



i-eval THINK Piece, No. 8

External Quality Appraisal

Implications for evaluation quality and
utilization

Judith Friedman
Naomi Blight of IOD Parc, UK

International Labour Office
Evaluation Office

December 2014



External Quality Appraisal

Implications for evaluation quality and utilization

By Judith Friedman and Naomi Blight
of IOD Parc, UK

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.

Contents

Acronyms	ii
Introduction	3
Methodology	3
Key findings.....	4
Overall Findings of the 2012-13 Quality Appraisal	4
Key challenges affecting quality of evaluation reports	7
Quality Appraisal Process	7
Evaluation Guidelines	8
Institutional capacity and resources: Internal	10
Technical Capacity: External.....	17
Process and Utilization.....	19
Conclusion.....	22

Acronyms

CO	Country Office
DAC	Development Assistance Committee
DEFP	Departmental Evaluation Focal Point
DWCP	Decent Work Country Programme
AC	Evaluation Advisory Committee
EVAL	ILO Evaluation Office
ILO	International Labour Organization
IPEC	International Programme on the Elimination of Child Labour
OECD	Organisation for Economic Cooperation and Development
PARDEV	Partnerships and Field Support Department
PROGRAM	Strategic Programming and Management Department
QA	Quality Assurance
REO	Regional Evaluation Officers
RO	Regional Office
SMART	Specific, Measurable, Achievable, Realistic, Time-bound
TOR	Terms of Reference
UN	United Nations
UNEG	United Nations Evaluation Group

Introduction

The ILO places strong emphasis on ensuring that credible independent evaluations of its strategies, programmes and policies are conducted in accordance with international norms and standards, and in accordance with the expectations of both its constituents and donors. The ILO holds a long-standing commitment to conducting external quality appraisals of its independent evaluation reports. Meta-evaluations of independent evaluations have been conducted from 2005 to 2013 as a way of helping the ILO to improve the quality of its reports.¹

IOD PARC was commissioned by the ILO's Office of Evaluation (EVAL) to conduct an external Quality Appraisal of the reports of the independent project evaluation conducted by the ILO from 2012-2013, as stand-alone documents, in order to determine the quality of its reports through an external assessment and rating. The Quality Appraisal found that the overall quality of ILO-managed independent project evaluations had plateaued from previous years. The majority of evaluations are of acceptable levels of quality, or approaching an acceptable level. This Think Piece identifies a number of strengths across evaluations and also key areas for improvements which emerged from those evaluations falling in the lower quality category.²

Based upon the findings of the Quality Appraisal, this subsequent Think Piece has been commissioned to both disseminate some of the findings of the Quality Appraisal to a wider audience and attempt to take the analysis further to understand the underlying reasons for the quality performance determined by the appraisal by highlighting challenges observed. Beyond analysis of what the quality is, this Think Piece aims to understand why the quality of evaluations has not significantly improved in recent years.

Methodology

The conclusions of this Think Piece are drawn from the 2012-13 Quality Appraisal Report.³ For that appraisal IOD PARC appraised a random sample of 60 per cent of the 2012-2013 project evaluations using an appraisal tool developed by EVAL. In addition, IOD PARC administered a short questionnaire to evaluation managers on compliance with evaluation policy guidelines, conducted interviews with five ILO staff, and consulted other relevant ILO documentation.

The Quality Appraisal identified areas of strength, weakness, and overall trends in quality performance through the lens of international evaluation standards and ILO's own evaluation policies and guidance. A clear picture emerged revealing insights on the quality of specific evaluation components and identifying specific areas for improvement.

¹ N. Blight, J. Friedman and R. Polastro. IOD PARC: *Quality Appraisal of Independent Project Evaluations 2012-2013*, (Geneva, ILO internal report, 2014). Hereafter referred to as Quality Appraisal Report 2012-13.

² Ibid.

³ Ibid.

In its appraisal, the IOD PARC team applied ILO evaluation principles to appraise the quality of the evaluation key components (executive summary, project background, evaluation background, methodology, findings, conclusions, and lessons learned, emerging good practices, recommendations and appendices). An evaluation report was assessed as ‘good quality’ when it addressed the evaluation purpose and objectives based on evidence, providing clear and practicable recommendations, and lessons for project improvement for the evaluation to be used with confidence by project management and stakeholders.⁴

To understand the underlying reasons for quality remaining near the same, not improving, or lagging behind in certain areas, the IOD PARC team considered ways that the different functions of ILO’s evaluation system help or hinder quality.⁵ Further enquiry into the characteristics of internal and external resourcing, utilization of guidance, capacity development, management, and quality assurance processes were explored through five open-ended interviews with ILO evaluation managers at different levels. In this way, the relevance of key findings of the Quality Appraisal were further explored and substantiated with additional evidence.

Limitations

A key limitation of this Think Piece is that there were limited opportunities to gather additional data, beyond the Quality Appraisal exercise. As mentioned above, the IOD Parc team conducted five staff interviews, used the results of the evaluation manager survey and analyzed additional relevant documentation to supplement the findings of the Quality Appraisal. The Think Piece team has therefore drawn conclusions from the evidence available and from tacit knowledge of EVAL’s staff.

Key findings

Overall Findings of the 2012-13 Quality Appraisal

Overall, the majority of the evaluations reviewed were found to be of an acceptable level of quality, or approaching an acceptable level of quality.⁶ The appraisal found that 11 per cent (6) of evaluations scored 1.5 or lower, 42 per cent (23) scored above 1.5-1.9 and 53 per cent (29) scored 2.0 or more; therefore the majority of evaluations are of or are approaching acceptable levels of quality.⁷

⁴ Blight and Friedman, op. cit.

⁵ System Dynamics, as defined by Sterman, looks at the behavior of complex systems; For the ILO, we have looked at the interaction between the different evaluation processes and systems to find relationships and interdependencies across the system. J.D. Sterman: *Business Dynamics: Systems thinking and modeling for a complex world*, (New York, McGraw-Hill, 2000).

⁶ Quality has slightly improved from 1.8 as noted in the 2009-11 Quality Appraisal, to 1.9 for the current 2012-13 Quality Appraisal exercise. It is important to note here that the mean aggregate scores presented below are calculated from the components that are present in the reports, rather than including those components that were missing. The appraisal found that 11 per cent (6) of evaluations scored 1.5 or lower, 42 per cent (23) scored above 1.5-1.9 and 53 per cent (29) scored 2.0 or more; therefore the majority of evaluations are of or are approaching acceptable levels of quality.

