Independent Evaluation of ILO's Evaluation Function 2011–2016

Independent Evaluation of ILO's Evaluation Function – IEE 2011–2016

Final Report

International Labour Organization

December 2016

_____

TABLE OF CONTENTS

ANNEXES – available upon request (eval@ilo.org)

List of annexes

ABBREVIATIONS

DCOMM	Department of Communication
DEFP	Departmental Evaluation Focal Point
DFR	Draft final report
DG	Director-General
DWA	Decent Work Agenda
DWCP	Decent Work Country Programme
EAC	Evaluation Advisory Committee
EEC	European Economic Community
EFP	Evaluation focal point
EQs	Evaluation questions
EMCP	Evaluation Manager Certification Programme
EU	European Union
EVAL	Evaluation Office
FAO	Food and Agriculture Organization of the UN
GB	Governing Body
GED	Gender, Equality and Diversity Branch
HLE	High-level evaluations
HQ	Headquarters
HR/GE	Human rights/Gender equality and women's empowerment
IEE	Independent Evaluation of the Evaluation Function
ILO	International Labour Organization
ILS	International Labour Standard
ILO	International Training Centre of the ILO
JIU	Joint Inspection Unit
KAP	Knowledge, attitude and practice
M&E	Monitoring and evaluation
OECD/DAC	The Organisation for Economic Co-operation and Development/Development Assistance Committee

P&B	Programme and Budget
PARDEV	Partnerships and Field Support Department
PROGRAM	Strategic Programming and Management Department
RB	Regular budget
RBM	Results-based management
RBSA	Regular Budget Supplementary Account
REO	Regional Evaluation Officer
RO	Regional office
SDG	Sustainable Development Goal
SPF	Strategic Policy Framework
ToC	Theory of change
ToR	Terms of reference
UN	United Nations
UN SWAP	The UN System Wide Action Plan on Gender Equality and the Empowerment of Women
UNEG	United Nations Evaluation Group
VOPE	Voluntary Organizations for Professional Evaluation

ACKNOWLEDGEMENTS

EXECUTIVE SUMMARY

This independent evaluation of the evaluation function (IEE) has been commissioned to provide an assessment of the overall performance of ILO's evaluation function during the period 2011–16 along with explanations for this performance, lessons and good practices.

More specifically, the evaluation will focus on assessing two main interconnected objectives:

1. the relevance of the ILO evaluation strategy, relating to its *quality and capacity to inform sound decision-making in the ILO*; and

2. the effectiveness of its operational arrangements and structures, relating to the *appropriateness of the organizational set-up and processes*, specifically focusing on enhancing the ILO's evaluation capacity.

The overall scope of the IEE is the *"evaluation function"* of the ILO. For the purposes of this evaluation, the evaluation function is made up of two main elements:

- **central evaluation function:** the Evaluation Office (EVAL), the Evaluation Advisory Committee (EAC), and the systems, structures and evaluations under the control of EVAL, which oversee the activities of the decentralized evaluation function, and the final phase before the approval of all evaluations within the ILO; and

- **decentralized evaluation function:** evaluations under the control of programme, technical and field units; and the regional evaluation officers (REOs), departmental evaluation focal points or persons (DEFPs), and certified evaluation managers that support decentralized evaluations.

Four evaluation criteria were applied based on the United Nations Evaluation Group (UNEG) guidance of independence, credibility, utility and enabling environment.

BACKGROUND

In 2005, ILO adopted an evaluation policy framework, inspired by internationally accepted norms and standards that aimed to improve and strengthen independent evaluation in the ILO. As per point 46 of the 2005 evaluation policy, the first IEE was carried out in 2010. Since then, the evaluation function has been significantly transformed through the *Results-based strategies 2011–2015 Evaluation strategy – Strengthening the use of evaluations*. Some major developments have taken place, including:

- the transformation of the central Evaluation Unit into an independent Evaluation Office (both have the acronym "EVAL"), reporting on evaluations directly to the Director-General (DG);

- the development of codified evaluation policy guidelines (issued in 2012);

- the development and implementation of an evaluation communications strategy, including the preparation of a wide range of newsletters, think pieces, meta-studies, synthesis reviews, etc.;

- the introduction of a publicly accessible database to monitor and track evaluations, and store evaluation findings and full reports;

- the enhancement of the evaluation structure through the identification of regional evaluation officers and departmental evaluation focal persons;

- the available resources for evaluation activities have increased, including the budget channelled through the fixed share of programme support and extra-budgetary sources; and

- systematic training of evaluation managers through a certified course – Evaluation Manager Certification Programme (EMCP).

The 2011–15 Evaluation Strategy (extended to cover 2016-17) is organized along **three main outcomes**, which contribute to Outcome B "Effective governance of the Organization" of the ILO Strategic Plan development results framework:

- Outcome 1: Improved use of evaluation by ILO constituents and management for governance.

- Outcome 2: Harmonized office-wide evaluation practice to support transparency and accountability.

- Outcome 3: Evaluation capability expanded through enhanced knowledge, skills and tools.

During the evaluation period 2011–2016, ILO carried out a total of 616 evaluations making an average of 123 per year. Independent project evaluations accounted for 44 per cent of the total number of evaluations followed by internal evaluation reports and joint project evaluations. The highest number of evaluation reports was carried out in Africa and Asia, with 178 and 153 reports, respectively.

