



i-eval THINK Piece, No. 10

Quality Assessments of ILO Project Evaluations: What are the next steps to better evaluations?

Brad R. Watts

**International Labour Office
Evaluation Office**

February 2016



Quality Assessments of ILO Project Evaluations: What are the next steps to better evaluations?

By Brad R. Watts

Assistant Director

The Evaluation Center, Western Michigan University, U.S.A

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

Table of Contents

1.	Introduction	2
2.	What has the quality assessment told us about ILO project evaluations?	3
3.	The challenge: What do these quality assessments tell us about evaluation quality?	4
4.	Ideas for future change and improvement.....	6
5.	Closing thoughts	10

1. Introduction

This Think Piece provides ideas for the ongoing modification and improvement of the system that is currently in place for evaluating projects that are supported by the International Labour Organization (ILO), along with providing a summary of the results of recent studies of these evaluations. The ideas presented here should be considered in the context of a decentralized, hybrid-system although much of the discussion focuses on issues related to the evaluation consultants in the field and the evaluation managers who provide oversight. Under the current system for ILO project evaluation, there are two main parties responsible for the evaluation: an evaluation manager and an independent evaluation consultant or evaluator. The evaluation manager has direct responsibility for selecting and managing the evaluator, who ultimately conducts the evaluation in the field and produces the evaluation report. The managers of decentralized evaluations receive support from a Regional Evaluation Officer (REO). Managers of centralized evaluations receive support from a Departmental Evaluation Focal Point (DEFP). Their work, in turn, is over-seen by a Senior Evaluation Officer from the ILO's Evaluation Office (EVAL). It is the Senior Evaluation Officer who ultimately approves the final report. At the local level, the work of the evaluator is also assisted by the project's manager and staff, who assist with data collection and other practical matters at the local level. All of these parties play a role in the quality of evaluations produced for ILO.

Over the past decade, EVAL has contracted with external organizations to conduct regular quality assessments (sometimes referred to as quality appraisals) of ILO-funded project evaluations. The quality appraisal consists of a systematic review process to rate the quality of evaluations being produced. The process uses EVAL's checklists and guidelines for evaluation as performance criteria.¹ Additionally, these quality assessments also include information on other aspects of project evaluation quality, such as the practices of ILO evaluation managers and whether or not the evaluations are meeting the United Nation's standards on gender equality and the empowerment of women. The most recent quality assessment process also examined the Evaluation Manager Certification Programme (EMCP) that was introduced in 2013 in order to determine if this training effort has been able to show any effect on the quality ratings of evaluations produced under the supervision of participating ILO managers.

This Think Piece is not intended as a simple rehashing of the results of the most recent quality assessment study. For those interested in the detailed, technical results, copies of the full quality assessment report² and the EMCP analysis brief³ may be obtained from EVAL.

¹ An overview of all ILO project evaluation checklists and guidelines may be accessed at http://www.ilo.org/global/docs/WCMS_176814/lang-en/index.htm

² Watts, B., Coryn, C.L.S., Fiekowsky, E.L., Wilson, L.N., & Mateu, P. (2016). *Quality Appraisal of ILO Project Evaluations*. Kalamazoo, MI: Western Michigan University, The Evaluation Center.

Instead, the purpose of this Think Piece is to identify some challenges to improving evaluation quality and offer several alternative ideas that EVAL might consider in order to obtain stronger evaluations in the future.

2. What has the quality assessment told us about ILO project evaluations?

The results of the most recent quality assessment examined 42 ILO project evaluations conducted during 2014 and the first half of 2015.⁴ The results indicated that the average rating score for the reviewed evaluations was around an “acceptable” quality level, based on a four-point rating scale that ranges from unacceptable (0) to high quality (3). Of course, there was some variation across report areas: the executive summary, project background, and recommendations sections were areas of relative strength, while the evaluation methodology was found to be an area of relative weakness. There was also some variation across regions: a country-level analysis found that evaluations in Asia scored somewhat higher than those in the Americas, while other regions fell in-between. Still, the differences were small and all average scores hovered around the “acceptable” level.