⁷ Items were scored as zero 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 the 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). Where the scores included 88 and 99, adjustments were made to ensure that the scores were not skewed.

The team looked for quality trends across years and by thematic areas and regions. However, this revealed no significant differences in quality across the data sub-sets. For example, the five highest scoring evaluations all scored 2.3 and were from different administrative units, different regions and covered different thematic areas. It is particularly challenging to compare average quality scores between regions or thematic area because of the small sample size of evaluations appraised. The Quality Appraisal team found that the overall quality of the 2012-13 ILO-managed independent project evaluations was slightly improved from the previous 2011-12 sample (1.8), with a mean aggregate average score of 1.9, based on a comparison of the aggregate mean score of the datasets.

It is important to note that the evaluations were judged against quality criteria, based upon checklists and guidance that were created and disseminated only after a number of the evaluations had been completed. Given that many of the evaluations assessed in the sample were not able to benefit from the most recent checklists and guidance, there are elements that may have been improved had the updated guidance and checklists been available at the time the evaluation was actually undertaken.

Below, we discuss key strengths identified by the Quality Appraisal as well as the weaknesses and areas for improvement identified. These weaknesses are addressed in the recommendations and are aimed at improving the issues that emerged from the number of evaluations which fell below the acceptable quality rating.

Key Strengths

The following characteristics and practices were noted as strengths in the ILO-managed independent project evaluations reviewed in the Quality Appraisal of 2012-13 evaluation reports:

- **Robustness in data analysis and clarity in articulating a context:** This provided a sound basis for understanding the intervention logic and the factors that contributed to success or failure of the project, therefore lending the evaluations to greater confidence to act on the basis of the findings, conclusions, recommendations and lessons.
- **Description of the project's objectives:** A thorough and analytical description of the object of the evaluation reflected competence in the evaluation team which added to the credibility of the evaluation report.
- **Description of the context and intervention logic:** This included background information on the national and project context, and how the project sits alongside national development priorities. Actions were clear and demonstrated the relevance of the project.
- **Discussion of factors that have contributed to the success/failure of the project (or elements) were identified and discussed:** External factors contributing to success/failure were identified and analysed, including the social, political or environmental situation.
- **Impartial and credible conclusions:** Generally, the conclusions helped to clarify the findings that were presented and set out issues that are relevant to wider considerations for the ILO or for the development intervention.

- **Recommendations:** Recommendations were supported by evidence and followed logically from the findings, conclusions, lessons learned and emerging good practices if applicable; specifying clear action and associated responsibilities.

Areas for Improvements

Many of the components that were identified as areas for improvement may have extended from incomplete use of ILO Guidelines for Evaluation, particularly Checklist 5 and Checklist 6⁸, or because these resources were not available when a number of the evaluations were conducted. This is also mirrored in the fact that many key components were missing from a number of the evaluation reports. Areas for improvement relating to the evaluation reports, the evaluation guidance and evaluation process included:

- **Lack of specific description of methodological approaches:** This meant that key aspects of data integrity and validity, and limitations of methodological approaches could not be fully interrogated by reading the evaluation report.
- **Limited reference to evaluation questions and (OECD/DAC) criteria:** This was a critical element for determining the quality of the evaluation undertaken as the evaluation questions are the most important lens for interpreting the findings and evidence presented.
- **Lack of description of norms, standards and ethical safeguards:** Given the often sensitive nature of the evaluations undertaken and the vulnerability of certain stakeholder groups, there is a strong rationale for outlining ethical considerations in evaluation reports.
- **Findings were not adequately supported by evidence or data:** Reports reviewed generally presented limited disaggregated data and did not refer adequately to the internal and external sources explicitly showing how the evidence had been gathered and the findings substantiated. In many of the reports, data was not sufficiently disaggregated by sex, age, ethnic group or other relevant demographic categories, to the degree that is expected.
- **Recommendations often do not distinguish priority or importance in broad terms:** Few of the evaluations demonstrated that recommendations had been developed in collaboration with stakeholders; which is crucial to engaging the 'buy-in' of stakeholders. Additionally, the recommendations often do not acknowledge resource implications in terms of time, staff, and budget.
- **Lessons insufficiently targeted:** Lessons often did not identify target users of the lessons or good practice and the impact on them. Lessons also did not demonstrate their link with specific impacts.
- **Reference in TORs to key guidance:** TORs do not consistently refer evaluators to ILO and UNEG guidance and did not always include key information.

⁸ ILO, Evaluation Office, March 2014. [I-eval Resource Kit, Checklist 5: Preparing the Evaluation Report and Checklist 6: Rating the Quality of Evaluation Reports](http://www.iilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_165967.pdf). [Online] Available at: http://www.iilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_165967.pdf

Key challenges affecting quality of evaluation reports

Quality Appraisal Process

Eval seeks to improve its evaluation quality by conducting regular appraisals and meta-studies in order to provide continuous feedback on the evaluation workflow, processes, and products. The ILO's initiative to learn more about the status and results of decentralized evaluations by commissioning these meta-studies is an example of good practice as it provides an opportunity for learning and accountability, to identify quality trends and to make recommendations for improvements.⁹

Independent project evaluations in the ILO are largely conducted to assess project success and improve project design. According to interviews with ILO staff, one of the key challenges potentially affecting the quality is that there are varying perceptions and definitions of quality for an evaluation across the ILO.

As mentioned in the limitations section, the Quality Appraisal consisted largely of the examination of evaluation reports as stand-alone documents. The tool for the Quality Appraisal is based upon prescriptive and precise UNEG standards and ILO checklists and guidance. The benefit of this approach is consistency against a set of standards but the limitation is the reliance on a single source of information (the evaluation report) to develop a view on the quality of an evaluation. This may not fully account for the evaluation process or the quality of the analysis that lead to evaluative judgments, or the use of evaluations to enhance accountability, performance, and learning in ILO.

Interviews with ILO staff highlighted the potential difficulty in seeing the quality of evaluation reports only through the assessment derived from the Quality Appraisal process; and that both the process of conducting evaluations and the utility of evaluations are important to consider in terms of an evaluation's quality. For example, evaluations may score well against the components of the checklists, but may not generate sufficiently useful recommendations and lessons for use by project teams. Similarly, reports may have low scores against the checklists but may be of high analytical quality, or have useful, actionable recommendations.

