



i-eval THINK Piece, No. 13

ILO EVALUATION QUALITY: CHALLENGES AND POTENTIAL STRATEGIES

REFLECTIONS BEYOND THE QUALITY
APPRAISAL

By Salvador Bustamante, Estela López & Inmaculada Román
Edited by C. Russon

International Labour Office
Evaluation Office

June 2018



ILO EVALUATION QUALITY: CHALLENGES AND POTENTIAL STRATEGIES

REFLECTIONS BEYOND THE QUALITY APPRAISAL

By Salvador Bustamante, Estela López & Inmaculada Román
Edited by C. Russon

June 2018

The responsibility for opinions expressed in this document rests solely with the authors. The publication does not constitute an endorsement by the International Labour Organization.

This document has not been subject to professional editing.

International Labour Office
Evaluation Office

Table of contents

ACRONYMS..... 4

INTRODUCTION 5

1. WHAT ARE THE MAIN RESULTS OF THE QUALITY APPRAISAL? 5

2. WHAT ARE THE MAIN CAUSES CONSIDERED AFFECTING THESE RESULTS?..... 7

3. WHAT STRATEGIES COULD BE CONSIDERED FOR THE FUTURE? 7

ANNEX 1..... 8

ACRONYMS

EM	Evaluation Manager
EMCP	Evaluation Manager Certification Programme
EVAL	ILO Evaluation Office
GEEW	Gender Equality and the Empowerment of Women
IEE	Independent External Evaluation
ILO	International Labour Organization
M&E	Monitoring and Evaluation
P&B	Programme and Budget
QA	Quality Appraisal
RBM	Results-Based Management
ToR	Terms of Reference
UN-SWAP	United Nations System-wide Action Plan
UNEG	United Nations Evaluation Group

INTRODUCTION

The present Think Piece was commissioned by the Evaluation Office (EVAL) of the International Labour Organization (ILO). The document reports the findings and conclusions from a Quality Appraisal (QA)¹ of independent evaluations that were submitted to the ILO from January 2015 to March 2017. QAs are performed by EVAL on a regular basis to measure progress towards improving the quality of the ILO's evaluation function.

With this purpose in mind, the QA 2015-2017², carried out by the authors of this paper, included a quality assessment of 86 evaluation reports. It also featured an analysis of the alignment of these evaluations with the United Nations System-wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP-GEEW). In addition, the appraisal analysed the evaluation managers' level of compliance with ILO guidelines and policies. Finally, the QA presented findings and recommendations to enhance the QA tool itself.

The purpose of this Think Piece is to go beyond a mere presentation of the evidence and to speculate about the root causes of the results that were obtained. To deliver its goal, the document is structured into three sections, within which ideas are developed for the Theory of Change. These sections are as follows:

- The QA's main results;
- Hypotheses surrounding the causes that might have contributed to the results; and
- Potential strategies that might address limitations and boost evaluation quality.

1. WHAT ARE THE MAIN RESULTS OF THE QUALITY APPRAISAL?

Before presenting the results, it is worth noting that, in advance of the study, EVAL made a number of important changes to its quality appraisal system. The changes were informed by the review of the quality appraisal systems of similar UN agencies, programmes, and funds and by the analysis of the results of past quality appraisals.³ In some respects, this quality appraisal was viewed as a pilot test.

The quality appraisal⁴ found that more than ninety per cent of the reports assessed met the minimum level of overall quality. The components that were rated as satisfactory were the executive summary, project background, evaluation background, findings, recommendations and lessons learned. However, other components such as criteria and questions, methodology, conclusions and good practices received lower ratings and this attenuated the overall quality. Despite this limitation, on the whole, the findings are considered to be satisfactory.

Since 2011,⁵ the Evaluation Office has undertaken a number of initiatives to improve the quality of evaluation reports. These initiatives include the introduction and regular updating of Evaluation Policy Guidelines, an Evaluation Manager Certification Programme and a gradual increase in consultancy rates to hire better qualified evaluators. The appraisal explored the effectiveness of some of these initiatives.

¹ Bustamante, S; López, E & Román, I. 2017. *Quality Appraisal of ILO Evaluations (2015-2017)*.

² Key findings, conclusions and recommendations of the QA are included in its executive summary.

³ Llabres, Iker, 2017. *Evaluation quality assessment methodology in the UN system and changes to the ILO's quality appraisal methodology. i-eval THINK Piece, No. 12*.

⁴ The complete QA report can be consulted at the ILO's Evaluation Office.

⁵ This coincided with the start of evaluation strategy 2011-2015 that called upon evaluation to strengthen knowledge development and accountability in the areas of decent work, international labour rights and standards and the ILO Declaration on Social Justice, as well as to enhance the relevance and usefulness of evaluation to constituents.

The appraisal found that Evaluation Managers (EM) considered the Evaluation Manager Certification Programme (EMCP) to have provided them with appropriate competencies to guide evaluations. In addition, the EMCP increased their understanding of the ILO's evaluation guidelines and policies. All of the EMs who took part in the EMCP concurred with its high quality and with the usefulness of its content.⁶

As Evaluation Managers' understanding increased, so did their use of ILO evaluation guidelines and policies. The analysis revealed that Evaluation Managers reported sharing the Evaluation Policy Guidelines with evaluation consultants more frequently. Also, Evaluation Managers reported providing more comments, based on the Guidelines, to draft evaluation reports.

