i-eval THINK Piece, No. 4

Leveraging Appraisal Findings to Improve Evaluation Quality

Kelly Robertson

Dr. Daniela Schröter

International Labour Office
Evaluation Unit

February 2014

Leveraging Appraisal Findings to Improve Evaluation Quality

Kelly Robertson
Dr. Daniela Schröter

TABLE OF CONTENTS

BACKGROUND AND CONTEXT

EVALUATION AT THE INTERNATIONAL LABOUR ORGANIZATION

The centralization of the evaluation function within the ILO was established by the Governing Body's 2005 evaluation policy,[1] which initiated the formation of the central evaluation unit, EVAL. Following an independent assessment of the ILO evaluation function in 2010, the policy was revised in 2011 to conform to international standards of good practice. The ensuing *ILO Results-Based Evaluation Strategy for 2011-15* complements the policy and is intended to guide current practice, which adheres to the Organization for Economic Co-operation and Development's (OECD) Development Assistance Committee (DAC) principles and the United Nations Evaluation Group's (UNEG) norms and standards. The ILO utilizes evaluation as a management and organizational learning tool that supports the governance and oversight of the Governing Body, as well as, program and policy improvements that further the ILO's decent work and social justice initiatives. Moreover, evaluation is viewed as an essential component of ILO's results-based management and is "a critical means to improve decision-making, generate knowledge, and provide verifiable evidence of effectiveness."[2] Further information about the context and logic for evaluation within the ILO can be found in the *ILO Policy Guidelines for Results-Based Evaluation: Principles, Rationale, Planning, and Managing for Evaluations.*

To assure consistent quality and credibility of evaluation reports, ILO EVAL has managed annual appraisals of independent project evaluation reports since 2006. In addition, the ILO conducts independent external evaluations of the ILO evaluation policy and overall evaluation function every five years. The 2010 *Independent External Evaluation of the ILO Evaluation Function* indicated that ILO evaluation policy was sound and that significant improvements were made to the evaluation function between 2005 and 2009. Furthermore, during the same period, while room for improvement was noted, the quality and consistency of independent project evaluation reports improved, meeting almost all international standards of good practice. Relatedly, as will be discussed in more detail, quality appraisals of independent project evaluation reports produced between 2008 and 2011—conducted by Western Michigan University's Evaluation

[1] The Reference section of this report lists complete bibliographic information for documents referred to herein.
[2] ILO. *ILO Policy Guidelines for Results-Based Evaluation* (Geneva, 2012) p.3.

Center—found that the number of required items included in the reports increased. The quality of report contents, however, remained roughly unchanged. The plateau in evaluation quality is primarily a result of low quality in areas requiring improvement (e.g., evaluation background and methodology). Many of these components identified as needing improvement in the most recent appraisal were also cited as areas for improvement in prior appraisals. Suggestions for different approaches to address these improvements are presented in the section *Leveraging Findings from the Appraisals* on page 16.

APPRAISAL BACKGROUND AND METHODOLOGY

Western Michigan University's Evaluation Center conducted two quality appraisals of ILO independent project evaluation reports. The first appraisal was conducted in 2009 and examined a sample of ILO's 2008 reports. The second appraisal focused on a sample of ILO reports produced between 2009 and 2011. Both appraisals followed ILO's policies and procedures under specific consideration of UNEG evaluation norms and standards and the OECD/DAC criteria. Additionally, the second appraisal followed *ILO Policy Guidelines for Results-based Evaluation* published in 2012, which includes *Checklist 5: Preparing the Evaluation Report*[3], which was not available during the first appraisal. The team also complied with the Joint Committee on Standards for Educational Evaluation's (Yarbrough et al., 2011) *Program Evaluation Standards* and the American Evaluation Association's (2004) *Guiding Principles for Evaluators*, which have been adapted in many evaluation communities around the globe.

The methodology used in both appraisals was largely prescribed by ILO EVAL. Changes made in the instrumentation used for the second appraisal, reflected reporting requirements outlined in ILO's *Checklist 5.* Changes were made with consideration of preserving comparability and consistency of items assessed in the first appraisal conducted by the Evaluation Center in 2009. As often the case, primary methodological limitations of the appraisals stemmed from time and resource constraints.

The key evaluation questions included the following:

1. What proportion of evaluation reports addressed key reporting components and related items required by ILO policy?

[3] ILO. *Checklist 5: Preparing the Evaluation Report (Geneva,* 2013). [Supplemental document to the *ILO Policy Guidelines for Results-Based Evaluation* (2012)].

2. How well are the key reporting components and related items addressed within the evaluation reports?

The primary evaluation criteria corresponded to *Checklist 5.* Data sources included a sample of 60 per cent (*n* = 93) of independent project evaluation reports produced between 2009 and 2011 and a census of all identifiable evaluation managers who were responsible for overseeing the sampled reports. Reports were selected utilizing a probability to proportional size sampling technique and drawn in a way that, when possible, 60 per cent of reports from each year, region, type (e.g., interim/mid-term or final), and language were selected.

> On average, 69 per cent of evaluation items were addressed within the reports.

Independent reviewers systematically reviewed the sampled evaluation reports and scored items on a set of quantitative scales that focused on the existence/inclusion of report components, related items and their quality. Components and items included in the *Checklist 6: Rating the Quality of Evaluation Reports*[4] largely corresponded to the items included in *Checklist 5*, in addition to items that were present in the 2008 quality checklist and input from EVAL. Similar to past appraisals, quantitative data (scores) were analyzed using descriptive statistics for report components and items. When possible and reasonable, aggregated data were examined by report year and region. Qualitative data were thematically analyzed and used to explain quantitative ratings. Manager survey data were analyzed and integrated with the findings from the document review. Data from different sources were triangulated where possible.

FINDINGS FROM THE 2009-2011 QUALITY APPRAISAL

EVALUATION QUESTION 1: WHAT PROPORTION OF EVALUATION REPORTS ADDRESSED KEY REPORTING COMPONENTS AND RELATED ITEMS REQUIRED BY ILO POLICY?

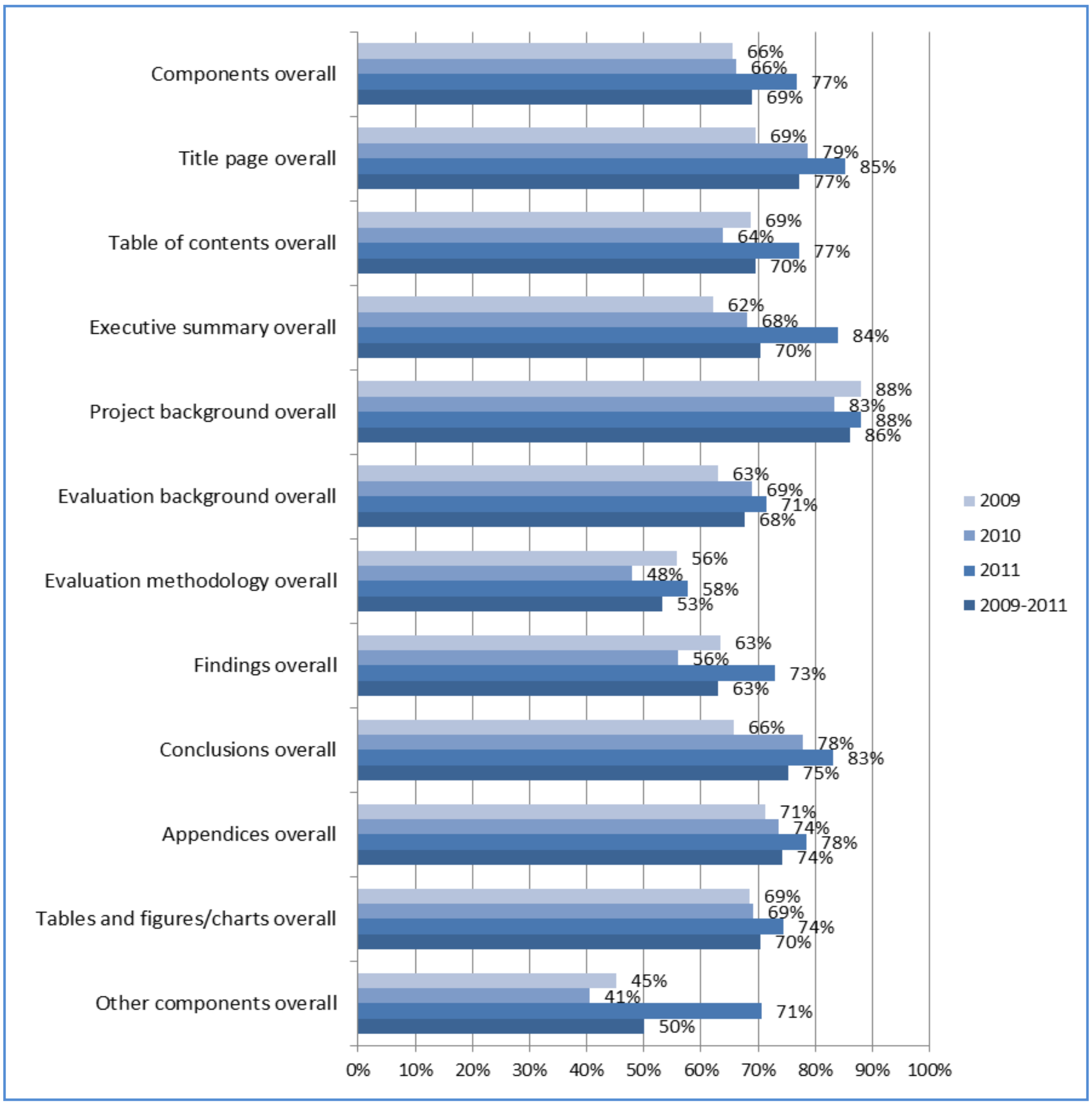
Reports were reviewed for inclusion of items in 11 key report components, namely title page, table of contents and lists (e.g., tables, figures, and acronyms), executive summary, project background, evaluation background, methodology, main findings, conclusions,