Weighting evaluation components

Furthermore, from the appraisal process itself, the review team found that components of the evaluation reports were not weighted according to which aspects of the evaluation were more important to delivering a credible product. Whilst all components of an evaluation articulated in Checklist 5 should be present, some are critical while others simply enhance the overall quality. This means that certain nuances related to this lack of weighting could not be reflected in the overall quality rating.

⁹ S. Prom-Jackson and G. Bartsiotas: *Analysis of the Evaluation Function in the UN System 2014, Volume I*, (Geneva, Joint Inspection Unit: United Nations, 2014).

Recommendation 1 (For EVAL): Consider weighting the component scores of the evaluation reports within the checklists and the Appraisal Tool to make explicit the critical elements of the evaluation report versus those that simply enhance the overall quality, (e.g. description of evaluation methods being weighted as more important than operational sequences of the evaluation).

Evaluation Guidelines

ILO evaluations are guided by ILO policy guidelines which set out principles, rationale and protocols for planning and managing evaluations. ILO guidance states that evaluations should '*meet ILO evaluation quality standards, which are consistent with and directly inspired from the UNEG norms and standards*'. The ILO has evolved substantially in the depth and breadth of its guidance since 2012 and has a strong and coherent evaluation policy.

It is important to note here that the evaluation policy guidelines, the supplemental guidance and the checklists were only introduced in 2012 and then updated in 2014 after the 2011-12 Quality Appraisal exercise. Consequently, some of the evaluations conducted in 2012 were undertaken before any ILO evaluation guidance checklists were in place and many of the evaluations conducted in 2013 that were part of the review did not have sight of the most recent checklists and guidance. This means that there are aspects or components that might have been improved had the updated guidance been available when the evaluations which were included in the quality appraisal were conducted.

Guidance and Checklists

ILO evaluation policy is supported by supplemental guidance which provides instruction for managing the evaluation process and product. There is the policy document, 16 guidance notes and an additional 10 checklists. This supplemental guidance is prepared by EVAL and continually updated to improve evaluation quality.¹⁰ [Guidance Note 6: The evaluation manager – Role and function](#) is intended for use by internal, volunteer evaluation managers to assess the compliance with guidelines, manage and advise evaluation consultants, and to control the quality of the evaluations before final submission for approval. [Checklist 5: Preparing the evaluation report](#) is addressed to evaluation managers and consultants to provide guidance on specific requirements for drafting the formal elements of the report, as well as details on how to formulate and present conclusions, recommendations, lessons learned, and emerging good practice in the report. Additionally available, for quality control, primarily for Terms of reference and reviewing finalized evaluations are: [Checklist 1: Writing terms of reference](#), [Checklist 2: Rating the quality of terms of reference](#), and [Checklist 6: Rating the quality of evaluation reports](#).

The Quality Appraisal found that despite the provisions of Guidance Note 6 and Checklist 5 in 2013, many of the evaluations reviewed did not follow the instructions in full. This resulted in missing components and evaluations that did not reflect ILO

¹⁰ ILO evaluation guidance and checklists were created in 2012 and revised in March 2013 following recommendations from the 2010-2011 Quality Appraisal process. See [i-eval Resource Kit - ILO policy guidelines for results-based evaluation: Principles, rationale, planning and managing for evaluations](#).

evaluation standards. As explained above, this is partially due to the fact that the checklist and guidance notes applied by the Quality Appraisal were finalized and came into force only in March 2013.

Although most evaluation managers surveyed indicated that they regularly referred to existing ILO policy and guidance, and had a practice of referring evaluators to the checklists, there is an apparent variability in the application of the checklist to evaluation practice.¹¹ Indeed, in interviews conducted, colleagues stated that the use of and 'sign-off' of evaluation reports against the checklist was not undertaken systematically.

The quality assessment tool for undertaking the Quality Appraisal of evaluation reports is closely aligned to the 2013 evaluation checklists. As such, adherence to and utilization of these checklists is essential to receiving high quality scores. The checklist content is based upon UNEG evaluations standards which present clear and demanding standards for the content of evaluation reports.

For the checklists to be used effectively, there also needs to be a level of 'buy-in' into their importance for use by evaluators. Additionally, evaluation managers must find them to be accessible and fit for use. For the checklists to be used effectively by both evaluators and evaluation managers alike there must be sufficient evaluation expertise to deliver and manage the evaluations. With this, the evaluations can be assessed not only with their compliance but in their fitness for use to provide findings and recommendations that can improve the project.

Mixed assessments of the suitability of the checklists and guidance for ILO evaluations were noted by the Quality Appraisal. Some interviewees felt that the guidance played an instrumental role in supporting quality EVAL processes, and others indicated that the guidance notes and checklists were excessive in their level of detail and could be confusing. Interviewees also discussed a reluctance to make the quality assessment of evaluations purely a 'tick-box' exercise.

Across the evaluation function, the checklists are used as a general guide for ensuring compliance with ILO and UNEG Evaluation Standards. Evaluation managers are encouraged to apply them to control quality. An approval checklist to control the quality of finalized reports is also available.¹² Although checklist use is recommended by EVAL, there is no standard protocol for oversight of the use of the ILO checklists. Compulsory sign-off would provide a mechanism for ensuring that the checklists are routinely used, thereby ensuring inclusion of all evaluation components.

Recommendation 2 (for EVAL): Consider making use of certain quality checklists compulsory (and for inception reports where applicable), with verified sign-off protocols implemented at the different quality appraisal and approval processes.

¹¹ Blight and Friedman, op. cit.

¹² ILO. Evaluation Office: [Checklist 6: Rating the quality of evaluation reports](#). (March 2014).

Terms of Reference (TOR)

Another critical element guiding evaluation quality is the evaluation TOR which should outline the scope of an evaluation, present the evaluation questions and introduce a proposed methodological approach. TORs for ILO-managed independent project evaluations are developed by project management and the evaluation manager, which should be guided by the quality checklists provided by EVAL.¹³ The TOR is then circulated to key stakeholders to provide comments. Other than the requirement inherent in applying the quality checklists for writing and rating their quality, the TORs are not otherwise standardized across ILO evaluations. Regional Evaluation Officers and Senior Evaluation Officers approve TORs as part of the quality assurance. The lack of standardisation was discussed in interviews where some ILO staff expressed that standardisation of the TORs might result in less flexibility, and consequently in evaluation reports that would be less responsive to individual projects.