There were some findings that appeared to be counter-intuitive. For example, the Evaluation Office has increased rates under the assumption that it would enable the ILO to hire better qualified consultants and that this, in-turn would ultimately lead to better quality reports. Analysis did not support the assumption. The correlation between evaluation budget and quality was not significant, nor was the correlation between consultants' years of experience and quality.

The one factor that did correlate significantly with quality was the duration of the evaluation. Evaluations that had a longer duration correlated positively with report quality. This finding was triangulated with data from the Evaluation Managers who expressed concern about the timeframes, as well as access to relevant information and sources of information.

Gender

All UN agencies, programmes and funds are required to report on the results of their efforts to mainstream gender into evaluations. Most UN organisations comply with this requirement by conducting self-assessments. This practice, however, is subject to self-reporting bias. The ILO Evaluation Office is one of a small number of evaluation entities in UN organisations to systemically hire outside consultants to evaluate its gender mainstreaming efforts. The ILO Evaluation Office bundles this together with its quality appraisals.

Despite the positive effort described above, analysis showed that, with regard to the ILO's efforts to mainstream gender into evaluations, there is ample scope for improvement. The overall average Gender Equality and the Empowerment of Women (GEEW) rating has been consistently low but was well below the satisfactory level in the latest quality appraisal. In fact, only 8% of the evaluation reports (seven in total) fully met the UN-SWAP requirements.

The authors speculated at the reasons for the retrogression. Gender is not among the criteria for selecting consultants to conduct independent evaluations. Gender issues are often not adequately addressed in project designs. The ILO generally lacks a comprehensive monitoring and reporting system to track progress in gender equality. As a result, gender concerns may be poorly represented in the evaluation designs contained in the ToRs and the final report.

⁶ In 2015, EVAL commissioned an impact assessment of the EMCP on the quality of evaluations conducted for the International Labour Organisation (ILO). The results showed no statistically significant difference between the average quality of evaluations produced under managers who had completed the EMCP versus those who had not completed the EMCP. Secondary analysis later showed that the quality of the independent evaluation reports submitted to EVAL over time may not have gone up, however, the comprehensiveness or completeness of the reports certainly did—which can be considered to be an improvement.

2. WHAT ARE THE MAIN CAUSES CONSIDERED AFFECTING THESE RESULTS?

In the previous section, the highlights of the quality appraisal report were presented. In this section, the authors go beyond a mere presentation of the evidence and speculate about the root causes of the results that were obtained. The point of departure for this discussion is the assumption that the various components of an evaluation report should be aligned with each other. The purpose of the evaluation, in large measure, determines the criteria and questions. The criteria and questions influence the choice of methods. The methods, in turn, lead to the findings.

Alignment of report components

In the previous section, it was noted that criteria and questions, in addition to methodology were both rated as being less than satisfactory. This makes perfect sense to the authors. When the Terms of Reference were examined, the authors often found large numbers of evaluation questions. The authors are of the opinion that there should be two or three questions per criteria—no more. It seems to the authors that the proliferation of evaluation questions may have confused the choice of methods.

In the previous section, it was also noted that findings were rated as being satisfactory while the ILO's evaluation methodology was not. This is a conundrum to the authors. If we accept the proposition that findings are derived from methods, then it should not be possible to obtain satisfactory findings using flawed methods. If the methods are flawed, the findings should, per force, be of low quality. However, that appears not to be the case.

The only explanation that the authors can come up with is that, perhaps, maybe the ILO's methods are not substandard, after all. Perhaps it is the *description* of the methodology that stands in need of improvement. The recommendation that would follow from this logic is that the ILO needs to find a way to ensure independent evaluators better describe their methods. It is possible that the tools already exist to help independent evaluators to do just that.

Among the guidance notes, templates and checklists that the ILO Evaluation Office provides to independent evaluators is Checklist 4, Validating Methodologies (Annex 1). The checklist is based on information contained in the Joint Committee on Educational Evaluation Standards, the UN Evaluation Groups Standards and the OECD DAC Evaluation Quality Standards. The authors suggest that emphasizing the use of this short checklist might help the ILO to leverage a disproportionate improvement of quality.

3. WHAT STRATEGIES COULD BE CONSIDERED FOR THE FUTURE?

The following section presents certain strategies considered essential to address the main causes affecting evaluation quality and to reinforce the identified positive results. In general terms, the evaluation culture of the ILO should transcend enforcing the evaluation policies and guidelines, which are already considered comprehensive and useful, to helping staff to internalise their contents and principles. This process could be facilitated by transitioning from purely intrinsic incentive structures to finding ways to incentivise staff extrinsically. From this inspirational approach, the Think Piece presents a number of specific strategies that could be explored to improve evaluation quality.