[4] See Annex A of this report.

appendices/annexes, tables and figures/charts, and other components. Each component had corresponding specific requirements. For example, the methodology component should include: evaluation questions, evaluation criteria, indication of the use of ethical safeguards to protect human subjects, data sources, sampling procedures, data collection methods, instruments, methods for data analysis, methodological limitations, and a description of and rationale for stakeholder participation.

Each item within a component was scored with either 0 for not provided, 1 for provided, or 99 for not applicable. As shown in Figure 1, the percentage of items addressed within each component increased between 2009 and 2011 with the exception of project background, in

Figure 1. Average percentage of items addressed in reports by year



which the same percentage of items were addressed in 2009 and 2011 (88%) despite dropping slightly in 2010 (83%). On average, the sampled evaluation reports included 69 per cent of items across the years. The reports most frequently included the items required for the project background section (86%), title page (77%), conclusions (75%), appendices/annexes (74%), table of contents and lists (70%), executive summary (70%), and tables and figures (70%). Items required in the evaluation background section (68%), findings (63%), evaluation methodology (53%), and for other components (50%) such as acknowledgements, affiliation of evaluator(s), and the type of evaluation were least frequently included. Notably, sampling procedures (12%), data collection instruments (30%), methods for data analysis (36%), stakeholder participation (46%) and rationales for stakeholder participation (9%)—all of which are items within the evaluation methodology section and vital aspects in the assessment of evaluation quality and credibility of findings—were more often missing than discussed within the sampled reports. Similarly, cross-cutting issues (26%) within the findings section and good practices (38%) within the conclusions section were discussed in less than half of the reviewed reports. Furthermore, data collection instruments (34%) were not usually appended to reports and data were not typically disaggregated by subgroups (e.g., sex, ethnic group, etc.) in tables and/or figures/charts when relevant (32%). Interestingly, similar proportions of reports disaggregated data in tables and/or figures/charts in 2009 (24%) and 2011 (25%), despite increasing significantly in 2010 (44%).

EVALUATION QUESTION 2: HOW WELL ARE THE KEY REPORTING COMPONENTS AND RELATED ITEMS ADDRESSED WITHIN THE EVALUATION REPORTS?

Once the existence of key items within given report components were established, raters scored the quality of the contents presented for the executive summary, project background, evaluation methodology, findings, conclusions, recommendations, lessons learned, good practices, and formatting items (i.e. references). Items were scored as 0 for unacceptable quality, 1 for insufficient quality, 2 for acceptable quality, and 3 for high quality. A score of 88 was applied when insufficient information was presented to judge quality of an item. A 99, for not applicable, was applied when the corresponding component rating was 0 (i.e. the item was not addressed in the report). As in past quality appraisals, scores were averaged across items and components. A cut score of 1.5 presented an *acceptable* level of quality and a cut score of 2.5 a *high* level of quality.

As shown in Table 1, while the included components in the majority 444of sampled reports were considered to be of acceptable quality overall, none of the components received an average score above the threshold for high quality (2.5); with the two most highly rated components being project background and conclusions. The project background section was most often rated as 1 (mode), indicating that the largest number of reports received an insufficient quality rating. For the conclusions section, it is important to note that the quality of conclusions was not assessed in terms of the validity of findings.

Average ratings suggest that the quality of the contents presented is acceptable overall.

Table 1. Report component quality rating descriptive statistics

	Mean	Median	Mode
Quality Overall	1.8	2	2
Executive Summary	1.9	2	2
Project Background	2.0	2	1
Evaluation Background	1.9	2	3
Evaluation Methodology	1.7	1	1
Findings	1.7	2	2
Gender Mainstreaming	1.7	2	2
All data is sex-disaggregated, where feasible	1.2	1	0
Conclusions	2.1	2	2
Recommendations	1.3	2	2
Lessons Learned	1.9	2	2
Good Practices	1.9	2	2
Formal Items	1.9	2	2

Rather, conclusions were assessed in terms of whether they were based on the findings section and whether judgments were fair and impartial. That means future appraisal tools should include validity of findings and conclusions under specific consideration of the evaluation methodology and related limitations. While the evaluation methodology section received an acceptable average score (1.7), only about half of the required items were included in the methodology section (see components section), and of those, more than half had ratings of insufficient quality, which has significant implications for the validity of conclusions and related recommendations. Further, while more than half of reports had ratings of acceptable quality, the recommendations section was the only component that received an average rating of insufficient quality overall. This can be attributed to the fact that many of the recommendations sections inadequately specified time frames, action needed to remedy the situation, priority or importance level, and acknowledgement of resource implications. Because of its importance within ILO policy, the two gender-focused items—located in *Checklist 5*—are highlighted here, namely gender mainstreaming and sex-disaggregation of data. While gender mainstreaming received an acceptable score overall (1.7), sex-disaggregation of data received an average score of insufficient quality (1.2) and as shown in Table 1, was most frequently scored as unacceptable.

KEY AREAS FOR IMPROVEMENT

Key areas for improvement identified in the most recent quality appraisal, and verified as longstanding needs in past appraisals, are identified and described in this section. These areas for improvement are framed in terms of credibility and use of evaluation, two interrelated concepts that serve as integral aspects of evaluation within the ILO. The OECD/DAC definition of evaluation adopted by the ILO states that "an evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of other recipients and donors."[5] In terms of credibility, the 2010 *Independent External Evaluation of the ILO Evaluation Function* states:

> Credibility of evaluation is produced by different factors. One is that the independence of the evaluations is accepted. A second is that the methods used in the evaluation meet the criteria of having valid indicators and reliable collection of data. Credibility is also assured if all the stakeholders are brought into the process. EVAL, once its independent position is secured, is in a position to provide that kind of quality assurance, based on the experience it has acquired over the past five years.[6]