METHODS

The IEE applies three main design principles:

- **theory-based:** assessing the performance of the evaluation function based on a reconstructed theory of change (ToC);

- **utilization-focused:** maximizing the utility of the evaluation process and products to the primarily intended users; and

- **systemic:** taking into consideration the multi-dimensional aspects of the evaluation function in the ILO, including elements of organizational efficiency, among others.

These principles were realized through a mixed-methods evaluation using multiple lines and levels of evidence (qualitative and quantitative) in parallel as a means to triangulate and interrogate findings. Three main methods were used: (1) case studies from each ILO region; (2) a purposively sampled portfolio analysis of 20 evaluations; and (3) a realist synthesis of secondary evidence. These were combined through the use of contribution analysis.

The main sources of evidence were:

- **desk reviews** of more than 150 documents;

- **field visits** to regional offices (Abidjan, Bangkok and Geneva), and two virtual country visits to regional offices (ROs) (Beirut and Lima);

- **interviews** with 138 key informants;

- **surveys** (architecture and users) in three languages; and

- **portfolio reviews** of a selection of 20 evaluations using checklists and interviews with the evaluation consultant, evaluation manager and an intended user.

The evaluation team identified a number of limitations, assumptions and constraints that will have implications for the primary intended uses (purpose) of the evaluation. These are:

- the **selected methods prioritize learning** and do not provide statistically significant attribution;
- the evidence available to the evaluation is naturally **limited by the perceptions** and knowledge of informants;
- **practical time constraints** limit the level of triangulation possible in the final synthesis; and
- not all **UNEG evaluation criteria** are covered by the evaluation.

OVERALL FINDINGS

The IEE (2010) strongly emphasized the structural independence of the central evaluation function. As a result of the changes made following that evaluation, and the subsequent work of EVAL, the ILO is now recognized as having one of the three most mature evaluation functions in the UN system. This is a significant achievement.

EVAL has understandably emphasized its independence during the course of this transformation. As a small centralized office within the ILO, it has established mandatory requirements and highly structured systems and processes to deliver the evaluation function. These include some common tools, such as guidance and manuals, and some innovations, such as the EMCP and linking the *i*-Track database to other corporate information technology (IT) systems.

The consequence of this success is a systemized evaluation function that is strongly independent and consistently delivers its technical requirements to UNEG standards. This has attenuated the need for an integrated evaluation planning system, since nearly all current evaluations are mandatory and are triggered at mid-term and/or at the end of a project depending on the size of its budget. In future, a more flexible, integrated and utilization-focused approach will be increasingly demanded of EVAL. Meeting this challenge with the same small team and without sacrificing current coverage and quality will be the major challenge for the next evaluation strategy.

The cumulative estimate of regular budget (RB) expenditure on evaluation over the period 2011–2016 is US$26 million, equivalent to $4.3 million per year. Based on the biennial budget for ILO 2014–2015 ($400.6 million per year), evaluation represents 0.8 per cent of total budgeted expenditure, i.e. excluding extra-budgetary resources such as technical cooperation (TC) projects. Based on the recommendations of the United Nations Joint Inspection Unit (JIU) (2014), and given that ILO is a medium-size UN entity in JIU classifications, this evaluation would expect to see expenditure in the range of 1.5–2 per cent of the total organizational budget being allocated to evaluation. ILO is currently committing only half of this amount, which might explain the general perception among interviewees that there are inadequate resources and staff capacity to manage evaluations and follow-up on their outcomes.

INDEPENDENCE

Whilst the independence of the evaluation function has increased considerably, several limitations remain. The evaluation acknowledges that the small number of interviewees who expressed concern about independence did so because of the fact that the head of evaluation reports to the DG, or because of the conditions attached to the position. Having considered these perspectives, and triangulated them with other data, the evaluation is of the view that the main priorities regarding independence are, however, not related to these issues.

The highest priority regarding independence is to transition the REOs to being full staff members of EVAL, located in the regions. The regional offices (ROs) may administratively support these

regional evaluation specialists, but independence requires that they would only report directly to the head of evaluation.

Interviews indicate that the main barrier to such an arrangement is the structural underinvestment by ILO field offices and projects in monitoring and evaluation (M&E) specialists to support the results-based management (RBM) requirements of projects and programmes, with little evidence of systematic application of the policy guidance to allocate 3 per cent of budgets to RBM/M&E. This is manifested in substantial demands being placed on the REOs and EVAL focal points to support evaluability and monitoring of programmes (which is included in their terms of reference (ToR). The implication of this arrangement is that there is insufficient time to support either monitoring or evaluation functions adequately, and a critical missing link in the 'independence-chain'. It is recognized that not all projects, especially smaller ones, would require dedicated M&E capacity (in terms of staffing), but could instead make use of external consultants for specific inputs.

A second priority regarding independence is to continue exploring ways in which to support the knowledge, skills and understanding of evaluation managers. The evaluation concludes that the current evaluation manager system is also primarily a response to the shortage of M&E specialists (rather than purely about independence), but that it nevertheless has real potential to support learning across the Organization. Fulfilling this potential requires that evaluation managers have a more nuanced understanding of how to achieve independence, and sufficient confidence to critically challenge evaluators where they see gaps in their methods or analysis.