Although the examination of individual evaluation TORs was beyond the scope of the Quality Appraisal, the Quality Appraisal found that UNEG norms and standards were not consistently applied in all evaluation TORs. EVAL held a Biennial Evaluation Workshop to review of ILO's Evaluation Strategy 2014-15 in 2013 and participants identified that better guidance for and model templates of TORs is needed.¹⁴

A TOR which is quality controlled will positively affect the overall quality of evaluation report components. This is particularly important in terms of the evaluation questions, which were found to be an area of quality concern in the Quality Appraisal. This is a critical element for determining evaluation quality as the evaluation questions are the most important lens for interpreting the findings and evidence presented.

Recommendation 3 (for EVAL): Consider implementing a more formal approval process for TORs that will control the quality of the TORs and ensure consultant compliance to the TORs. This would subsequently lead to a more complete evaluative product that includes all critical evaluation components (as detailed in Checklist 5).

Institutional capacity and resources: Internal

The rise in number of evaluations

Since 2009, there has been a gradual increase in the total number of independent evaluations conducted by the ILO, including ILO-managed independent project evaluations commissioned by the ILO, the Regular Budget Supplementary Account (RBSA) evaluations and externally-managed joint evaluations, as seen in the table below:

¹³ ILO. Evaluation Office: [Checklist 1: Writing terms of reference](#), and [Checklist 2: Rating the quality of terms of reference](#). (March 2014).

¹⁴ ILO, Biennial Evaluation Workshop, Review of ILO's Evaluation Strategy 2014-15, (ILO/EVAL internal report, 2013).

Table 1: Number of Evaluations and Mean Aggregate Quality Scores 2009-2013

Year	No. of ILO Managed Independent Evaluations ¹⁵	Externally Managed Joint Evaluations	RBSA Evaluations	Total number of Independent Evaluations	Mean Aggregate Score of ILO Managed Independent Evaluations
2009	52	2	1	55	1.9
2010	54	6	8	68	1.9
2011	47	47	2	96	1.7
2012	47	31	6	84	1.9
2013	56	34	3	93	1.8

Whilst the **ILO-managed independent project evaluations** were the only reports assessed as part of the 2012-13 Quality Appraisal, and there has not been a notable increase in the number of these, there has been a significant increase in the number of externally-managed joint evaluations (2 in 2009 to 34 in 2013), meaning an increase in the total number of independent project evaluations from 55 in 2009, to 93 in 2013.

In practice, this has led to a greater demand on ILO human resources allocated to ILO-managed evaluations as well as the time necessary to support the big increase in externally-managed joint evaluations. This continuous rise in the number of evaluations - requiring attention from evaluation officers and evaluation managers - has not been matched by a proportionate increase in human or financial resources. Correspondingly, in 2011, when there were significantly more evaluations conducted than the preceding two years, there was a dip in the aggregate mean score of the quality of ILO-managed independent project evaluations. A similar pattern occurred in 2013 with a peak (in ILO-managed independent evaluations) in the number of evaluations and a decrease in the mean aggregate score of the quality of those evaluations.¹⁶

Management of evaluations

Through the ILO's decentralized evaluation function, there are five Regional Evaluation Officers (REOs) and eight Departmental Evaluation Focal Points (DEFPs). These officers and focal points are further supported by three Senior Evaluation Officers based in the ILO Headquarters in Geneva. An evaluation manager is appointed to each evaluation. The evaluation manager's role¹⁷ includes working with the project team to identify the relevant project data, to participate in the drafting of the TOR, to work with senior evaluation officials to recruit and select the evaluation team, to support project management to coordinate and manage any field study, guide the consultant through reference to EVAL policy guidelines, guidance and checklists, and perform the initial quality appraisal¹⁸ before submission of the finalized report to EVAL. As indicated in Figure 1, the decentralized evaluation manager function is overseen by a small number

¹⁵ The ILO-managed independent project evaluations were the reports examined for the Quality Appraisal.

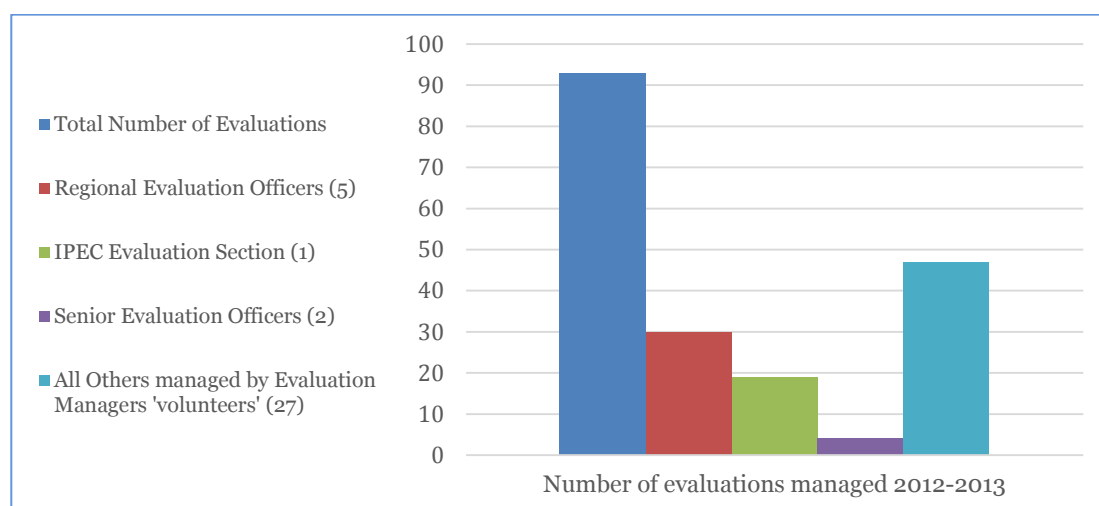
¹⁶ It is also important to note that the different appraisal methodologies applied by different appraisal teams over the years may also have had a bearing on the quality scores given to the reports, although efforts were made by this year's appraisal team to ensure consistency with previous scoring, so far as was possible.

¹⁷ ILO. Evaluation Office, op. cit. [Guidance Note 6: The Evaluation manager: Role and Function](#).

¹⁸ ILO. Evaluation Office: [Checklist 6: Rating the quality of evaluation reports](#), op. cit.

of REOs, DEFPs and EVAL officials who also manage a high volume of project evaluations. Evaluation managers are volunteers and are not necessarily evaluation specialists. They might manage between one and three evaluations per year in addition to their full-time commitments.