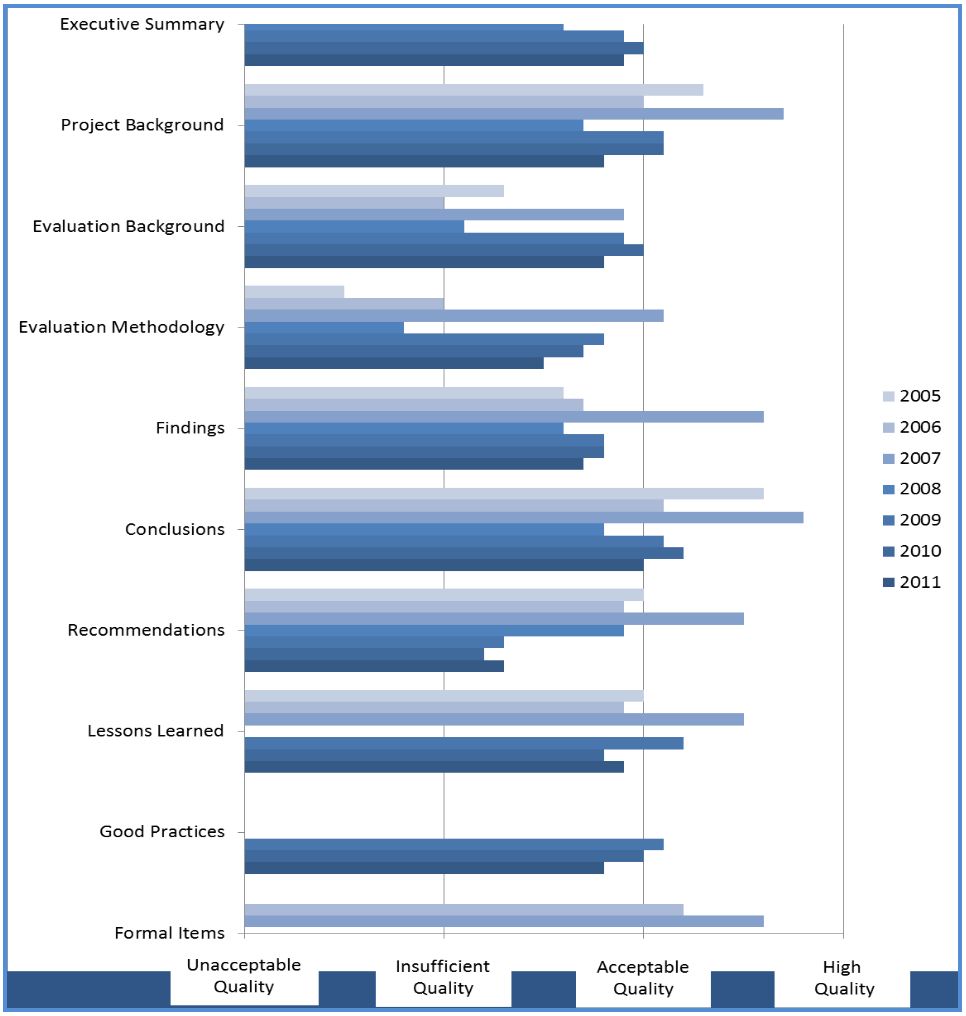
If credibility is thought to consist, in part, of the quality of evaluation methods, then credibility of ILO independent project evaluations can be questioned given that the evaluation methodology section was one of the lowest rated components in all appraisals, as shown in Figure 2. Further, the current plateau in evaluation quality is related to the low

[5] ILO, 2012, p. 4.

[6] Mathiason, J. et al., *Independent External Evaluation of the International Labour Office Evaluation Function* (Geneva, ILO, 2010), p. 41.

quality of key component sections, primarily evaluation methodology and evaluation background, which then in turn, impacts report credibility and produces a chain reaction in terms of lower quality throughout the remainder of the report. Moreover, it is speculated

Figure 2. Overall component quality per year[7]



that while the quality of reports has increased, the longstanding areas for improvement—particularly the key areas for improvement impacting credibility (i.e. methodology and

[7] Appraisal teams, methods, and instruments changed over the years, thus, direct comparison should be made with great caution. Appraisals of 2008 and 2009-2011 reports are the most comparable since they were conducted by the same institution (i.e., Western Michigan University). Although, it should be noted that changes were made to the appraisal instrument used for the 2009-2011 reports to reflect reporting requirements outlined in ILO's *Checklist 5*.

evaluation background)—may in part, impact use, as the 2010 *Independent External Evaluation of the ILO Evaluation Function* found the following:

> Evaluations that are produced by the ILO are of generally high quality, measured in terms of UNEG standards, but they have not been used as much as implied by the 2005 Policy. Managers do not yet see them as essential ingredients either in policy formulation and analysis or in strategic planning and performance reporting under RBM.[8]

Table 2 shows components which were frequently omitted as cited in the appraisal of 2009-2011. Those with the lowest quality of content ratings have been identified as priority areas for improvement.

Table 2. Most frequently omitted and/or lowest rated items - 2009 and 2011 appraisal

Report Component	Items	Per cent of Reports Assessed	Quality Ratings		
			Mean	Median	Mode
Project Background	Project logic	90%	2.0	2	1
Evaluation Background	Primary uses of the evaluation	53%	1.7	1	1
	Clients of the evaluation and the main audience of the report	65%	2.4	3	3
Evaluation Methodology	Evaluation Question	58%	1.9	2	2
	Evaluation Criteria	70%	1.7	2	1
	Protection of human subjects/ethical safeguard	20%	1.5	1	1
	Sampling procedure	14%	1.2	1	1
	Data collection instruments	30%	1.5	1	1
	Data analysis method	38%	1.3	1	1
	Limitations	56%	2.0	2	3
	Description of stakeholder participation	46%	1.5	1	1
Findings	Poverty alleviation	83%	1.4	1	2
	All data is sex-disaggregated, where feasible	61%	1.2	1	0
	Data is broken down by other social categories, if relevant	28%	0.8	0	0
Recommendations	Recommended time frame	98%	0.3	0	0
	Action needed to remedy situation	99%	1.4	2	2
	Priority or importance	99%	0.1	0	0
	Acknowledges resource implications	99%	0.1	0	0

Note: A mean of 1.5 or higher indicates acceptable quality and a mean of 2.5 indicates high quality.

[8] Mathiason, J. et al, p. ix.

PROJECT BACKGROUND

While the project background section was one of the highest rated sections across all appraisals, the project logic item was cited as a key area for improvement across multiple appraisals. Particular areas of concern included insufficient differentiation and lack of clear linkages between project activities, outputs, and outcomes. Better clarity on project logic could enhance report credibility. When evaluation questions, criteria, and methods are better linked with project components, this in turn, improves the findings and conclusions sections where confusion over outputs and outcomes and/or lack of inclusion of outcomes was also noted as an issue over time. Without a clear understanding of what happened or what was evaluated, findings cannot be easily referenced for comparable projects.

EVALUATION BACKGROUND

The evaluation background section received some of the lowest quality ratings over time, as shown in Table 2. Items that require specific attention, as indicated by the most recent appraisal, include indicating the primary uses of the evaluation, clients and main audience.

Primary Uses of the Evaluation

As shown in Table 2, the most recent appraisal indicated that primary uses of evaluation were only identified in a little over half of the sampled reports, and when it was addressed, room for improvement remained. Evaluation use was not included as an item in past appraisals, but was added for the 2009-2011 appraisal because both evaluation purpose and use were included in the description of the evaluation background section in ILO's *Checklist 5*. Greater clarity in terms of the intended uses of the evaluations, on the part of the evaluation team, may lead to changes in the evaluation process that result in reports that are deemed to be more useful by ILO staff.

Clients of the Evaluation and Main Audience

Similarly, clients of the evaluation and main audiences were only identified in 65 per cent of reports; however, when it was addressed, it received an average quality rating (2.4) just shy of the high quality cut off (2.5). Greater clarity of the evaluation clients and audiences may alter evaluation practice and result in reports that are more useful to the respective readers.

EVALUATION METHODOLOGY

The evaluation methodology section was identified as one of the key areas for improvement across all appraisals. Evaluation methodology is an integral part of establishing evaluation credibility and validity, and ultimately the use of findings. Items that require specific attention, as indicated by the most recent appraisal, include evaluation questions, criteria, indication that the evaluator applied ethical safeguards and the protection of human subjects, sampling procedures, data collection instruments, methodological limitations, and stakeholder participation.

Evaluation Questions

As shown in Table 2, evaluation questions were included in less than 60 per cent of reports. Because evaluation questions are foundational pieces of any evaluation, it is very difficult to accurately assess the quality of an evaluation report and subsequent findings without knowing the questions to be addressed, thus reducing both credibility and potential use. It might be argued that evaluation questions are listed in the terms of reference (TOR), and therefore should not be reiterated in the body of the report. However, *Checklist 5* clearly states that evaluation questions should be included in the report and thus, it is suggested that a sentence be added in the checklist that states that omission of questions in the report with a reference to the TOR or inception report is not acceptable.

First, it is essential that the evaluation questions are included in the report as almost every aspect of the evaluation revolves around these questions (e.g., criteria and indicators). It was noted that many of the evaluation terms of reference (TORs) list more questions than could possibly be answered with a high degree of quality within the given time frames and budgets of the evaluations. Additionally, many of the TORs include a statement such as the following:

> The evaluation is an independent evaluation and the final methodology and evaluation questions will be determined by the evaluation team in consultation with the evaluation manager.[9]

This statement suggests that not all of the evaluation questions listed in the TORs are addressed in the evaluation reports and thus, for this reason alone, it would be necessary to require that evaluation questions, actually addressed by the evaluation team, be clearly listed in the reports.