CREDIBILITY

One of the key questions asked of this evaluation was to examine the persistent level of quality ratings for evaluation reports despite the wide range of systems strengthening activities being undertaken by EVAL. The main observations of the evaluation in this regard are that the current portfolio is relatively homogeneous in terms of its size, methods, evaluators and objective. The dominant evaluation approaches in ILO are somewhat conservative, and focused on examining the achievement of results frameworks rather than on examining the underlying theories of change (ToCs).

There is also the case that quality standards themselves are somewhat reductionist, and only measure certain aspects of quality. This gives emphasis to standardization, rather than appropriateness or fit-for-purpose, which are different hallmarks of credibility. Examination of a broader range of evidence leads this evaluation to conclude that the three main aspects of quality that are challenging ILO are:

■ the need to expand the use of participatory methods (such as mechanisms for self-assessment) that both model social dialogue, and engage social partners and government counterparts in examining the impact of normative changes;

■ the flexibility to ensure credibility by commissioning the right evaluation with the right purpose at the right time; and

■ supporting both a culture and practice of evaluation.

The existence of these challenges is driven by the real and perceived demand for set levels of accountability from donors and the Governing Body (GB). Thus, in the absence of a mandate from donors to commission more strategic clustered evaluations and/or sufficient staff with the expertise to make informed professional judgements on commissioning evaluations, this has been achieved through large numbers of small-budget mandatory project evaluations (which cannot afford the time required to be highly participatory).

The implication of continuing this model, however, is that opportunities are being missed to: address constituents' questions more comprehensively; build the ownership of interventions through participation; and transfer evaluation capacity to social partners at national level. Given

the constraints observed during this evaluation, EVAL's work in support of some of the flagship programmes seems to be the most promising entry point to pilot more flexible and participatory evaluation approaches.

The combination of technical areas, normative work, and tripartite working means that ILO has carved out a niche in the international evaluation arena. Evaluators that combine experience of these dimensions with geographic, lingual and evaluative experience are relatively rare. This is why many ILO interviewees spoke during the evaluation about the challenges in finding evaluators, or about having to continuously work with the same individual evaluation consultants. The relatively low rates being offered by ILO, and the easier administrative procedures for contracting individuals rather than firms, narrows the field still further. Although the IEE has not allowed for a more in-depth study of the procurement of evaluation consultants, observations suggest that the process could be streamlined and more transparent.

Whilst it would be useful for EVAL to provide guidance on the minimum viable costs of evaluation in projects, the evaluation finds reasons to be cautious in replacing the 2 per cent allocation in projects with fixed budgets (despite the high demand from management to do so). Given that the evaluation is structurally underfunded and that TC projects account for at least half of evaluation expenditure, any action that would reduce the overall resourcing of evaluation would have significant negative implications. Current RB and regular budget supplementary account (RBSA) allocations are not sufficient to cover the evaluation needs of different organizational entities, including for Decent Work Country Programme (DWCP) evaluations and quality assurance impact evaluations.

UTILITY

The focus of the evaluation strategy (under Outcome 1) has been on the use of evaluations for governance at central level. The EAC plays two important roles: firstly, it helps to distribute learning among senior managers; and secondly, it demonstrates that accountability through evaluation is the responsibility of the whole of the Organization (and not just EVAL). The system for the follow-up of management responses at the decentralized level is far inferior to the EAC arrangement and needs to be more strongly represented in the indicators in a future evaluation strategy.

The evaluation found that strengthening the field-level system for the follow-up of management responses could not only make a significant contribution to use, but would also bolster EVAL's ability to more systematically track the outcomes of evaluation processes to meet the GB's interest in hearing about the impact of evaluation beyond the percentage of recommendations that were addressed.

An underlying assumption of Outcome 2 is that increasing evaluation quality will also lead to an increase in their use both in the field and by technical departments at headquarters (HQ). Whilst the evaluation agrees that this may be partially true, it finds that the primary driver of utility is not quality. A strong factor that emerged from the portfolio analysis was timing (the opportunity to implement recommendations immediately). However, communications and knowledge management were also highlighted as key ingredients in enhancing the use of evaluations outside the context in which they are completed.

The importance of communications is recognized by EVAL and inter alia reflected in the efforts made to produce meta/synthesis studies and other knowledge products, and the development of the *i*-Track platform. At the same time, evidence suggests that the ILO could do more to support EVAL in ensuring that evaluations are shared: with different audiences; with the right people in an appropriate form; through relevant channels; and in a timely manner. The development of a communications strategy and the appointment of a knowledge manager have laid important groundwork for achieving this.

ENABLING ENVIRONMENT

Many of the requests and suggestions encountered during interviews for this evaluation were essentially to address gaps in the wider RBM system (especially regarding evaluable indicators or incorporation of evaluation recommendations). Whilst these are relevant and necessary programming needs, they are not typically the mandate of a central evaluation function.

Thus, it is important to recognize that the boundaries of the evaluation function are strongly influenced by the environment. The lack of M&E capacity within projects and programmes limits the results data available to evaluations, and thus the range of methods that can be used or the sophistication of the questions that can be addressed.

Similarly, the weak organizational incentives for ILO staff to engage in evaluations – as witnessed by evaluation management – limits the capacity of the function. Despite the independence of EVAL, the evaluation function in ILO does not exist in isolation: it is embedded within a cultural framework and a set of institutional drivers that will also need to adjust if the performance of the function is to improve.