Figure 1: Evaluation Management by Type (2012-2013)



Interviews with ILO staff suggest that human resources for the management of evaluations are stretched as regional or departmental focal points have a high volume of evaluations and have limited time to backstop the volunteer evaluation managers who must integrate this additional work into their already full-time workload. Managing evaluations effectively means extensive contact with project staff, and ensuring that the consultant fully adheres to ethics, contract and content requirements. According to ILO staff, finding and engaging volunteer evaluation managers with the adequate skills, capacity and availability to manage independent evaluations is a continuing challenge facing the ILO.

As shown in Table 2, there are also regional disparities between the number of evaluations managed across ILO; with a significantly higher number of evaluations taking place in the Africa and Asia regions.¹⁹ There was no significant difference, however, found between regions regarding the mean aggregate Quality Appraisal score.²⁰ The Biennial Evaluation Workshop Review of ILO's Evaluation Strategy 2014-15 stated that work volume analysis needed to take place in Asia and Africa and the necessary adjustments made in staff allocations. Given the size of the disparity of evaluation numbers between regions (1 evaluation planned for Europe in 2014, 28 in Africa), it seems likely that there is a need for further analysis of staffing and reinforcement of evaluation capacity. Additionally, where there are complex evaluations of outcome-based programmes and global products, additional capacity may be required.

¹⁹ It is particularly challenging to compare average quality scores between regions because of the small sample size.

²⁰ The Biennial Evaluation Workshop Review of ILO's Evaluation Strategy 2014-15 stated that work volume analysis needed to take place in Asia and Africa and the necessary adjustments made in staff allocations. Given the size and the disparity of evaluation numbers between regions (1 evaluation planned for Europe in 2014, 28 in Africa), it seems there is a need for further reinforcement of evaluation capacity.

Table 2: Number of Evaluations by Region (2013 and currently in the pipeline)²¹

Americas		Africa		Arab States		Asia		Europe	
2013	Planned for 2014	2013	Planned for 2014	2013	Planned for 2014	2013	Planned for 2014	2013	Planned for 2014
24	14	30	28	5	2	21	21	6	1

Recommendation 4 (REGIONS): Consider further resourcing the evaluation capacity in both the Africa and Asia regions to reflect their significantly larger evaluation workload. Consideration for further resourcing should also be given to departments with complex evaluations of outcome-based programmes.

Volunteer evaluation managers

The REOs and DEFPs are responsible for recruiting additional volunteers to manage evaluations where they themselves have insufficient capacity. All evaluation management takes place within this scope. Senior management²² in EVAL believes that this can compromise evaluation quality, as the volunteer evaluation managers are often non-specialists and have little experience with evaluation. Volunteer evaluation managers are recruited by REOs or EFPs based upon their availability, and more often willingness, to take on this role on a voluntary, uncompensated basis. Interviews with ILO staff indicated that the time required to manage an evaluation is not always explicit and is, in practice, highly variable. This role must often be added to their routine responsibilities and the additional work time required to carry out this role is not consistently recognized in work plans.

Discussion at the Biennial Evaluation Workshop ²³revealed that some of the participants felt that the current recruitment conditions for volunteer evaluation managers are unsustainable and that a more formalized acknowledgement of this increased workload needed to be considered. The workshop participants also highlighted an urgent need for clarification of the role of the evaluation manager at the launch of the evaluation process, specifically for all officials involved in the evaluation, especially project management and Regional and Departmental Directors.

There is therefore a need to recognize this contribution of volunteer evaluation managers. In addition to providing capacity development opportunities, it would also be advantageous to incentivize this level of engagement. The role of the evaluation manager demands an investment of time which should be recognized within their work portfolio or within annual work plans in order to earmark the time spent managing the evaluation. Although there is wide variation in the amount of time required by each evaluation, it is estimated that it would range from 10 to 30 days, depending on the

²¹ These figures were obtained directly from ILO.

²² From interviews conducted by the authors.

²³ ILO. Evaluation Office, Workshop, op. cit.

complexity, the number of iterations required of the TOR and the number of drafts for each evaluation (or 4-13 per cent of a full-time post).²⁴ This is further detailed in Table 3 below. Where the evaluation capabilities of the evaluation manager are at a lower level of experience, it could be anticipated that more time would be required to deliver a quality evaluation product.

The ability of the evaluation manager to be able to carry out the role and function effectively has a direct and inevitable impact upon the quality of the evaluation reports. This relates both to capacity and to time allocated to the role.

Table 3: Estimated time investment for Evaluation Manager Role Responsibilities

Evaluation Manager tasks²⁵	Estimated Time
Works with project management to draft evaluation TOR	1-4 days
Circulates TOR to stakeholders and makes revisions	0.5 – 1 day
Procures evaluation consultants (consults roster, open procurement)	2-5 days
Negotiates terms and finalizes consultant arrangements	0.5-2 days
Works with project staff to ensure consultant is provided with adequate documentation and access to data	1-4 days
Ongoing dialogue with evaluation consultant	2-4 days
Field Visit	0-5 days
Carries out quality assurance of draft report	1-2 days
Circulates draft report with consolidated comments to stakeholders	1 day
Submits draft report to REOs or DEFPs	1 day
Endorses payment for consultant and forwards report onto stakeholders	1 day
Total	10-30 days

Recommendation 5 (for EVAL, HQ Departments, IPEC, and Regional Offices): Evaluation Managers' time investment should be consistently anticipated and integrated into annual individual work plans; thereby warranting recognition and acceptance from senior managers of the importance of this role.

Supporting volunteer evaluation managers

Because of the constraints discussed above, the role of evaluation manager is often occupied by a technical expert rather than an evaluation practitioner. To help to improve and incentivize this function, EVAL has developed a new training and

²⁴ Calculations were based upon a rough estimate of 225 working days per year.

²⁵ This is based upon [Guidance Note 6: The evaluation manager – Role and function](#), op. cit.

certification programme for evaluation managers.²⁶ The ILO Evaluation Manager Certification Programme is a part of the ILO's Evaluation strategy "*to expand evaluation capability in the form of knowledge skills and tools.*"²⁷ The programme involves a three-day workshop and a guided practice where trainees manage an evaluation under supervision of Senior EVAL officers, REOs and EFPs. Participants of the training receive certification after successful completion of the practicum which has recognition and credibility across the ILO. According to ILO staff, so far, the certification programme has been well received; representing a step to support institutionalization of evaluation with the ILO.²⁸ The certification programme is positioned to address a key capacity gap among evaluation managers and will lend itself to greater oversight of evaluation practice. For the certification to remain effective, certified evaluation managers are included in an internal evaluation network which promotes feedback, updates on training and improvements in guidance, and ensures representation and participation from across regions and departments.