Second, overlap and lack of differentiation of evaluation questions and criteria were also noted as quality issues in over the years. For example, rather than listing evaluation questions as required, some reports listed OECD/DAC criteria instead (i.e. relevance, effectiveness, efficiency, impact, and sustainability). These criteria might be too broad,

---

[9] Kanneganti, S., *Reducing Vulnerability to Bondage in India through Promotion of Decent Work (RVBIPDW) – Midterm Evaluation (*Geneva, ILO*,* 2012), p. 53.

unspecific, or even inadequate for each unique project evaluation. Rather, the OECD/DAC criteria should serve as a guide for the development of evaluation questions. More guidance might be needed on formulating evaluation questions that can be answered within specific program contexts and which appropriately suit the definitions of the DAC criteria and associated assumptions about what the program can realistically achieve.

Evaluation Criteria

The inclusion and degree to which criteria are defined has been a major weakness within the methodology section across multiple appraisals. Having evaluation questions, criteria, and indicators that are clearly linked are foundational pieces of any quality evaluation. Lack of identification or clarity on the links between any of these components makes it difficult to assess the quality of the evaluation report and subsequent findings, which impacts both credibility and use. Multiple appraisals indicated that reports often made general reference to the OECD/DAC criteria, but failed to provide any description as to how these criteria were operationalized and measured. For example, the appraisal of 2006 reports states:

> From a broader perspective, it is evident that standard evaluation criteria are used in most evaluations, but the application is often implicit rather than explicit. Most evaluations do not supply clear definitions of the evaluation criteria or of special focus areas with the term 'sustainability' being a typical case in point where there is considerable ambiguity.[10]

Protection of Human Subjects and Ethical Safeguards

In the most recent appraisal, less than a quarter of reports addressed protection of human subjects and ethical safeguards to any degree, a similar finding also noted in past appraisals. Additionally, while the UNEG norms and standards encourage inclusion of a list of interviewees in the appendix of reports, it is suggested that this practice be reconsidered within the ILO. The inclusion of a list of individuals interviewed as part of an evaluation would not be supported by many ethical review boards and is problematic in terms of the quality of data gathered. That is, if respondents in fact are aware that their names will be disclosed in reports, they may not be as open or forthcoming with information as when their identity is kept confidential. One suggestion would be to include job titles or other demographics of interviewees rather than names, to partially address both accountability and confidentiality.

Sampling Procedures

Sampling procedures in general, or particular details thereof, were often missing from reports over the years. For example, in the most recent appraisal, sampling procedures were

[10] Goodchild, M., p. 13.

only documented in 14 per cent of reviewed reports and when addressed, there was often need for improvement. ILO staff indicated that sampling might not be addressed in some reports because a census was taken rather than a sample. However, if that is the case, it could be noted in the report to satisfy requirements and appropriately inform method. Further, even if census surveys were employed, a discussion of the population characteristics in relation to the respondents would be important. One consideration going forward would be to reword the *Checklist 5* sampling item to read something similar to the following: "Selection of data sources (e.g., sampling, census)".

Data Collection Instruments

Data collection instruments were only discussed in 30 per cent of reports sampled as part of the most recent appraisal, as was similarly noted in past appraisals. Further, while the average rating score met the threshold for acceptable quality, there was much room for improvement, as was also noted across the years. For example, in the appraisal of 2006 reports, it was indicated that "the instruments and tools used in the evaluations were seldom documented to a high standard."[11] Failure to discuss what instruments were used and how they were developed threatens the credibility of the report and findings. Relatedly, only a third of reports included a copy of the data collection instruments in the appendix as required. Inclusion of the instruments would likely increase evaluation credibility and use, and provide opportunities for institutional learning and development, insofar as a repository of data collection instruments could lead to increased evaluation efficiency and encourage learning regarding the development, refinement, and validation of instruments.

Data Analysis Method

Data analysis methods were only addressed in 38 per cent of sampled reports and when they were, the average quality rating was considered insufficient (1.3). Guidance and examples may need to be provided to improve both the inclusion and quality of this item.

Limitations

The 2009-2011 appraisal indicated that over half of sampled reports addressed evaluation limitations and that while the average score was considered acceptable overall, there was often room for improvement. More thorough documentation of limitations would allow readers to better assess the credibility or degree of confidence they should put in the findings, conclusions, and potential for use.

[11] Goodchild, M., p. 13.

Description of Stakeholder Participation

In the most recent appraisal, a description of stakeholder participation was included in less than half of the reports. When the item was addressed, while it received a quality score just at the cut off for acceptable quality (1.5) overall, much room for improvement remains. For example, the descriptions were often vague regarding the degree to which stakeholders participated in the evaluation process beyond providing data for interviews.

FINDINGS

Overall, a general disconnect or lack of clarity regarding the project logic and/or the evaluation methods was identified across the years. For example, confusion between outputs, outcomes, and impacts within the findings section was commonly observed or as was stated in the appraisal of 2005 reports, "some reports only looked at the outputs and not the outcomes of the project which made them read more like progress reports."[12] Similarly, the link between data sources and data collection methods and findings was not always clear. For example, if data collection methods included interviews and a systematic review of project documents, findings would often not be delineated in terms of data collection method or source but rather would be reported at a general level. Additionally, as was observed over the years and stated in the appraisal of 2006 reports, "often the validation of findings and referencing of evidence is poor, particularly in relation to evaluation instruments."[13]

Poverty Alleviation

While poverty alleviation was addressed to some degree in over 80 per cent of reports, when it was, the quality of the content was not considered acceptable. In part, this may be related to the fact that only a little over a quarter of reports sampled as part of the most recent appraisal included a separate section on poverty alleviation or other ILO cross-cutting issues (i.e. gender mainstreaming, labour standards, and workers' education, tripartism, and social dialogue). ILO EVAL staff report limitations in addressing poverty within the findings section to be a result of project design, in that many projects do not include an indicator to measure their impact on poverty. Clear delineation of the project logic, evaluation questions, criteria, and indicators might help to identify not applicable items in future appraisals.

[12] ILO. *Quality Appraisal of 2005 Independent Project Evaluations (Geneva, ILO,* 2006), p. 3.

[13] Goodchild, M., p. 14.

Disaggregation of Data

The disaggregation of data by sex and social categories was only observed in 61 and 28 per cent of sampled reports respectively; when it was observed, the quality with which it was executed was deemed less than sufficient. Very rarely did reports explicitly discuss the opportunities and limitations of data disaggregation.

RECOMMENDATIONS

While many of the evaluation reports produced between 2009 and 2011 included a recommendations section, the quality with which the following items were addressed was not considered acceptable: recommended timeframe, action needed to remedy situation, priority or importance level, and acknowledgement of resource implications. ILO staff has noted that independent evaluators may not be equipped to provide such detailed recommendations, and perhaps these items should be reconsidered for inclusion in *Checklist 5* or a new approach should be taken to creating the recommendations section. For example, the formulation of recommendations might lend itself as an opportunity for collaboration and stakeholder participation, which could strengthen credibility and use of the findings. Additionally, ILO staff suggested that several of the items mentioned may not be relevant depending on the timing of the evaluation (e.g., midterm or final) and it is suggested that evaluators note this within the recommendations section of the report.

CONCLUSIONS

Confusion regarding the difference between the findings and conclusions sections was observed in multiple appraisals. Overall, evaluative conclusions were often omitted and instead the text focused on findings or general statements not directly backed up by evidence or directly related to the evaluation questions, which greatly diminishes the credibility of the evaluation. This failure is partially thought to be attributed to inadequate links and descriptions of evaluation questions, criteria, and indicators, noted in earlier sections of the evaluation, and were often accompanied by a failure to discuss performance levels, benchmarks, standards of success, or other interpretative frameworks of merit and worth in the ILO context.

LEVERAGING FINDINGS FROM THE APPRAISALS

Findings from the quality appraisals present important lessons for improving the ILO evaluation infrastructure with the ultimate goal of increasing the quality of evaluations, and the quality of reporting in particular. To that end, this section provides suggestions on how to leverage findings to address key areas for improvement, described in the previous section. These suggestions are based on themes identified across key areas for improvement in not only the most recent quality appraisal, but past appraisals as well.