The evaluation policy (2005) was an immediate response to the UNEG norms and standards that were agreed in that year, and provided a strong framework for the institutionalization of evaluation in ILO. Similarly, the evaluation strategy (2011) was an immediate response to the IEE (2010). However, the results framework of the current strategy has become insufficient as a management tool and requires a fundamental revision. This evaluation sees three other main challenges facing the design of the next evaluation strategy:

- determining a means to strengthen the focus of the evaluation activity on utility and national capacity in what is a structurally underfunded function in a way that does not jeopardize the gains already made in systematizing ILO's independent evaluation responsibility;

- expanding the use of strategic, joint, thematic and DWCP evaluations in a way that does not leave TC projects with insufficient evaluative coverage; and

- progressively integrating the large body of internal and self-evaluations into ILO's quality assurance and system-strengthening process, and dealing with the implications of this on both human resources and the probable near-term drop in average report quality that will result from this integration.

Missing from the current strategy is a clear ToC for the evaluation function, and indicators of the outcome of the evaluation itself (beyond the proportion of recommendations adopted). Also missing from the current policy is a clear mission for the evaluation function, and guided flexibility to maximize the value of each evaluation (for example, by avoiding duplication with external evaluations).

The evidence from this evaluation indicates that the one of the main contributing factors to this situation was the top-down development of the current strategy, based primarily on the recommendations (and framework) of the IEE (2010). The ToR for this evaluation can be interpreted to imply that a similar process is intended for the next evaluation strategy. However, whilst this evaluation should serve as a useful input to that process (along with the UNEG norms and standards, 2016), it concludes that in order to be owned and understood, the next strategy (and policy) should be developed through a participatory process facilitated by EVAL.

CONCLUSIONS

Overall conclusion: In the course of the evaluation strategy (2011–2017), the evaluation function in ILO has been transformed in terms of its structural independence, the institutionalization of evaluation practice, and the development of material. It is highly regarded for having achieved this with

limited resources, including becoming a more consistent and visible champion for gender mainstreaming, and introducing good practices that are in some ways ahead of its UN evaluation peers.

- Conclusion 1: Priorities for enhancing the independence of the evaluation function are the integration of regional evaluation officers (REOs) as full staff members of EVAL, strengthening the capacity of evaluation managers and expanding quality assurance systems to internal evaluations.

- Conclusion 2: Whilst independent evaluation reports largely meet UNEG standards, there is much to be gained from increasing the diversity of evaluations, mainstreaming social dialogue, and deepening constituents' participation.

- Conclusion 3: Enhancing evaluation budgets, rosters, networks and procurement processes is necessary to consistently secure evaluators with the combination of evaluation skills and technical knowledge that ILO requires.

- Conclusion 4: There is significant value to be realized from strengthening communications and knowledge management to enhance the utility and use of evaluation in the field.

- Conclusion 5: At the decentralized level, there is a need to replicate the same success that the EAC has had in ensuring management responds effectively to recommendations from high-level evaluations (HLEs).

- Conclusion 6: Despite high-level support for evaluation in the ILO, the emergence of an evaluation culture is inhibited by underinvestment in M&E specialists, and too few institutional incentives for programme and technical staff to engage with evaluation.

- Conclusion 7: The evaluation policy (2005) and the evaluation strategy (2011) have served their purpose well, but both now need to be updated to meet the challenges of a changing context.

RECOMMENDATIONS

The following 13 recommendations have been developed by the evaluation team based on the conclusions, findings, and consultations with the technical committee for the IEE.

- Recommendation 1: Recognizing ILO's comparative leadership in evaluation in the UN system, the necessity for evaluative thinking to meet the 2030 Agenda, and the need to consolidate an emerging evaluation culture, it is recommended that the GB and ILO's senior management strongly reassert organizational commitment to the evaluation function. Given the current trajectory and the recommended level of budgetary and political support, it would be possible to set an ambitious target of ILO beginning the transition to the highest level of the JIU maturity matrix.

- Recommendation 2: Transition REOs into full staff members of EVAL.

- Recommendation 3: Incentivize and strengthen the evaluation manager and focal person system.

- Recommendation 4: Establish an integrated evaluation planning system.

- Recommendation 5: Further develop collaboration with other agencies of the UN system to advocate for and support a diverse community of evaluators and national constituents with expertise in evaluating decent work and promoting social dialogue.

- Recommendation 6: Enhance evaluation value added and relevance by promoting participatory, gender-responsive and mixed-methods evaluation.

- Recommendation 7: Expand the quality assurance system to include internal evaluations, and switch to an annual or real-time independent quality assurance system.

- Recommendation 8: Diversify and elevate the overall portfolio of evaluations to include more DWCP evaluations and thematic evaluations.

- Recommendation 9: Strengthen the decentralized evaluation management response mechanism.
- Recommendation 10: Prioritize EVAL's communications capacity and coaching function.
- Recommendation 11: Strengthen RBM and M&E systems to promote DWCP, programme and project evaluability.
- Recommendation 12: Update and align the evaluation policy to IEE recommendations and current organizational structure and processes.
- Recommendation 13: Develop the new evaluation strategy in a participatory manner to promote ownership and visibility.