Recommendation 6 (for EVAL): Continue to use regular feed-back processes from the Evaluation Manager Certification Programme participants to monitor quality improvements and gaps in knowledge for on-going refinement and improvement of program.

Choice of Evaluations

The ILO mandates evaluations based upon a threshold of expenditure, where all projects over US\$1 million²⁹ are subject to an independent evaluation. It should be noted that this threshold was increased from US\$500,000 to US\$1M in 2011. Additionally, if a project is perceived to be very complex, or a new and innovative approach, a strategic decision to evaluate these projects may also be taken. Projects below US\$1m are also subject to evaluation; but only internal evaluation which is not subject to quality control.

Specific funding modalities also make it reasonable to cluster evaluations around a thematic technical area. This can, however, pose challenges linked to different project starting and ending dates, synchronicity, and donor demand. The implication of this is that by remaining focused and selective in its evaluations, ILO can achieve strong results in terms of quality and continue to make the best use of evaluation technical assistance provided by regional and technical experts. EVAL should consider adjusting the independent project evaluation threshold to reduce volume of work in order to raise quality in general.

²⁶ The ILO Evaluation Manager Certification Programme was developed based on a need analysis carried out in 2012-13.

²⁷ ILO. Evaluation Office. *Evaluation Manager Certification Programme Handbook*, internal document, 2013.

²⁸ S. Prom-Jackson, op. cit.

²⁹ This was increased from US\$500, 000 in 2011.

Recommendation 7 (for ILO project management): Where possible, the ILO should find opportunities to carry out cluster evaluations (e.g. by country, thematic area, departmental) as an efficiency measure to reduce the volume of individual project evaluations; in favour of clustered, and more strategic and thematic evaluations.

(For EVAL): Where possible, EVAL should improve the selection of evaluations to be undertaken; ensuring this meets EVAL guidelines and donor requirements while remaining strategic. Options to undertake evaluations by country, thematic area or donor should be explored to the fullest extent possible. Ultimately EVAL should try to do fewer project evaluations but increase its central capacity to undertake these to a higher level of quality.

Resources

The ILO, like many UN agencies, is increasingly faced with the paradigm of doing more or the same amount of work with an increasing demand from donors for efficiency savings and resource constraints. EVAL has also seen the number of evaluations rise year-on-year, but this rise in the number of evaluations has not been met with proportionately increased capacity or increased spending.

Though it is clear that resource constraints (financial and human) are a key challenge to improving the quality of ILO evaluations, it must be acknowledged that there is not currently any potential to increase resources and to invest in certain areas that would result in improved evaluation quality (for example, increased paid and professional evaluation capacity in regions, higher evaluation budgets or fee rates for evaluation consultants).

As per ILO evaluation policy, a minimum of two per cent of the total project funds should be reserved for ILO-managed independent project evaluations. Because of economies of scale, larger projects can negotiate with EVAL to adjust the investment in an evaluation downwards. Interviews with ILO staff indicated that costing evaluations based on real evaluation and monitoring costs was indeed more effective than just using a flat rate.

Recommendation 8 (for ILO project management): Project design should ensure that 'real' evaluation budgets are built into projects and these are reflective of the project type, need and complexity.

Time Crunch

As mentioned above, time is a key challenge affecting the quality of evaluation reports; both for the managing and conducting of evaluations. Evaluation managers may struggle to take on the time demands of evaluation management due to capacity issues. This may impact on the quality of an evaluation in a number of ways: limited time to prepare and validate/seek feedback on a TOR, difficulties in procuring appropriate

evaluation teams in a short time frame, challenges of supporting the evaluating team during the field study, and the difficulties of conducting a thorough quality appraisal.

While evaluation process was not assessed as part of the Quality Appraisal, there were a number of evaluations citing time constraints as a challenge to quality. Examples were insufficient time to review project documentation before the field visit, insufficient duration of field visits which resulted in less time for data collection, and delays in project start-up which caused problems with deadlines for completion of a final evaluation report.

Embedding provisions for evaluations into project planning from the outset of project design will increase the evaluability and subsequent quality of evaluations. Similarly, allocating sufficient time to evaluation managers to conduct evaluations will create an enabling environment that will result in greater attention to the importance of evaluations and more considered analysis at all stages of evaluation management.

Technical Capacity: External

Evaluation Consultant Expertise

The ILO commissions independent project evaluations for highly technical and policy specific projects. There is sometimes a dichotomy between the need for strong technical experience with the need for strong evaluative skills. Limited financial resources and short time frames often mean that an evaluation consultant is chosen from a small pool or at a lower cost than a more specialized evaluator.

Often a choice is made between a i) consultant with technical expertise in a relevant thematic area, ii) an evaluation technical expert, or iii) an evaluation expert with vastly differing outcomes and approaches to the evaluation exercise. The choice between the three options also has a bearing on what the expectations are from the evaluation; for instance, delivering technically targeted recommendations, at the expense of the evaluation's credibility versus a well-targeted evaluation, without the depth of technical insight. Limited resources often mean that a team with combined experience is beyond reach. While a team that is equipped with technical and evaluation expertise may be the ideal, an evaluation professional with appropriate skills in inquiry and analysis would be well placed to carry out a high quality evaluation in a wide range of disciplines.³⁰

Recommendation 9 (for EVAL): It is recommended that EVAL include explicit guidance for the selection of evaluation consultants; striking a clear balance between the need for technical expertise and the need for evaluation expertise and how to make that judgment. The ILO should specifically consider the strength of the evaluator's inquiry and analytical skills as these are transferrable across technical areas. Where possible, technical staff backstopping a project could be used to support the drafting of TORs and the assessment of the inception reports.

³⁰ B.R. Worthen and J.R. Sanders: *Content Specialization and Educational Evaluation: A Necessary Marriage?* (2011, Occasional Paper Series, No. 14), Kalamazoo, MI, Evaluation Center, Western Michigan University.

Consultant Procurement

The ILO carries out an open procurement process to recruit evaluation consultants for independent project evaluations. Rather than relying on a small pool of known evaluation consultants, EVAL seeks to draw consultants from a wide network in order to promote independence and prevent bias. Recruitment of evaluators, and the time allocated for an evaluation, is at times rushed due to delays in the overall timeline of a project's implementation.