CHECKLIST 5: PREPARING THE EVALUATION REPORT - REQUIREMENT OR GUIDELINE?

- ✓ Make it well known that *Checklist 5* is a requirement and that each item should be addressed accordingly.
- ✓ Make sure that evaluators receive a copy of *Checklist 5* as part of the contracting process.
- ✓ Have evaluation managers complete and sign an itemized and revised version of *Checklist 5* at the time the draft report is reviewed so that improvements can be made in the final draft.
- ✓ Also, have evaluators and managers complete and sign an itemized and revised version of *Checklist 5* at the time of report submission to increase accountability.

ILO staff indicated some confusion as to whether or not *Checklist 5* should be regarded as a guideline or requirement despite the fact that the document states, "Adherence to this checklist should be considered a ***contractual requirement*** when submitting evaluations ***to ensure full remuneration*** of the contract."[14] Thus, there appears to be a need to increase accountability by (a) making the obligation known throughout the organization in general and among evaluation managers in particular, and (b) holding evaluators accountable for addressing the checklist as part of their evaluation assignment. Additionally, evaluation TORs should include a clause stating that all *Checklist 5* items must be addressed within the report, and if an item is not applicable, a justification should be provided within the body of the text, the limitations section, or within an appendix.

Furthermore, it is suggested that a point-by-point checklist, based on *Checklist 5*, be created that evaluators would be required to sign when submitting evaluation reports to verify that all items have been addressed and/or the reason why items were not addressed was documented (e.g., Table 3). Similarly, ILO EVAL could request that all ILO evaluation managers submit a copy of *Checklist 5* or *Checklist 6: Rating the Quality of Evaluation Report*,

[14] ILO. 2013. *Checklist 5*, p.1.

as part of the review of the draft report. This would provide evaluators with feedback on missing components or poorly addressed content during the normal review process so that reports could be improved in real time—versus learning about deficiencies during the appraisal process when changes cannot be readily made—and would serve as a means of holding evaluators and managers accountable for the quality of deliverables. Having managers review and sign the checklist would also provide a means for tracking the managers who were responsible for given evaluations, which in turn would allow for identification of all managers at the time of an appraisal.

PROPOSED REVISIONS OF CHECKLIST 5

- ✓ Align *Checklist 5: Preparing the Evaluation Report* and the Quality Appraisal Instruments
- ✓ Provide a point-by-point format
- ✓ Add more definitions and examples
- ✓ Add items related to areas for improvement

While *Checklist 5* is a helpful tool, further refinement is recommended to help improve the quality of evaluation reports overall. Similarly, it is also proposed that *Checklist 6* be updated and kept consistent with any changes made to *Checklist 5*. The following are suggestions for improving *Checklist 5* and subsequently, improving report quality, credibility, and use.

Table 3. Example from the suggested checklist for submission with evaluation reports

Addressed	NA and explained why in report	Items	Comments
☐	☐	Evaluation questions	
☐	☐	Evaluation criteria to answer the questions	
☐	☐	Protection of human subjects/ethical safeguards	
☐	☐	Use of standards and norms	
☐	☐	Data sources	
☐	☐	Data collection instruments	
☐	☐	Data analysis methods	
☐	☐	Limitations	
☐	☐	Description of stakeholder participation	

I certify that all items in this checklist have been addressed accurately and sufficiently.

____ (Evaluator Signature) _______________ (Date)

____(Manager Signature) _______________ (Date)

Align *Checklist 5* and the Quality Appraisal Instrument

Evaluation reports can only truly meet high standards of quality when each requirement or standard is clearly communicated and transparent to evaluators and evaluation managers. While the instrument used for the most recent quality appraisal is very similar to *Checklist 5*, differences still exist, for example, in terms of disentangling double-barreled items. Aligning *Checklist 5* and the Appraisal Instrument would be an easy way to improve report quality.

Point-by-Point Format

Overall, it is recommended that a more consistent point-by-point format of *Checklist 5* be developed that more closely overlaps with the appraisal instruments. A more consistent point-by-point format would provide a clear and quick way to communicate each requirement, to increase accountability, and to have greater alignment between the stated requirements and the requirements by which report quality is assessed. For example, the executive summary section of *Checklist 5* should be formatted point-by-point as is done in other sections. Further, some of the section descriptions should be considered for inclusion in the point-by-point list. *Checklist 5* explains that the background section "describes the purpose and primary use of the evaluation."[15] However, use and purpose are not included in the list of bullet points, but are considered separate items in the evaluation appraisal instrument; thus, it is suggested that use and purpose be listed separately as bullet points.

Add Short Definitions/Examples

While some bullet points provide examples, more bullet points could benefit from a brief definition or example. For instance, under project background, in addition to listing project logic as a bullet point, it could be accompanied by a definition and/or description such as the following: project logic (i.e. key activities linked to anticipated outputs [direct results] and outcomes [changes as a result of activities and outputs]. Use of a logic model or theory of change is highly suggested).

Add Items Related to Areas for Improvement

Additional items should be added to the list of bulleted items, as is suggested above. Further, EVAL may want to consider the inclusion of additional items related to the identified areas for improvement. The following list includes several suggestions to consider and shows the corresponding sections in brackets:

- Intended users [Evaluation background]
- Indicators for each criterion [Evaluation methodology]

[15] ILO. 2013. *Checklist 5*, p. 3.

- Evaluation performance levels, benchmarks, and/or standard [Evaluation methodology]
- Selection of data sources (e.g., sampling procedure or census) [Evaluation methodology]
- Response rate [Findings]
- Triangulation of conclusions (e.g., evaluators, theory, data collection method, data sources) [Conclusions]
- Evaluative judgment based on performance levels, benchmarks, and/or standards [Conclusions]

DEVELOP SUPPLEMENTARY TRAINING MATERIALS RELATED TO CHECKLIST 5

✓ Uniform report structure

✓ Define and differentiate terms

✓ Clearly demonstrate expectations for high quality

Supplementary materials that elaborate on the sections and items listed in *Checklist 5* should be developed that encourage more uniform reporting, define and differentiate terms, and provide examples that clearly demonstrate expectations for high quality reporting. This information could be incorporated into the existing supplementary text or in appendices of the ILO Policy Guidelines for Results-Based Evaluation. Such materials could include a template, a template/guide, a detailed guide, glossary, model reports, logic model and/or theory of change templates, and an evaluation matrix, all of which could be supported by brief online webinars to increase awareness, use, and accessibility of materials.

Uniform Report Structure
A uniform report structure communicated via a template or detailed guide could encourage higher quality of contents and usability—something similar to what was created for the title page. A uniform report structure could greatly benefit ILO staff in locating relevant information (i.e. use) and evaluation managers in efficiently assessing evaluation quality and providing feedback when draft evaluation reports are submitted, especially if a revised version of *Checklist 5* is used as a feedback mechanism and format for a template.

Templates for items such as log frames, logic models, and/or theories of change and an evaluation matrix may also encourage uniform structure, quick access to important information, and increased quality of contents.

Log Frames/Logic Model/Theory of Change Template

Templates for developing program log frames, logic models, or theories of change could be a simple way to improve the quality of the project and evaluation background sections, as well as findings and conclusion sections, by encouraging clear links and definitions of project inputs, activities, outputs, and outcomes. Inclusion of a basic program log frame, logic model, and/or theory of change could be a requirement of all evaluations.

Evaluation Matrix

Similarly, an evaluation matrix template could aide managers and evaluators in communicating and linking evaluation questions, criteria, indicators, and performance levels of success (i.e. benchmarks or standards) or even evaluation purpose and use; thus providing the basis for developing and documenting rigorous evaluation methodology. A similar suggestion was made in the appraisal of 2007 reports. Strengthening the methodology sections to enhance conclusions and recommendations could take place using a specific table formatted with the following items: the OECD/DAC criteria (i.e. relevance, efficiency, effectiveness, impact, and sustainability), stakeholders/project sites sampled and identified, and a list of actual methods employed. The table may facilitate one to link a) evaluation questions/criteria to b) sampling and stakeholder participation to c) findings. EVAL should reinforce the use of a clear system in documenting methods in future reports.[16]

Define and Differentiate Terms

Given that evaluation is a relatively young profession, inconsistency and variation in the use of evaluation terms is common. While the OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management is a great resource, it does not include all terms utilized in *Checklist 5.* For example, the term "criteria" is not defined in the OECD/DAC Glossary and was a weakness of many of the reports. Further, some terms that are included in the glossary, such as purpose, are defined in reference to projects and not evaluation.