1 INTRODUCTION

This report elaborates the results of the inception, field and reporting phases undertaken by the evaluation team, during the period October–December 2016. The main activities are shown in table 1 below.

Table 1. Inception, field and reporting phases undertake by the evaluation team

PHASES	INCEPTION OCTOBER		FIELD NOVEMBER		REPORTING DECEMBER–JANUARY
Tasks	Structuring of the evaluation and definition of the methodology Mapping of stakeholders Data collection and inventory Analysis of strategic documents (desk study) Evaluability assessment Theory of change (ToC) Evaluation questions and indicators Selection of regional offices (ROs) for field visits		Field missions (observations, interviews, etc.) Collection of further documents Verification of hypotheses Presentation of preliminary findings to the ROs Three field missions: – RO Africa – RO Asia – RO Europe		Drafting the Draft Final Report (DFR) with complete answers to the evaluation questions (EQs) and conclusions and recommendations Circulation and comments from key stakeholders Finalization of the Final Report after discussion of the DFR with the Evaluation Advisory Committee (EAC) and stakeholders Preparation of the full set of annexes
Meetings	Preliminary meeting in Geneva		Debriefing with regional evaluation officers (REOs)		Final validation workshop
Deliverables	Inception report		Short summary of conclusions of interviews		DFR slide presentation for the final workshop Final Report

The report is structured as follows:

- Section 2 presents the background of the evaluation
- Section 3 presents the overall purpose of the evaluation, its objectives and design constraints, its context and guiding questions
- Section 4 focuses on the evaluation object
- Section 5 discusses the theoretical basis for the evaluation
- Section 6 describes the evaluation design and methods
- Section 7 presents the findings
- Section 8 presents the conclusions
- Section 9 presents the recommendations
- Section 10 presents lessons for evaluation practices.

2 BACKGROUND

2.1 ILO EVALUATION POLICY AND STRATEGY

The ILO's evaluation function has increased its significance and coverage since the first actions of the Governing Body (GB) in 2000[1] and 2002[2] that laid down the foundations of *A new policy and strategic framework for evaluation at the ILO* published in 2005 (table 2).

The 2002 document defined the objectives of evaluation in the ILO, set the basic principles, methods and strategy of the evaluation function within a results-based management (RBM) context and outlined the core capacities needed to implement the proposed evaluation framework over a period of three years. Several aspects of the evaluation strategy proposed in 2002 have been implemented during that period. Among these, as recognized by the 2005 policy,[3] are: the annual reports to the GB on the outcome of regular budget (RB) programme and technical cooperation (TC) project evaluations; the consultation with the GB constituent groups in the selection of evaluation topics and terms of reference (ToR); the allocation of RB resources for training activities on monitoring and evaluation (M&E).

Table 2. Timeline of ILO evaluation strategy evolution

ILO evaluation strategy milestones	Year	Contextual factors
■ ILO evaluation strategy, Nov. 2000, GB.279/PFA/8,	**2000**	
■ ILO evaluation framework: Evaluation within a strategic budgeting context, Nov. 2002, GB.285/PFA/10,	**2002**	■ 2002–05 ILO Strategic Policy Framework
■ A new policy and strategic framework for evaluation at the ILO, GB.294-PFA-8-4 ■ Creation of the central Evaluation Unit	**2005**	■ United Nations Evaluation Group (UNEG) Norms for Evaluation
	2006	■ 2006–09 ILO Strategic Policy Framework
Evaluation of the evaluation function (IEEF)	**2010**	■ 2010–15 ILO Strategic Policy Framework
■ Results-based strategies 2011–15: Evaluation strategy – Strengthening the use of evaluations	**2011**	
Evaluation of the evaluation function (IEE)	**2016**	■ 2016–17 ILO Transitional strategic plan ■ UNEG Norms for Evaluation updated
■ New evaluation strategy	**2018**	■ *2018–2021 new ILO Strategic Plan (SP)*

---

[1] ILO evaluation strategy, Nov. 2000, GB.279/PFA/8.

[2] ILO evaluation framework: Evaluation within a strategic budgeting context, Nov. 2002, GB.285/PFA/10.

[3] Evaluation: (d) A new policy and strategic framework for evaluation at the ILO, Nov. 2005, GB.294/PFA/8/4.

The 2005 evaluation policy framework was inspired by internationally accepted norms and standards that aimed to improve and strengthen the practice of independent evaluation in the ILO. The policy also established principles for systematic self-evaluation of programme performance in order to cover all ILO activities supporting the ILO's vision articulated in the 2006–09 Strategic Policy Framework (SPF).

The 2005 policy has provided a <u>vision</u> for the evaluation function, which is largely recognized: "A sustained and expanding institutional culture of accountability, transparency and quality improvement is a strong vision that both the ILO Governing Body and the Office share. Evaluation for better performance and effectiveness in the pursuit of the Decent Work Agenda (DWA) is at the core of this commitment."[4]

Its key <u>objectives</u> and <u>principles</u> are as follows:

- Objectives: (i) improve Office-wide transparency and accountability for impact of ILO actions to support its constituents; (ii) strengthen the decision-making process; (iii) contribute feedback for learning and ongoing improvement of ILO's work.

- Principles: (i) adherence to international good practice; (ii) upholding the ILO mandate and mission; (iii) ensuring professionalism; (iv) transparency and learning; (v) independence of process.