1. Strengthen incentive structures

The implementation of a supportive and rewards-based system for ILO staff is recommended with a view to bolster innovative action for the continuous improvement of evaluation quality. These actions should

be specifically oriented towards reinforcing Departmental Evaluation Focal Points, Evaluation Managers and internal evaluators. A supportive system comprising capacity building, mentoring, technical and methodological guidance with respect to evaluation should be put in place. This system should, among other strategies, promote horizontal learning among participants. The incentive structure would acknowledge staff by means of incentives, such as internal promotion, recognition awards, publications, etc. The new [evaluation strategy of the ILO](#) already includes plans to review how incentives for ILO volunteers involved in evaluation can be improved.

2. Put gender at the centre of the ILO's work, considering an inter-sectorial approach to discrimination

In order to address the challenge of providing valuable, meaningful and relevant gender information through the ILO's independent evaluations, it might be necessary to effectively put gender at the centre of the ILO's mission and work. This path of action has recently been initiated by the agency through the adoption of a new gender mainstreaming procedure by which all gender contributions must be identified, monitored and evaluated. However, this measure is not expected to produce tangible results in the short term.

3. Guarantee monitoring quality

The agency should ensure compliance with a standardised monitoring and evaluation (M&E) system at project level. The agency works with outputs and outcomes that mainly reflect the overall work carried out in the realm of policy development, capacity building and partnership generation. This shared framework across interventions should be reflected in the M&E system. Obviously, the M&E system will need to be adapted to the particular requirements and contextual factors that affect each project, but the shared framework will enable data comparability and homogeneity to a certain extent. In addition, this common framework would also simplify M&E activities across interventions.

4. Ensure strong evaluation capacities in consultants

Consultants should play an essential role in boosting evaluation quality. In this regard, the ILO should implement a strategy to leverage their capacities and guarantee the selection of competent evaluation professionals. This could be achieved by ensuring that consultants hold specific and formal evaluation training and master a wide range of evaluation approaches, methods and techniques. Moreover, consultants should demonstrate adequate gender expertise and/or incorporate a gender specialist within the team. These should be considered as important as the sectorial and geographical expertise presented. Consultants should understand that gender skills are not only complementary, but are also a requirement that must be met to work with the agency. (To enhance consultants' understanding of how to conduct evaluation for the ILO, EVAL has recently developed an on-line, induction programme.)

5. Enlarge and improve the QA approach

Finally, it might be interesting to include meta-evaluative case studies in the QA approach. The execution of in-depth analysis to determine all the elements that significantly affect the report quality may provide meaningful information. This would allow appreciations regarding general quality to be transcended, by uncovering the 'black box'. This concept refers to a commonly forgotten inquiry into how key elements and processes provide added value to achieve outputs, outcomes and impact. Moreover, this work would provide the opportunity to jointly analyse each evaluation report with its corresponding programme planning document (PRODOC), its baseline, and even the mid-term and final evaluation reports, to trace the evidence from one report to another. The evaluation's coherence could be better understood by examining the contributions and logics exchanged between all these documents.

Annex 1

Validating Methodology⁷

The information contained in this checklist can guide evaluation managers on how to validate methodological soundness of an evaluation proposal.

☐ *Clear Project Description*

Having a clear description of the project being evaluated is important because it permits stakeholders to determine if the proposed methods are appropriate.

☐ *Described Purposes and Procedures*

Clearly stated purposes and procedures permit stakeholders to determine if there is good alignment between the purpose of the evaluation and the proposed methods. In the ILO, evaluations are usually conducted in order to demonstrate accountability to donors for funding received and in order to promote organizational learning.

☐ *Evaluation Criteria*

The evaluation design should clearly spell out the criteria against which the project to be evaluated will be assessed. The OECD DAC Criteria of relevance, efficiency, effectiveness, impact and sustainability guides evaluations in the ILO.

☐ *Methods Answer Evaluation Questions*

The methods selected should permit information to be collected that address pertinent questions that stakeholders want the evaluation to answer.

☐ *Valid and reliable evaluation methods*

Valid and reliable evaluation methods are free from a consistent alignment with one point of view that results in a lack of objectivity, fairness, or impartiality. To ensure validity and reliability, a combination of methods such as document review, interviews and direct observation are often used.

☐ *Methods are Appropriate for the Sources of Information*

The evaluation methods selected should be appropriate for the sources from which information will be collected. For example, in Muslim cultures, it may not be appropriate for a man to interview an unaccompanied woman.

☐ *Complete and Fair Assessment*

Evaluation methods should be implemented in a rigorous manner to permit strengths and weaknesses of the project to be identified in complete, fair and unbiased manner.

☐ *Quantitative and Qualitative Data Analysis*

The techniques for analysing quantitative information include descriptive and/or inferential statistics. The techniques for analysing qualitative information include content analysis (i.e., determining patterns, categories, taxonomies, themes, etc.)

☐ *Cost-effectiveness*

The methods selected should be efficient and produce information of sufficient value that the outlay of resources can be justified.

⁷ This ILO checklist is based on information contained in the Joint Committee on Educational Evaluation standards, the UN Evaluation Groups Standards and the OECD DAC Evaluation Quality Standards.