A glossary and/or a template could provide definitions and additional information that could help evaluators understand how EVAL differentiates terms included in *Checklist 5*. Based on observations of the sample of 2009-2011 reports, the following are some of the terms that should be considered for inclusion: outputs, outcomes, impacts, evaluation questions, criteria, indicators, evaluation purpose, evaluation use, evaluation clients, evaluation audience, protection of human subjects and the use of ethical safeguards, stakeholder participation, performance level/benchmarks/standards, and evaluative conclusions.

---

[16] Nichols, T. *Quality Appraisal of Independent Project Evaluation Reports 2007 (*Geneva, ILO, 2008), p. 7.

Clearly Demonstrate Expectations for High Quality

Guidance and examples that clearly demonstrate expectations for how reports can attain high quality ratings for all items in *Checklist 5*, especially those listed in Table 2, that relate to the key areas for improvement, may help to improve report quality. Depending on the item, guidance may need to focus on the "how to" and/or what details need to be included to address each item with high quality. For example, "how to" instructions may be beneficial for items such as the following:

- Developing evaluation questions based on DAC criteria (add to existing section in the ILO evaluation policy guide by showing examples of statements in question format);
- Identifying criteria and measurable indicators;
- Making evaluative conclusions based on performance levels, benchmarks, and standards—including distinguishing the difference between expectations for the findings and conclusion sections;
- Formatting findings and conclusion sections so they clearly relate back to evaluation questions;
- Validating and supporting findings and conclusions;
- Overlap between general findings and findings within the cross-cutting sections;
- How to address cross-cutting issues such as poverty alleviation;
- Identifying unexpected or unintended effects; and
- Identifying relevance and sources of data to facilitate disaggregation of data by sex or social categories.

Additionally, guidance on what information needs to be presented to address an item with high quality may be relevant for items such as data collection instruments, the protection of human subject/ethical safeguards, and description of stakeholder participation. Guidance could be detailed or brief, incorporated into a template and/or a separate guide. Such documents could include or be supplemented by model reports and/or sections of reports, log frames, logic models, and/or a theory of change template, and an evaluation matrix template.

Examples/Model Reports

Real or fictitious examples of model sections or entire reports could be useful in helping evaluators and/or evaluation managers translate guidelines into practice while also demonstrating expectations of high quality. Examples of model reports or use of model sections of reports could be included in a template and/or detailed guide.

FINAL CONCLUSIONS

While the overall quality of ILO evaluation reports has increased over the years, key areas for improvement—highly related to evaluation credibility and use—have plateaued. Thus, several suggestions for leveraging findings from the most recent appraisal of independent project evaluation reports have been provided to not only improve the quality, credibility, and use of ILO evaluation reports but evaluation within the ILO overall.

- Results can best be leveraged by clarifying and increasing accountability for adherence to ILO's *Checklist 5: Preparing the Evaluation Report*.
- To increase accountability, it is suggested that evaluators and evaluation managers be required to submit a signed and itemized version of *Checklist 5* along with the final report.
- It is suggested that evaluation managers also complete the checklist during the draft report review process to facilitate real-time updates.
- Findings of this *Think Piece* should be used to revise *Checklist 5*, the most important being alignment with the appraisal instrument.
- Findings of this *Think Piece* should also be used to develop supplementary training materials related to *Checklist 5.* The two most important issues that new training material should emphasize relate to the development of a uniform report structure and clear demonstration of expectations for high quality.

REFERENCES

American Evaluation Association. 2004. *Guiding Principles for Evaluators*. Available at: http://www.eval.org/p/cm/ld/fid=51

Goodchild, M. 2007. *Quality Appraisal of Independent Project Evaluation Reports 2006.* (Geneva, ILO, Internal Document).

International Labour Organization. Governing Body. 2005. *A new policy and strategic framework for evaluation at the ILO*. (Geneva). Available at: http://www.ilo.org/public/libdoc/ilo/GB/294/GB.294_PFA_8_4_engl.pdf

International Labour Organization. 2006. *Quality Appraisal of 2005 Independent Project Evaluations*. (Geneva, ILO, Internal Document).

International Labour Organization. Governing Body. 2011. *Results-based Strategies 2011-15: Evaluation Strategy—Strengthening the Use of Evaluations*. (Geneva). Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---relconf/documents/meetingdocument/wcms_152025.pdf

International Labour Organization. 2012. *ILO Policy Guidelines for Results-Based Evaluation: Principles, Rationale, Planning and Managing for Evaluations*. (Geneva). Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_168289.pdf

International Labour Organization. 2013. *Checklist 5: Preparing the Evaluation Report*. [Supplemental document to the *ILO Policy Guidelines for Results-Based Evaluation].* (Geneva). Available at: http://www.ilo.org/eval/Evaluationguidance/WCMS_165967/lang--en/index.htm

International Labour Organization. 2013. *Checklist 6: Rating the quality of evaluation reports.* [Supplemental document to the *ILO Policy Guidelines for Results-Based Evaluation*]. (Geneva). Available at: http://www.ilo.org/eval/Evaluationguidance/WCMS_165968/lang--en/index.htm

Kanneganti, S. 2012. *Reducing Vulnerability to Bondage in India through Promotion of Decent Work (RVBIPDW) – Midterm Evaluation*. (IND/11/02/CAN). (Geneva, ILO).

Mathiason, J.; Williams, F.; Arora, P.; and Sutherland, A. 2010. *Independent External Evaluation of the International Labour Office Evaluation Function*. (Syracuse, NY, Associates for International Management Services). Available at: http://www.ilo.org/gb/GBSessions/WCMS_145748/lang--en/index.htm

Nichols, T. 2008. *Quality Appraisal of Independent Project Evaluation Reports 2007.* (Geneva, ILO, Internal Document).

Organization for Economic Co-operation and Development. Development Assistance Committee. 2002. *Glossary of Key Terms in Evaluation and Results Based Management. (Paris).* Available at: http://www.oecd.org/development/peer-reviews/2754804.pdf

Organization for Economic Co-operation and Development. Development Assistance Committee. 2010. *Evaluating development co-operation: Summary of key norms and standards* (2nd ed.), *(Paris).* Available: http://www.oecd.org/development/evaluation/dcdndep/41612905.pdf

Schroeter, D.; Cullen, A.; Robertson, K.; and Coryn, C. 2009. *Independent Quality Appraisal of ILO's 2008 Independent Project Evaluation Reports*. Western Michigan University, The Evaluation Center, Kalamazoo, MI. (Geneva, ILO, Internal Document).

Schroeter, D.; Robertson, K.; Arbour, G.; Mateu, P.; Roth, Y.; and Proctor, C. 2013. *Independent Quality Appraisal of ILO's 2009-2011 Project Evaluation Reports and Evaluation of the Status of Implementing the UN System-wide Action Plan*. Western Michigan University, The Evaluation Center, Kalamazoo, MI. (Geneva, ILO, Internal Document).

United Nations Evaluation Group. 2012. *Norms for Evaluation in the UN system*. (New York, UNEG Secretariat). Available at: http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=21

United Nations Evaluation Group. 2012. S*tandards for Evaluation in the UN system*. (New York, UNEG Secretariat). Available at: http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=22

Yarbrough, D. B.; Shulha, L. M.; Hopson, R. K.; and Caruthers, F. A. 2011. *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users.* (3rd Ed.). (Thousand Oaks, CA, Sage).