The policy also includes an overall strategy and operational approach in order to guarantee the independence of the evaluation function within ILO through adequate management and organization. One innovative element of the strategy was the establishment of a central Evaluation Unit (EVAL) responsible for consolidating and coordinating the efforts towards a strengthened evaluation framework and capacity throughout the entire Organization. EVAL works in consultation with the Strategic Programming and Management Department (PROGRAM) and the Department of Partnerships and Development Cooperation (PARDEV), and also with the regional management and technical departments. Moreover, the EAC was established as a mechanism to oversee the use, implementation and follow-up of lessons learned and recommendations resulting from ILO evaluation activities, and provide advice to the Director-General (DG) on the adequacy of progress made.

As per point 46 of the 2005 evaluation policy, in 2010 the first independent evaluation of the evaluation function (IEE) was carried out. The IEE provided a series of conclusions, lessons learned and recommendations. The main conclusions were:

- evaluations that are produced by the ILO are of generally high quality, measured by UNEG standards, but they have not been used as much as implied by the 2005 policy;

- the resources available from all sources for evaluation in the ILO are comparable to other organizations of the United Nations (UN) system, although they are comparatively lower from RB;

- between 2005 and 2009, evaluations did not play a significant role in shaping policies and strategies within a RBM context;

- the degree of independence of evaluations has not been reflected in organizational location, one of the standards used in the UN system; and

- the EAC has taken a narrow view of its mandate relating to follow-up, and does not see it in terms of RBM.

In line with IEE recommendations (2010), the evaluation function was substantively reformed through the 2011–2015 evaluation strategy. Some major developments have taken place, including:

- the transformation of the central Evaluation Unit into an independent Evaluation Office (both have the acronym "EVAL"), reporting to the GB on evaluations;

---

4   Results-based strategies 2011–15: Evaluation strategy – Strengthening the use of evaluations, 2011, GB.310/PFA/4/1(Rev.).

- the development of codified evaluation policy guidelines (issued in 2012);
- the development and implementation of an evaluation communications strategy, including the preparation of a wide range of newsletters, think pieces, meta-studies, synthesis reviews, etc.;
- the introduction of a publicly accessible database to monitor and track the evaluations, and store evaluation findings and full reports;
- the independence of the evaluation function and architecture has been enhanced through the policy guidelines and protocols for HLE and DWCP, identification of REO/DEFP and evaluation managers ensuring the independence of the evaluations;
- the available resources for evaluation activities have increased, including the budget channelled through the fixed share of programme support and extra budgetary sources;
- systematic training of evaluation managers through a certified course – Evaluation Manager Certification Programme (EMCP); and
- the introduction of a peer review mechanism to enhance quality of the reports.

The list of specific activities carried out since the first IEE is provided in Annex F while the performance assessment of the activities is provided in Annex G.

The evaluation strategy (2011–15) is organized along three main outcomes, which are aligned with those set out in the 2005 policy and strategy. However, they placed greater emphasis on areas that needed more focused attention as recommended by the previous IEE. The refined outcomes are as follows:

- Outcome 1: Improved use of evaluation by ILO constituents and management for governance;
- Outcome 2: Harmonized office-wide evaluation practice to support transparency and accountability; and
- Outcome 3: Evaluation capability expanded through enhanced knowledge, skills and tools.

The strategy also defines indicators, baselines and targets linked to temporal milestones for 2010–11, 2012–13, and 2014–15, respectively. With the extension of the strategy in 2015, additional milestones for 2016–17 were added.

The present evaluation is, therefore, part of the whole evaluation strategy, directly deriving from one recommendation in the 2014–15 Annual Evaluation Report (AER), and also from one of the recommendations of the previous IEE. The present evaluation strategy is in line with the 2010–2015 ILO SPF and its transitional period of 2016–2017. Therefore, the present evaluation will inform the implementation of the new Strategic Plan that will be in line with the new planning cycle, 2018–2021, as foreseen by the UN General Assembly.[5]

2.2 EVALUATION WITHIN THE ILO STRUCTURE AND SYSTEMS

ILO's evaluation function was part of an analysis carried out by the Joint Inspection Unit (JIU) in 2014 of the evaluation function in the United Nations system. According to the study, the overall level of maturity of ILO's central evaluation function is placed at a high level transitioning to level 4 and thus appearing in cluster I, which contains nine organizations where most of the expected systems and policies supporting the central evaluation function are in place and well-institutionalized. This means that ILO's evaluation function operates as a completely separate office, which guarantees independence, and has an adequate amount of allocated resources. In terms of performance,

---

[5] "The UN General Assembly has mandated funds and programmes, and encouraged specialized agencies to adopt the quadrennial comprehensive policy review (QCPR) planning and reporting cycle. Eleven out of 19 UN entities have aligned their strategic plans with the QCPR. Currently this cycle is for 2014–17, followed by 2018–21. Given that the current Strategic Policy Framework ends in 2015, the ILO would be in a position to align itself with the four-year cycle in 2018", GB.294/PFA/8/4.

ILO has integrated and institutionalized the components[6] of the evaluation function and the focus for improvement has shifted to enhancing the value added effect for the Organization and enlarging partnerships as alternative ways of enhancing effectiveness, efficiency, sustainability and impact.[7]

According to the study, there are several actions that all assessed UN agencies (including ILO) should undertake to improve evaluation quality and to support better use of evaluation within the UN system as a tool for learning and evidence-based decision-making procedures.