Independent evaluations are always undertaken by an external consultant and are selected through a consultation process, based on the objective criteria set out in the TOR, and with appropriate collaboration of the REOs, DEFPs, and evaluation managers, then approved by EVAL. ILO staff indicated that consultant selection criteria are based on experience with the UN, evaluation expertise, and relevant technical expertise. Although responsibility for finding and proposing an evaluation consultant rests with the evaluation managers, REOs and DEFPs are responsible for verifying the expertise of the consultant. Insufficient consultant expertise leads to poor quality findings, and poorly formulated recommendations and lessons learned.

The difficulties of finding and contracting good consultants were identified as important challenges. This could be addressed by improving the selection process for evaluators by ensuring technical and evaluation expertise. This also requires that budget and other logistical expectations are set realistically and that more expensive evaluators and teams of evaluation consultants be considered for larger assignments.

EVAL uploads its consultant information in an existing database which records all data on contact information, expertise, and provides links to any ILO evaluation work. Internal staff can review full reports and obtain the name of the evaluation manager who recruited the consultant. This system could be more routinely used to improve the recruitment of reliable consultants through providing more thorough assessment of the consultant's work performance.

The Australian Department of Foreign Affairs and Trade (DFAT), for example, have developed an 'Aid Advisor Performance Assessment'³¹ which aims to ensure that all parties are meeting mandatory reporting responsibilities. This allows an opportunity to identify current or emerging issues and confirms all parties are aware of the outcome as it is signed by both the government official and by the Advisor, who has the opportunity to respond to any issues raised in the assessment.

Recommendation 10 (for EVAL and Regional Offices): Consider how the existing database and knowledge management processes could be used more effectively to disseminate knowledge about the quality of evaluation consultants across the organization to identify a high quality pool of consultants - within the legal and data protection restrictions around rating and storing the performance of consultants.

³¹ Australian Government. Department of Foreign Affairs and Trade (DFAT): *Aid contractor performance assessment* <http://aid.dfat.gov.au/Publications/Documents/performance-assessment-template-contractor.pdf> (hyper link within this document to the Advisor Performance Assessment form).

Process and Utilization

Organizational Culture

Evaluations are conducted for three main purposes: accountability, organizational learning, and project improvement. The majority of independent project evaluations are carried out for the latter purpose and as such, their recommendations, lessons, and examples of good practice are meant to improve organizational performance and also to contribute to cross-organizational learning and sharing. EVAL has implemented a three-tiered quality assurance process applying the levels of quality checks:

- ✓ the evaluation manager level;
- ✓ the Regional Evaluation Officer or Departmental Focal Point level; and
- ✓ the Senior Evaluation Officer at headquarters

Additionally, there is a mandatory management response exercise initiated by EVAL to generate response from project management on recommendations.³² The exercise covers an action plan and a secondary follow-up to record conclusion of the action taken by project management. Though current analysis of this exercise shows an increasing percentage of recommendations being followed-up in a timely manner within one year of the evaluation completion, a small percentage of evaluations only record *commitments* on the part of management.

Aside from the immediate action taken by project management on evaluation recommendations, a broader use of evaluation findings as feedback in future programme or project design remains a challenge. EVAL's database can create data sets on lessons learned and evaluation recommendations based on technical topics and reports of this type have contributed to general discussions on the ILO's formulations of strategies on "areas of critical importance".

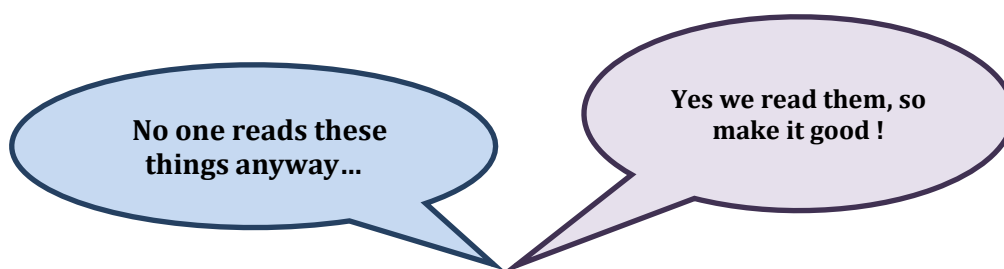
There is some suggestion that within the ILO's organizational culture, evaluation is seen as an obligation rather than an opportunity for active learning and engagement across project and institutional conduits. It is important to mention that this finding has been reported across UN agencies. While in the analysis of the evaluation function conducted across the UN system by the Joint Inspection Unit, ILO came out as one of the three top performers, it was nevertheless also found to face the same challenge as other UN agencies in developing an organizational culture where evaluation results not only serve accountability, but also feed into new project design and improvement.³³

How evaluations are used, and how they are perceived to be used, is a critical issue in considering quality. If evaluations are seen as an obligatory step of a project, then quality considerations will not be of great concern. Where evaluations are seen to be an integral component for project improvement across the organization, attention and effort will be placed into ensuring that they produce findings, recommendations, and lessons that are relevant and useful to stakeholders. Several interviewees reflected that evaluations could be better integrated in the organizational culture of the ILO.

³² ILO. Evaluation Office: [Guidance Note 15: Management follow-up for independent project evaluations](#) (Geneva, 2014).

³³ S. Prom-Jackson, op. cit.

Figure 2: The Evaluation Conundrum



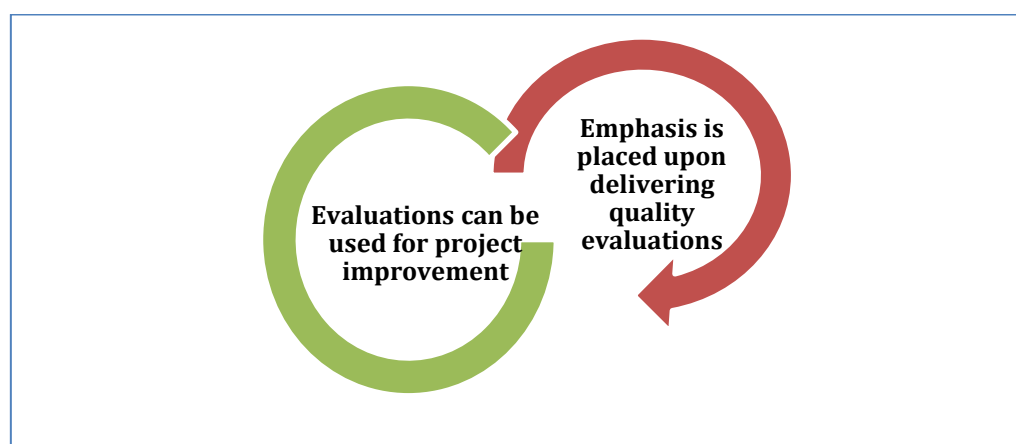
The establishment of an effective communication strategy would deepen the evaluation culture in ILO. EVAL's newly drafted communication strategy has planned information events, training and awareness-raising activities to increase the utility of evaluation products, which will include broader dissemination of evaluation findings. This was identified as a key area for action in 2014-15 and a communication strategy has now been drafted but was not, however, reviewed by the Quality Appraisal team as part of this Think Piece.