ANNEX A. APPRAISAL TOOL

PROVISION OF KEY COMPONENTS			
Report Title: TC Code:			
0 = Not provided 1 = Provided 99 = Not applicable. Please use this option sparingly. Please provide comments/justifications of ratings as appropriate. Sometimes information may be displaced in other report sections. Beware and make a note.			
TITLE PAGE			
*1. Title page (first two pages)** [17]	0	1	99
2. Title of project(s) evaluated	0	1	99
*3. Uses the same title as cited in the TOR**	0	1	99
4. Project reference/TC code(s)	0	1	99
5. Timing of evaluation (final or interim/midterm)	0	1	99
*6. Country(ies)/Region(s)**	0	1	99
*7. Project end date**	0	1	99
*8. Evaluation manager**	0	1	99
*9. ILO administrative unit responsible for project**	0	1	99
*10. ILO technical unit responsible for backstopping the project**	0	1	99
11. Name(s) of evaluator(s)	0	1	99
12. Date evaluation completed (date or time frame)	0	1	99
*13. Keywords**	0	1	99
Comments/Justification of ratings:			
TABLE OF CONTENTS AND LISTS			
1. Table of contents (TOC)	0	1	99
2. Contains at least all first- and second- level headers in the report.	0	1	99
3. TOC lists appendices/annexes	0	1	99
*3.1. ...Terms of reference (TOR)**	0	1	99
*3.2. ...List of persons interviewed/Interviewees/etc.**	0	1	99
*3.3. ...Data collection instruments (e.g., survey, questionnaire, interview protocol, focus group or observation protocol/etc.)**	0	1	99
*3.4. ...List of publications cited/References/Bibliography**	0	1	99
4. Lists of tables (include if more than 5 in report)	0	1	99
5. Lists of figures/charts (include if more than 5 in report) (pictures and text boxes should be considered as figures)	0	1	99
Comments/Justification of ratings:			
ACRONYM/ABBREVIATION LIST			
1. List of acronyms or abbreviations (99 - if clearly indicated in text via footnotes, brackets, etc.)	0	1	99
Comments/Justification of ratings:			
EXECUTIVE SUMMARY			
*1. Executive summary**	0	1	99
*2. Project background & context **	0	1	99
*2.1. ... project purpose (i.e., objectives, geographic coverage)**	0	1	99
*2.2. ... project logic (i.e., key activities, anticipated outputs, and/or outcomes)**	0	1	99
*2.3. ... project management structure**	0	1	99
2.4. ...present situation/status of project (e.g., ongoing, finished, level of accomplishment, etc.)	0	1	99
3. Evaluation purpose (intention, goal, aim)	0	1	99
4. Evaluation scope (e.g., geographic coverage, information on phases of the project when the evaluation covers more than one project, special focus areas such as gender, collaboration, exit strategy, etc.)	0	1	99
5. Evaluation clients/users	0	1	99

[17] *Indicates an item added to the checklist in 2013 based on the ILO "*Checklist 5: Preparing the Evaluation Report*" and/or to improve the checklist, as compared with the last checklist used for the 2008 meta-evaluation.

6.	Evaluation methodology	0	1	99
7.	Main findings	0	1	99
8.	Conclusions	0	1	99
9.	Main recommendations	0	1	99
10.	Important lessons learned	0	1	99
11.	*Good practices (Are presented as activities and strategies that performed well, achieved marked results or benefits and are determined to be specifically useful for replication or up-scaling to other ILO projects.)**	0	1	99
Comments/Justification of ratings:				
BODY OF REPORT				
Consider elements of sections within the body of the report. Beware and make a note.				
PROJECT BACKGROUND				
1.	*Project background section**	0	1	99
2.	Project context (i.e., background, problem situation, needs, socio-political context)	0	1	99
3.	Project purpose (i.e., objectives, geographic coverage)	0	1	99
4.	Project logic (i.e., key activities, anticipated outputs, and/or outcomes)	0	1	99
5.	Funding arrangements (including donor)	0	1	99
6.	Organizational arrangements for project implementation (i.e., project management structure/contributions and roles of ILO, the project partners, and other stakeholders).	0	1	99
7.	*Project major events and milestones**	0	1	99
Comments/Justification of ratings:				
EVALUATION BACKGROUND				
1.	*Evaluation background section**	0	1	99
2.	Evaluation purpose (intention, goal, aim)	0	1	99
3.	*Primary uses of the evaluation**	0	1	99
4.	Evaluation scope (e.g., geographic coverage, information on phases of the project when the evaluation covers more than one project, special focus areas such as gender, collaboration, exit strategy, etc.)	0	1	99
5.	*Operational sequence of evaluation, dates of evaluation mission**	0	1	99
6.	Clients of the evaluation and the main audience of the report (e.g., donors, constituents, implementing parties)	0	1	99
7.	Name(s) of evaluation team member(s)	0	1	99
8.	Name of evaluation manager is included with the citation of the ILO administrative unit	0	1	99
Comments/Justification of ratings:				
METHODOLOGY				
1.	*Evaluation methodology section**	0	1	99
2.	Evaluation questions	0	1	99
3.	Evaluation criteria to answer the questions (if say OECD criteria=1, not have to specify the individual criterion)	0	1	99
4.	Protection of human subjects (use of ethical safeguards)	0	1	99
5.	Use of standards and norms (UNEG and OECD/DAC Evaluation Quality Standards)	0	1	99
6.	Data sources (e.g., primary data sources such as interviewees; secondary data sources such as available documents, questionnaires, literature, etc.)	0	1	99
7.	Sampling procedures (for selecting documents, sites, respondents to questionnaires, interviewees, etc.)	0	1	99
8.	Data collection methods (e.g., desk study, interviews, focus groups, observations)	0	1	99
9.	Data collection instruments (e.g., description, development)	0	1	99
10.	Data analysis methods	0	1	99
11.	Limitations (e.g., of methods, case selection, data sources, etc.)	0	1	99
12.	*Description of stakeholder participation (e.g., instrument development, analyzing findings, interpreting information)**	0	1	99
13.	Rationale for stakeholder participation	0	1	99
Comments/Justification of ratings:				
MAIN FINDINGS				
1.	Findings	0	1	99
	*1.1. Separate section on cross-cutting issues such as: (i) the mainstreaming of gender; (ii) poverty alleviation; (iii) labour standards; and (iv) workers' education, tripartism and social dialogue**	0	1	99
Comments/Justification of ratings:				
CONCLUSIONS				
1.	*Conclusions**	0	1	99
2.	Recommendations	0	1	99
	*2.1. Recommendations are numbered**	0	1	99
3.	Lessons learned (for projects)	0	1	99

*4. Good practices (for projects)**	0	1	99
Comments/Justification of ratings:			
APPENDICES/ANNEXES			
*1. Appendices/Annexes**	0	1	99
1.1. Terms of reference	0	1	99
1.2. List of persons or organizations interviewed	0	1	99
1.3. Data collection instruments (e.g., questionnaires, interview protocols)	0	1	99
1.4. List of publications referenced, cited, or mentioned in report (list of working documents, Web sites, etc.)	0	1	99
Comments/Justification of ratings:			
TABLES AND FIGURES/CHARTS			
*1. Contains tables**	0	1	99
*2. Contains figures/charts (pictures and text boxes should be considered as figures)**	0	1	99
3. Tables are labeled (if 1 is done = 1)	0	1	99
4. Figures/Charts are labeled (pictures and text boxes should be considered as figures) (if 1 is done = 1)	0	1	99
*5. Tables and figures/charts are referenced correctly within the text**	0	1	99
6. Data are disaggregated by subgroups (e.g., sex, age, ethnic group, etc.), if relevant	0	1	99
Comments/Justification of ratings:			
OTHER COMPONENTS			
1. Acknowledgments	0	1	99
2. Name organizational affiliation of evaluator(s)[18]	0	1	99
3. Type of evaluation (independent or internal)	0	1	99
Comments/Justification of ratings:			

QUALITY OF CONTENTS						
Report Title:	TC code:					
0 = Unacceptable quality 1 = Insufficient level of quality 2 = Acceptable level of quality 3 = High quality 88 = Insufficient information 99 = Not applicable. If you indicated 0 in the components. Please use "Not applicable" and "Insufficient information" options sparingly. (C) – Quality of communication (Q) – Quality of content (CS) – Item rated in components section Please provide comments/justifications of ratings as appropriate and for low ratings (i.e., 0 and 1).						
EXECUTIVE SUMMARY						
Note that these items must be in the executive summary.						
*1. Project purpose (i.e., objectives, geographic coverage) (C) (CS)**	0	1	2	3	88	99
*2. Project logic (i.e., key activities linked anticipated outputs and outcomes) (CS)**	0	1	2	3	88	99
*3. Project management structure (CS)**	0	1	2	3	88	99
*4. Present situation/status of project (e.g., ongoing, finished, level of accomplishment, etc.) (CS)**	0	1	2	3	88	99
*5. Evaluation purpose (intention, goal, aim) (CS)**	0	1	2	3	88	99
*6. Evaluation scope (e.g., geographic coverage, information on phases of the project when the evaluation covers more than one project, special focus areas such as gender, collaboration, exit strategy, etc.) (CS)**	0	1	2	3	88	99
*7. Evaluation clients/users (CS)**	0	1	2	3	88	99
8. Evaluation methodology (CS)	0	1	2	3	88	99
9. Main findings (CS)	0	1	2	3	88	99
10. Conclusions (CS)	0	1	2	3	88	99
11. Main recommendations (CS)	0	1	2	3	88	99
12. Important lessons learned (CS)	0	1	2	3	88	99

[18] This is not part of the checklist, but logically the organizational affiliation might be more informative than the actual evaluator(s) doing the work.