In terms of structure, the independence of the evaluation function within the ILO is guaranteed by the Evaluation Office (EVAL), which is the central body responsible for independent evaluations of ILO's strategies, policies and programmes. EVAL is supported by regional evaluation officers (REOs) in five regional offices and a network of department evaluation focal points. In addition, ILO has an EAC, which oversees the use and implementation of evaluation recommendations and lessons learned. More details are provided in section 4.4.

In terms of activities, the evaluation function is organized around four inter-related main groups:

- <u>production and management of evaluations:</u> strategy and policy evaluations, Decent Work Country Programme (DWCP) evaluations, thematic evaluations, DWCP internal review, impact evaluations, joint evaluations, project evaluations, organizational reviews[8] (see below an overview of the evaluations carried out during the last six years – 2011–2016);

- <u>institutional accountability:</u> communication with the GB through reports such as high-level/ strategic evaluations, DWCP evaluations, high-level evaluations (HLEs), synthesis reviews, and annual reports;

- <u>communications and knowledge management:</u> production of newsletters, guidelines, think pieces, meta-studies, synthesis reviews, *i*-Track database constantly monitored and updated with new evaluation products and recommendations, keeping track of all outstanding evaluations, management of the evaluation consultant CV database, information service management, website maintenance, and other information-sharing activities (public and intranet site); and

- <u>capacity development:</u> policy guidelines, guidance documents, training modules organized with the support of the International Training Centre of the ILO.

The above list of activities has a pure descriptive purpose and is not exhaustive. The activities have been clustered around the three outcomes and the performance criteria (see Annex F).

Overview of evaluations carried out 2011–2016

During the evaluation period 2011–2016, ILO carried out a total of 616 evaluations making an average of 123 per year (table 3). The independent project evaluations accounted for 44 per cent of the total number of evaluations followed by the internal evaluations reports and the joint project evaluations, which accounted for 26 per cent and 17 per cent, respectively. In terms of geographical coverage, the highest number of evaluation reports was carried out in Africa and Asia, with a total of 178 and 153 reports, respectively. In terms of temporal distribution, the trend was positive between 2011 and 2013 when the highest number registered reached a peak of 145 reports. Subsequently, the activity gradually decreased and, by 2016, the total number of evaluations amounted to 37 (figure 1). This trend is linked to revisions in the thresholds for independent evaluations (previously US$500,000, currently $1 million).

---

[6]   (i) Enabling environment – institutional framework and support system; (ii) independence with inclusion for enhanced credibility; (iii) quality – technical and managerial rigour for enhanced credibility; (iv) utility – use of evidence; (v) relevance and readiness to support the organization to address emerging changes and challenges.

[7]   Joint Inspection Unit: *Analysis of the evaluation function in the United Nations System* (Geneva, United Nations, 2014).

[8]   ILO: *Results-based strategies 2011–15: Evaluation strategy – Strengthening the use of evaluations*, Governing Body, 310th Session, Geneva, Mar. 2011, GB.310/PFA/4/1(Rev.).

At the same time, it is recognized that a significant number of evaluations were being completed in the last quarter of 2016 and, consequently, the total number of evaluations conducted in that year may be significantly higher.

Table 3. ILO's evaluation reports by type of document and region, 2011–2016

Document type[9]	Africa	Asia	Europe	Latin America		Middle East	Inter-regional	Grand total
DWCP evaluation	1	1		2		1	–	**5**
DWCP internal review	14	5	7	3		1	–	**30**
Organizational review	–	–	–	–		–	2	**2**
Project / external evaluation report	5	4	–	2		–	5	**16**
Project / independent evaluation	89	62	17	38		11	54	**271**
Project / internal evaluation report	38	57	10	26		6	21	**158**
Project / joint evaluation report	24	23	9	45		4	1	**106**
RBSA evaluation	7	–	3	5		–	–	**15**
Strategy / policy evaluation report	–	–	–	–		–	9	**9**
Thematic evaluation report	–	1	–	–		–	3	**4**
Grand total	**178**	**153**	**46**	**121**		**23**	**95**	**616**