Placing a value on evaluations

Active knowledge sharing of the lessons and recommendations from evaluations can eventually have an impact on the quality of evaluations. Quality evaluations should likely lead to better projects and better projects should subsequently increase the value of quality evaluations (Figure 3). If there is emphasis on the use of evaluation findings, there would be a vested interest in making sure that evaluations are good in order to benefit others. This is a particularly important opportunity within the ILO given that evaluation users are evaluation managers themselves. This means that when more evaluations are actively read and used, the greater the interest in their quality.

It has been suggested that the participation of ILO staff who are technical experts in the evaluation process (and specifically in the Evaluation Manager Certification Program) could heighten awareness of the value of evaluations to different levels of the organization. Increased utilization of evaluation products linked with a level of quality that ensures their usefulness can have a cyclical or reinforcing impact upon strengthening evaluation culture in the ILO.

Figure 3: Evaluation Quality Feedback Loop



All project evaluations are kept in the EVAL *i-Track* database, with their recommendations, lessons learned and emerging good practices kept as separate data sets. EVAL and the Partnership and Field Support Department (PARDEV) collaborated to introduce a protocol in 2013 to ensure that relevant project evaluations are reviewed when a new project proposal on a similar topic is put forward for approval. New project proposal submissions should include a section that requires a citation of any evaluations which were reviewed on the related topic. However, as some interviewees pointed out, having good data sets on lessons learned doesn't mean that staff members will avail themselves of this information. Further collaboration on adhering to this protocol should strengthen this practice, and could be an activity covered by the EVAL communication strategy.

Recommendation 11 (for EVAL and PARDEV): Using its new communications strategy as a springboard, ensure institutional recognition of the utility of evaluations by developing mechanisms for better dissemination of evaluation findings. Using clear channels provide a facility where recommendations and lessons learned could be reviewed, discussed, and implemented to embed evaluation within ILO's organizational culture.

Project monitoring plans within the ILO

The Quality Appraisal found that "lack of evidence" was cited as a weakness in the evaluations reviewed. The dearth of data available to evaluators was either due to weak project monitoring systems that generated little or no data, or weak project design, all of which had a direct bearing on the quality of evaluations that consultants were able to produce. One other reason for this lack of data is that some projects are designed to influence change over the long-term. For instance, the monitoring of policy impact has limitations in its contribution to short-term contexts.

EVAL could enhance collaboration with PARDEV and PROGRAM to ensure adherence to appropriate monitoring components in project design. This should lead to better project data and a more evaluable program logic. Interviews with ILO project staff suggested that in many cases, even with strong evaluation management and a strong evaluator, evaluations are often weak because they are based upon weak evidence or projects that are built on an insubstantial logical framework.

Improving the evaluability of projects, particularly through improved project design, logical frameworks, data collection instruments, and monitoring components is critical to ensuring evaluation quality. Carrying out an evaluability assessment after one year for projects over US\$5 million is also particularly important.³⁴ A strong commitment by project managers to invest in proper monitoring of progress and results would further improve the quality of evaluations and is also likely to lead to improved use.

³⁴ ILO. Evaluation Office: [*Guidance Note 16: Procedure and tools for evaluability review of ILO projects overs US\\$5 million*](#) (Geneva, 2013).

Recommendation 12 (For REOs, EVAL staff, PROGRAM and PARDEV): Collaboration with REOs, EVAL staff, PROGRAM and PARDEV should be strengthened to ensure that project design has good monitoring practices and provisions embedded from the outset of projects and that evaluability assessments are undertaken as required.

Conclusion

The quality of the ILO-managed independent project evaluations is affected by resourcing issues and an organizational culture where the learning and accountability that extends from the evaluation function is not fully embedded. The lack of a fully embedded evaluation culture means that evaluation is not given sufficient priority and evaluation utility is not fully appreciated or realized.

It must be acknowledged that the ILO has been recognized by the UN Joint Inspection Unit (2014) for having taken strong initiative to improve both the internal evaluation culture and to institutionalize measures for efficiency and effectiveness in its evaluations. Examples of good practice include its attention to independence of the evaluation function, developing and recognizing the contribution of volunteer evaluation managers, and the continual provision and refinement of evaluation guidance for evaluation managers, project staff and evaluation consultants.

The relative uniformity in the quality of the evaluations across all ILO regions, where evaluations grapple with the same quality challenges, suggests that EVAL is working effectively across regional and country offices. Generally evaluation processes are being implemented systematically across the ILO's decentralized evaluation function and there is a certain level of uptake of the tools that are in place.

However, even with an evaluation system that is supported by clear and relevant guidance, there are constraints placed by resourcing and the challenges presented by an organizational culture that does not prioritize or suitably recognize the utility of evaluations. This means that quality issues persist.

Building upon the proactive steps that EVAL has taken to support its decentralized evaluation function and foster a culture of learning within the ILO, there is room to use current structures and existing policy more effectively to expand the potential of the tools that are in place, and build upon existing capacity. This will require increased recognition at all levels in the ILO of the value of the evaluation function and the potential this represents to improve outcomes and impact of ILO projects.

Recognizing that in the current environment, there is limited scope for increasing financial resources, small adjustments in the deployment of time and human resources and in dialogue and communication about evaluation can impact positively upon the quality of both evaluation reports and ultimately on the effectiveness and impact of ILO projects. It is hoped that the recommendations above will provide practical suggestions for using the ILO's existing systems and capacity to improve evaluation practice and thereby leading to a steady improvement in evaluation quality.