A comparison of the results of the most recent quality assessment with previous studies indicates that an “acceptable” evaluation quality has persisted over time, with little significant improvement or decline. Table 1 compares the mean scores derived from the most recent quality assessment on project evaluations conducted during 2014 and the first half of 2015, with the results published in the previous two quality assessment studies.^{5& 6} The previous quality assessment studies were conducted by different authors and each examined a sample of project evaluations during different analysis periods; however, the rating systems and methodologies used were roughly equivalent to the approach used in the most recent study.

The average overall scores are consistently in the range of 1.8-1.9, which can generally be interpreted as representing an “acceptable” level of quality. Although the scores for some sections have ticked up or down between assessments, there is no consistent pattern of either improvement or decline in evaluation quality over time (note that differences of less than 0.2 scale points are generally not meaningful).

³ Watts, B., Coryn, C.L.S., Fiekowsky, E.L., Wilson, L.N., & Mateu, P. (2016). *Quality Assessment Brief: Analysis of Evaluation Manager Certification Programme Impact on Quality and Practice*. Kalamazoo, MI: Western Michigan University, The Evaluation Center.

⁴ Ibid.

⁵ Friedman, J., Blight, N., & Polastro, R. (July 2014). International Labour Organisation – External Quality Appraisal of Independent Evaluation Reports: Draft Report. Sheffield, UK: IOD PARC research report.

⁶ Schroeter, D. et al. (July 2013). *Evaluation Appraisals for ILO Technical Cooperation Projects (CEB/2006/2)*. Kalamazoo, MI, USA: The Evaluation Center.

Past quality assessments have also examined the views and practices of ILO evaluation managers through an online survey. Over the course of the past three quality assessments, the questions and general focus of these surveys have varied, which makes a direct comparison impossible. However, generally speaking the participating evaluation managers

Table 1. Historical comparison of mean quality assessment scores

Report Section	2014- 2015	2012- 2013	2009- 2011
Overall	1.8	1.9	1.8
Executive Summary	2.0	1.9	1.9
Project Background	2.0	2.1	2.0
Evaluation Background	1.7	1.9	1.9
Evaluation Methodology	1.5	1.6	1.7
Findings	1.8	1.9	1.7
Conclusions	1.8	2.1	2.1
Recommendations	1.7	1.8	1.3
Lessons Learned	1.8	1.8	1.9
Good Practices	1.9	1.5	1.9
Formal Elements	1.9	n.a.	1.9

have indicated mostly positive views on project evaluations in the field. The most recent survey of evaluation managers who oversaw programs during 2014 and the first half of 2015 indicated that managers: consider a wide range of factors when selecting evaluation consultants; involve a wide range of stakeholders in the evaluation; and mostly view the resources for evaluation as being adequate.

Finally, the most recent quality assessment was also expanded to include an impact evaluation of the EMCP, which is a training program for ILO evaluation managers. The analysis looked at whether or not the quality assessment scores of evaluations conducted under the supervision of managers who had completed the EMCP were higher than those conducted under the supervision of other evaluation managers. Second, the survey results were examined to determine whether respondents who had completed the EMCP held views or engaged in practices that differed from other managers. Ultimately, no significant differences were found that could be attributed to the EMCP, although managers who had completed the training generally indicated that they felt it was beneficial.

3. The challenge: What do these quality assessments tell us about evaluation quality?

The quality assessments that have been conducted to date all agree that ILO project evaluations have met the basic standards laid out by EVAL. However, the consistency of the quality assessment results also illustrates a concern, which is a lack improvement in the

quality of project evaluations that are being produced in the field. In short, evaluation quality has stagnated at a level below its highest potential. Ideally, evaluation ratings should indicate continuous improvement toward the top rating level (“high quality” or ‘3’ on the 0-to-3 scale).

During the years that have passed since 2010 when the ILO conducted an Independent External Evaluation of the ILO’s evaluation function, substantial effort has been invested in developing a comprehensive set of policy guidelines, tools and training intended to boost the quality of evaluations being conducted for ILO. Examples include evaluation checklists⁷, templates for specific portions of evaluation reports⁸ and the EMCP. Additionally, according to the evaluation managers who completed the most recent survey, these resources are generally made widely available to their respective target audiences of evaluation consultants in the field and ILO evaluation managers. The hypothesis is that these investments should be contributing to a steady improvement in evaluation quality over time.

