadvantaged workers, human rights and gender equality and adapted to the ILO's specific mandate and context. EVAL has made significant progress in recent years in ensuring that this evaluation principle is applied in practice. With the recent update (4th edition) from EVAL of the ILO [policy guidelines for results-based evaluation](#) and more detailed guidance notes, these guidance documents call for evaluation managers to put more emphasis on the way evaluations include considerations for gender equality, non-discrimination and the SDGs in the evaluations' executive summaries, the description of the project background and evaluation background, methodology, and findings. In 2020, EVAL also updated its [Guidance Note 3.1 on integrating gender equality in monitoring and evaluation](#).

While the updates to these guidance documents will help address the main areas of improvement identified above, the fact that evaluation managers almost systematically share the guidance documents produced by EVAL with independent evaluators suggests that their inconsistent application in evaluation reports may not be due to dissemination issues. Instead, **evaluation managers and independent evaluators may not have a shared understanding of what the guidance documents entails and how they can be readily applied in a concrete evaluation context.**

A possible way to strengthen a shared understanding and application of the guidance documents could be to include a more exhaustive overview in the current training module of the Evaluation Manager Certification Programme developed and implemented by EVAL. EVAL could also consider sharing the guidance documents produced by other organizations more systematically, such [UNEG Guidance Document on Integrating Human Rights and Gender Equality in Evaluations](#), the [UN-SWAP Evaluation Performance Indicator Technical Note](#), [UN Women's Evaluation Handbook on Managing Gender-Responsive Evaluations](#) and [UN Women Guide for the Evaluation of Programmes and Projects with a Gender, Human Rights and Intercultural Perspective](#). In particular, the UN-SWAP Technical note is a short user-friendly document that includes annexes that can be used as short fact sheets that synthesize how independent evaluators are expected to address gender considerations in the evaluation reports.