*13. Good practices (CS)**	0	1	2	3	88	99
Comments/Justification of ratings:						
PROJECT BACKGROUND						
1. Project context (i.e., background, problem situation, needs, socio-political context) (CS)	0	1	2	3	88	99
2. Project objectives (CS)	0	1	2	3	88	99
3. Project logic (key activities linked to anticipated outputs and outcomes) (CS)	0	1	2	3	88	99
4. Funding arrangements (including donor[s] and project budget) (CS)	0	1	2	3	88	99
5. Organizational arrangements for project implementation (i.e., project management structure/contributions and roles of ILO, the project partners, and other stakeholders) (CS)	0	1	2	3	88	99
*6. Project major events and milestones (CS)**	0	1	2	3	88	99
Comments/Justification of ratings:						
EVALUATION BACKGROUND						
1. Evaluation purpose (intention, goal, aim) (CS)	0	1	2	3	88	99
*2. Primary uses of the evaluation (CS)**	0	1	2	3	88	99
3. Evaluation scope (e.g., geographic coverage, information on phases of the project when the evaluation covers more than one project, special focus areas such as gender, collaboration, exit strategy, etc.) (CS)	0	1	2	3	88	99
*4. Operational sequence of evaluation, dates of evaluation mission (CS)**	0	1	2	3	88	99
5. Clients of the evaluation and the main audience of the report (e.g., donors, constituents, implementing parties) (CS)	0	1	2	3	88	99
6. Description of the evaluation team	0	1	2	3	88	99
Comments/Justification of ratings:						
EVALUATION DESIGN/METHODOLOGY						
1. Evaluation questions (CS)	0	1	2	3	88	99
2. Evaluation criteria to answer the questions (CS)	0	1	2	3	88	99
3. Protection of human subjects (use of ethical safeguards) (CS)	0	1	2	3	88	99
4. Use of standards and norms (UNEG and OECD/DAC Evaluation Quality Standards) (CS)	0	1	2	3	88	99
5. Data sources (e.g., primary data sources such as interviewees; secondary data sources such as available documents, questionnaires, literature, etc.) (CS)	0	1	2	3	88	99
6. Sampling procedures (for selecting documents, sites, respondents to questionnaires, interviewees, etc.) (CS)	0	1	2	3	88	99
7. Data collection methods (e.g., desk study, interviews, focus groups, observations)	0	1	2	3	88	99
8. Data collection instruments (e.g., description, development) (CS)	0	1	2	3	88	99
*9. Data analysis methods (CS)**	0	1	2	3	88	99
10. Limitations (e.g., of methods, case selection, data sources, etc.) (CS)	0	1	2	3	88	99
11. Description of stakeholder participation (e.g., instrument development, analyzing findings, interpreting information) (CS)	0	1	2	3	88	99
Comments/Justification of ratings:						
FINDINGS						
1. Relevance to evaluation purpose and scope (Q) (CS)	0	1	2	3	88	99
2. The extent to which evaluation questions have been addressed (cross-check with questions in TOR, if not listed in report) (Q) (CS)	0	1	2	3	88	99
3. Findings and supporting evidence are consistent with methods and data (C) (CS)	0	1	2	3	88	99
4. Effectiveness (achievement of outcomes, objectives, and progress made) (C) (CS)	0	1	2	3	88	99
*5. Unintended or unexpected effects (CS)**	0	1	2	3	88	99
6. Contributing and/or inhibiting factors to project success (C) (CS)	0	1	2	3	88	99
7. Implementation fidelity and the delivery process (e.g., activities and outputs) (C) (CS)	0	1	2	3	88	99
8. Relevance/strategic fit/validity of design, including adequacy of objectives and indicators (C) (CS)	0	1	2	3	88	99
9. Efficiency and adequacy of resource use (C) (CS)	0	1	2	3	88	99
*10. Effectiveness of management arrangements are discussed (including implementation monitoring, outcome monitoring, backstopping, and collaboration with other projects)(C) (CS)**	0	1	2	3	88	99
11. Likelihood of the intervention to have a longer-term impact (C) (CS)	0	1	2	3	88	99
12. Sustainability (C) (CS)	0	1	2	3	88	99
13. Gender mainstreaming (C) (CS)	0	1	2	3	88	99
14. Poverty alleviation (C) (CS)	0	1	2	3	88	99
15. Labour standards (C) (CS)	0	1	2	3	88	99

16. _Workers' education*_, tripartism (between business, labor, and state), and promotion of social dialogue (C) (CS)	0	1	2	3	88	99
17. Promoting employment by creating a sustainable institutional and economic environment (e.g., capacity building, growth and income opportunities) (C) (CS)	0	1	2	3	88	99
18. Discussion of human rights and ethics (e.g., nondiscrimination, gender equity) (C) (CS)	0	1	2	3	88	99
19. All data is sex-disaggregated, where feasible* (Q) (CS)	0	1	2	3	88	99
20. Data is broken down by other social categories (age, ethnic group....) if relevant* (Q) (CS)	0	1	2	3	88	99
Comments/Justification of ratings:						
CONCLUSIONS						
1. Follow from findings (i.e., are based on evidence presented in the findings) (Q) (CS)	0	1	2	3	88	99
2. Judgments are fair, impartial and consistent with the findings (Q) (CS)	0	1	2	3	88	99
Comments/Justification of ratings:						
RECOMMENDATIONS						
1. Are based on findings and conclusions of the report (CS)	0	1	2	3	88	99
2. Are clear and concise (i.e., one or two sentences followed by explanatory text, if needed) (CS)	0	1	2	3	88	99
3. Are relevant and useful (reviewers judgment of perceived relevance for ILO) (CS)	0	1	2	3	88	99
4. Specifies who is called upon to act (CS)	0	1	2	3	88	99
5. Specifies recommended time frame (CS)	0	1	2	3	88	99
6. Specify action needed to remedy situation (CS)*	0	1	2	3	88	99
7. Distinguish priority or importance (e.g., high, medium, low) (CS)*	0	1	2	3	88	99
8. Acknowledges resource implications (CS)*	0	1	2	3	88	99
Comments/Justification of ratings:						
LESSONS LEARNED: In the explanatory text, the consultant should list any positive or negative insights gained through the life of the project which had substantial impact on operations, achievement of outcomes, or impact on sustainability. These can be aimed at the administrative aspects of the project or the technical context of the intervention. These should highlight strengths and weaknesses and provide decision-makers with relevant information to help them avoid common mistakes and promote an enabling environment.						
1. Are based on findings and conclusions of the report (CS)*	0	1	2	3	88	99
2. Are clear and concise (i.e., one or two sentences followed by explanatory text, if needed) (CS)*	0	1	2	3	88	99
3. Are relevant and useful (reviewers judgment of perceived relevance for ILO) (CS)*	0	1	2	3	88	99
4. Highlight strengths and weaknesses (CS)*	0	1	2	3	88	99
5. Provide relevant information to help avoid common mistakes and promote an enabling environment (CS)*	0	1	2	3	88	99
Comments/Justification of ratings:						
GOOD PRACTICES: Good practices are activities or strategies that performed well, achieved marked results or benefits and are determined to be specifically useful for replication or up-scaling to other ILO projects. Good practices are also a good resource for establishing ILO tool kits or model interventions.						
1. Are based on findings and conclusions of the report (CS)*	0	1	2	3	88	99
2. Are clear and concise (i.e., one or two sentences followed by explanatory text, if needed) (CS)*	0	1	2	3	88	99
3. Are relevant and useful (reviewers judgment of perceived relevance for ILO) (CS)*	0	1	2	3	88	99
4. Are presented as activities and strategies that performed well, achieved marked results or benefits and are determined to be specifically useful for replication or up-scaling to other ILO projects. (Good practices are not mean to be an itemization of administrative strengths of the project). (CS)*	0	1	2	3	88	99
Comments/Justification of ratings:						
FORMAL ELEMENTS						
1. A consistent system is used for references to literature, documents, comments, and similar text components (e.g., via brackets, footnotes, endnotes, etc.)	0	1	2	3	88	99
2. Referenced documents have sufficient source information to enable readers to locate the texts.	0	1	2	3	88	99
3. The formatting of the report is consistent.	0	1	2	3	88	99
Comments/Justification of ratings:						