As examined by an outsider to the ILO, it appears that the consultants who conduct project evaluations in the field have been provided with a comprehensive set of the resources necessary to assist them in conducting a high-quality evaluation. The expected evaluation components are well laid out ahead of time by the terms of reference (ToR) and the evaluation checklists. Additionally, templates are available for several sections of the evaluation reports (e.g. the title page, lessons learned, and emerging good practices sections), which should help ensure that all required information is provided. This is likely far more structure and guidance than is provided for a typical evaluation project.

Of course, the reality is that most project evaluations consistently fall short in at least some areas. Although the mean scores reach the acceptable level, a review of the detailed data in the full report⁹ shows that in each section there are typically individual criteria items where a sizable portion of evaluations score poorly. Additionally, although the quantitative quality assessment scores are usually within the “acceptable” range, the reviewers who participated in the most recent quality assessment study indicated that there may be deeper problems that are not fully reflected in the scores. One issue was that the evaluations do not always successfully integrate the different required evaluation sections into a cohesive finding on the project being evaluated. This can be described as evaluations that meet the “letter of the law” but not the “spirit of the law,” which is to say that sections may be included or topics may be addressed only because they are known to be expected as part of the report, but they are not truly integrated into the evaluation. An example of this

⁷ See for example Checklist 5: Preparing the Evaluation Report available at http://www.ilo.org/global/docs/WCMS_165967/lang-en/index.htm

⁸ See for example the Lessons Learned http://www.ilo.org/global/docs/WCMS_206158/lang-en/index.htm and Emerging Good Practices templates http://www.ilo.org/global/docs/WCMS_206159/lang-en/index.htm

⁹ See note 2.

situation may be found in reports that specifically discuss a topic or include a section that is required by the checklist, but show little evidence of weaving the topic into the broader activities of the evaluation. This suggests that the evaluation consultants are familiar with the guidelines, but may not have covered the topic or included the section had it not been required.

These issues leave us to speculate as to why the quality of evaluations has not improved along with the introduction of all the aforementioned tools and resources by EVAL. For discussion purposes, here are several possibilities to consider:

1. The consultants who conduct project evaluations are not making use of the checklists, templates and other tools provided by EVAL.
2. The checklists, tools, and other resources provided by EVAL are insufficient, misunderstood, or are missing the key elements necessary to directly affect evaluation quality.
3. There are not enough resources (time, money, data) available for consultants to be able to conduct high quality project evaluations.
4. The quality assessment process itself is flawed; the “true” quality of the evaluations is not being accurately measured or communicated by the quality assessment process.

Unfortunately, there is insufficient empirical evidence to either completely support or completely refute any of these possibilities. Instead, it is perhaps reasonable to start from a point of assuming that all of the above possibilities should be considered if evaluation quality is to improve.

4. Ideas for future change and improvement

The past quality assessment studies have all offered recommendations to address pressing evaluation problems or to work toward improving ILO’s project evaluations. However, to significantly improve the quality of project evaluations may require moving beyond incremental improvements, such as making checklist modifications or encouraging managers to push the use of templates in the field. As an alternative approach, the following four ideas are presented as possible steps to more broadly reshape evaluation policy and practice.

A. Interact directly with the independent consultants who conduct the project evaluations to teach skills and identify high-quality evaluators.

The current practice is to provide training to the ILO evaluation managers and have them communicate the goals and expectations of EVAL; however, it may be more effective for EVAL to directly reach out to the evaluators who conduct the evaluations in the field. While

evaluation managers bear responsibility for selecting consultants with the help of a consultant's roster held by EVAL and providing oversight, this role is limited by the evaluators who are available and willing to conduct the work at each country location. Ultimately, the individual evaluation consultant determines the quality of the evaluation that is produced; therefore, it may be beneficial to take a more "hands-on" approach for identifying and developing highly skilled evaluators.

Examples of options for bolstering the quality and capacity of independent evaluation consultants include creating a training program specifically for evaluation consultants and/or certifying evaluation consultants who meet specific quality and training standards. It may be possible for EVAL to directly offer training to evaluators (perhaps similar to the EMCP) or to compile a formal list of trainings, certifications, or other credentials that are desirable for the evaluation consultants to possess. The result should be an increase in the skills of the consultants, along with increased information for EVAL regarding the capabilities of the consultants who bid on evaluations.

Another option is to more directly involve consultants who have experience evaluating ILO projects in the development of checklists, guidelines, and standards for evaluation. Although EVAL should retain control of the process, encouraging input from those directly using the checklists, guidelines, and standards may provide the consultants with a better understanding of how the evaluation results are used and why certain aspects of the projects need to be addressed in their evaluations.

B. Substantially revamp the checklists and guidelines to narrow the focus onto a few core aspects of the evaluations that are the most important.

The current guidelines and checklist items for ILO project evaluations are quite thorough in terms of the coverage of desired components. For example, EVAL Checklist 6¹⁰, which is intended for rating evaluations, includes 63 different items across 11 different component and sub-component areas. Checklist 5¹¹, which provides guidance on what should be covered in a project evaluation, includes just over 100 bullet points and topic areas for consultants to consider. These two main checklist documents do not even include the full spectrum of guidance for the evaluations, such as those that are provided elsewhere for the title page¹², or for completing templates on lessons learned and emerging good practices. While each of these aspects may arguably be beneficial or important to include, it should also be apparent that these represent a large number of subjects to be addressed in all evaluations. Additionally, the extensive checklists do not address whether or not some components of an evaluation are more important and carry more weight than others.

¹⁰ Available at http://www.ilo.org/global/docs/WCMS_165968/lang--en/index.htm

¹¹ Available at http://www.ilo.org/global/docs/WCMS_165967/lang--en/index.htm

¹² Checklist 7, available at http://www.ilo.org/global/docs/WCMS_166363/lang--en/index.htm

It would not seem unreasonable to hypothesize that one reason why evaluation consultants may be producing evaluations of only acceptable quality overall is that: 1) there are too many topics to be covered within the current evaluation framework, timeline, and budget; and 2) there is no clear differentiation regarding the importance of each topic. For example, is the description of the project background really as important as the evaluation findings? And is it reasonable to expect that recommendations, lessons learned, and emerging good practices be addressed to the same extent as other sections? How does the consultant conducting the evaluation know where to focus their efforts? Under the structure of the current checklists, there is no weighting scheme or guidance on what aspects of an evaluation are most important, as all of the checklist items have a quantitatively equal weight.

Past efforts to improving evaluation quality have generally taken the view that evaluation issues should be addressed through the addition of new and more comprehensive rules and guidelines. This includes many of the recommendations issued by past quality assessments (including the most recent one). However, there also exists a very real risk that the continued addition of new rules and guidelines may ultimately result in an evaluation policy that is over-formalized and ultimately begins to harm the quality of evaluation practice.¹³ This can occur when the rules, guidelines, and forms that were originally intended to support better evaluation actually become overly bureaucratic and hinder the evaluation process. While the exact tipping point is unclear, at some point, the benefits of guidelines and rules—such as improved consistency—may be outweighed by downsides, such as diminished flexibility or increased time/costs to meet the requirements of the guidelines. It may seem counterintuitive to reduce guidelines in order to improve quality; however, given the stagnation of evaluation quality observed during years of policy expansion it may be time to consider an alternative to the current approach.

C. Rethink how and when evaluators are involved in projects and how they are paid.

When the survey asked evaluation managers about the sufficiency of resources, nearly 75 percent indicated that the funding for evaluation is “completely adequate” and 55 percent stated that the amount of time was “completely adequate.” However, a quick examination of actual resources relative to the size of projects being evaluated suggests that perceptions may not match reality. Among the reports reviewed during the most recent quality assessment, evaluation expenditures ranged from a low of US\$5,400 to US\$65,000, with an average of about US\$28,000.¹⁴ This represents an average of just less than 1.5% of the project costs, which is far short of the 3.7% average portion of private foundation

¹³ See Trochim, W.M.K. (2009). Evaluation Policy and Evaluation Practice. In Trochim, W.M.K., Mark, M.M., & Cooksy, L.J. (Eds.), Evaluation policy and evaluation practice. *New Directions for Evaluation*, 123, p.16.

¹⁴ Author’s calculations based on sample of 11 reports that provided total and evaluation expenditures consistent in \$US from the 2014-2015 analysis period.

investments that are dedicated to evaluation¹⁵ or the 5%-20% “rule of thumb” that is broadly recognized in the field.¹⁶ Additionally, in terms of time resources, the final project evaluations from the last quality assessment review typically took place during the last months of a project and involved one period of travel to the project site. To provide some context, all ILO projects subject to independent evaluation have budgets in excess of US\$1 million and typically take place over one or more years.

How can issues of limited resources be addressed without substantial budget increases? One way could be to restructure how evaluation consultants are selected and funded. For instance, evaluators could start earlier during the project and collect data over a longer period of time. Early involvement would allow the evaluation consultant to develop connections with the project team and donors, as well as to potentially identify sources of outcome data that could be collected during the course of the project. Instead of simply conducting a retrospective examination that relies on whatever project documents, data, and stakeholders are available near the end of the project, the evaluator could instead collect data or conduct interviews periodically throughout the project. Getting evaluation consultants involved at an earlier stage would not necessarily cost more, but would simply stretch the involvement of the consultant out over a longer period of time.

Second, EVAL may wish to consider minimum levels of evaluation investment, based on the size of the project. As mentioned previously, actual evaluation expenditures represented slightly less than 1.5% of project budgets during the analysis period. Although it is understood that ILO evaluation policy guidelines require 2% of the project budget to be devoted to evaluations, the actual expenditure ratios are low.

D. Undertake some research to determine whether or not the quality assessments are measuring the right concepts.

The final question to consider is whether or not the quality assessments are correctly measuring evaluation quality. This statement is not a reflection on the individuals and organizations that have conducted past quality assessments; indeed, the consistency of scores and findings over time actually suggest that the studies have been conducted in a manner that is well-aligned with EVAL’s own guidelines. Instead, it is meant to suggest that some consideration should be given to both the practical and philosophical aspects of quality. In a practical sense, the quality assessment process may need to be revised if other steps are taken to modify how evaluations are conducted. As the prior ideas have alluded to in this article, there may be a case for making substantial changes to the checklists, guidelines, standards, and even the managers’ survey.

¹⁵ Evaluation Roundtable. (2010). Use of Evaluation Information in Foundations: Benchmarking Data. Presentation by Patrizi Associates to the Evaluation Roundtable. Available online at <http://www.evaluationroundtable.org/documents/et-benchmarking.ppt>

¹⁶ For one discussion of evaluation costs as a portion of project budgets, see: http://betterevaluation.org/plan/manage_evaluation/determine_resources

In a philosophical sense, it may also be beneficial to ask whether the current manner of assessing quality truly aligns with the views and needs of all project evaluation stakeholders. Although ILO already has established policy¹⁷ regarding the official purpose of evaluation, the policy does not explicitly set criteria for how a “good” evaluation is defined, nor does it necessarily capture all the expectations that stakeholders have for an evaluation. Any discussion should begin with defining what a “good” evaluation should look like, both in terms of how evaluative findings are conceptualized and how those findings will be used by stakeholders. Questions should also be asked regarding whether or not the existing model fits with the reality of the evaluations that are actually produced and how they are used. For example, how does the current approach to quality assessment reflect how ILO uses the evaluations? Does an evaluation that meets quality standard provide the information that donor countries want to learn from an evaluation? Finally, how do the evaluation consultants in the field view their roles as evaluators? The answers to these questions should then drive how quality is viewed and assessed in the future.

5. Closing thoughts

This Think Piece has presented ideas for several different approaches to boosting the long-term quality of ILO project evaluations. Compared to the recommendations typically offered following a quality assessment, these ideas might seem to be too broad, costly, or untested. However, if a large and measureable jump in evaluation quality is desired, it seems evident that the previous approach of incremental changes in evaluation policy will be insufficient. At a minimum, it is hoped that the ideas presented here inspire a wide-ranging discussion regarding how best to define and pursue the evaluation goals of ILO in the future.

¹⁷ See ILO-EVAL. (2013). *ILO policy guidelines for results-based evaluation: principles, rationale, planning and managing for evaluations*, p. 3-4. http://www.ilo.org/eval/Evaluationpolicy/WCMS_168289/lang-en/index.htm