Figure 1. ILO's evaluation reports by region and year, 2011–2016
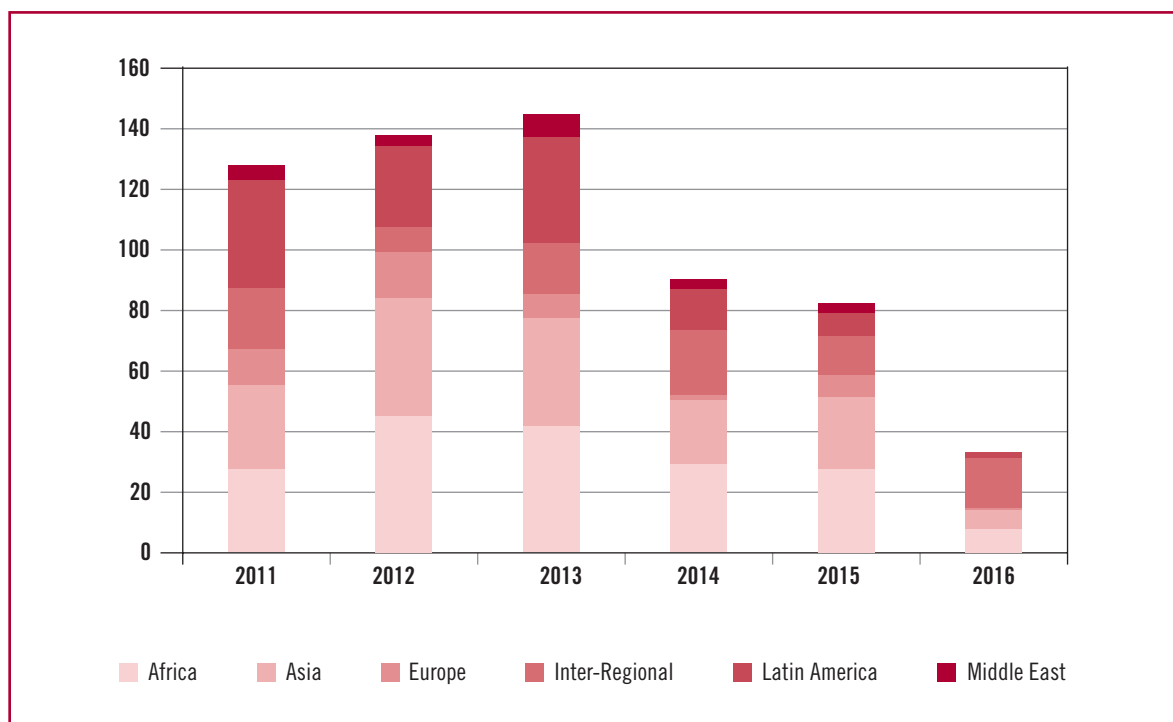
9   Categorization of evaluations used in ILO's database.

It is worth mentioning that during the evaluation period, three quality assessments (QAs) of the independent evaluation projects were carried out covering a total of 106 evaluations. In particular, the QA for 2013, covering the period 2009–2011, included 22 evaluations; the QA for 2014, covering the period 2012–2013, included 42 evaluations; and the QA for 2015, covering 2014 and part of 2015, included 42 evaluations. The results of the assessment revealed an overall average quality of 1.8 (where 1 is the threshold for insufficient quality and 2.5 is the threshold for high quality). However, according to the 2014 appraisal, the overall quality of ILO evaluation reports has increased over the years even though there are still key areas for improvement mostly related to evaluation credibility, and the methodology applied and its use.[10]

---

[10] K. Robertson and D. Schröter: "Leveraging appraisal findings to improve evaluation quality", in *i-eval Think Piece* (2014, No. 4).

3 EVALUATION PURPOSE, OBJECTIVE AND SCOPE

3.1 PURPOSE AND OBJECTIVES

This evaluation was commissioned at this time due to the forthcoming completion of the ILO evaluation strategy (extended) 2011–2017. It, thus, has a dual purpose: to enable retrospective accountability for the delivery of the evaluation strategy (including, but not limited to EVAL); and to support potential lessons and recommendations for a new evaluation strategy. The evaluation will thus aim to provide an assessment of the overall performance of ILO's evaluation function during the period 2011–15 (including the transitional extension covering 2016–17) along with explanations for this performance, and lessons learned and good practices.

More specifically, the evaluation focused on achieving two main interconnected objectives:

1. assessing the relevance of the strategy, relating to the *quality of the strategy and its actual capacity to inform sound decision-making in the ILO*; and

2. assessing the effectiveness of the operational arrangements and structures relating to the *appropriateness of the organizational set-up and processes*, specifically focusing on the enhancement of the ILO's evaluation capacity.

The evaluation assessed the independence, credibility, and utility of the evaluation function, as well as the overall enabling environment. This included the *structure* and the different and complementary roles of central and decentralized evaluation arrangements within the ILO, the sharing of knowledge and the harmonization of evaluation tools and practices.

On the whole, this evaluation is meant to respond to the three major priorities of the evaluation strategy, i.e. *learning, performance strengthening and accountability*. The results of this evaluation (findings, conclusions, lessons learned and recommendations) will inform a new evaluation strategy, aiming to provide insights to improve the operational and decision-making frameworks and contribute to the alignment of the evaluation strategy with the ILO Strategic Plan for 2018–2021.

3.2 SCOPE OF THE EVALUATION

The overall boundaries of the evaluation are outlined by the ToR and referred to as the *evaluation function* of the ILO. As confirmed through the consultations and desk reviews conducted during the inception phase, it is framed by the 2005 evaluation policy and the evaluation strategy for 2011–2015 (extended to 2017). The scope of analysis is developed around four main aspects of the evaluation function, that is independence, credibility, utility, and enabling environment:

- independence: assess the **adherence** of ILO's systems, structures and practices to ILO evaluation policy, UNEG norms and standards; relevance and effectiveness of organizational relationships, reporting arrangements and of the evaluation portfolio;

- credibility: assess the **accountability** of the evaluation function and whether it meets the needs of the constituents; determine whether ILO evaluation respects UNEG norms and standards and has the required capacities and competencies to do so;

- utility: assess the **usefulness** of the evaluation products (reports, reviews, newsletters, training materials, etc.); the actual **contribution** of the evaluation function and its results in informing the ILO's strategic directions, policies and programmes, and whether they have been incorporated into and used in follow-up activities; and

- enabling environment: assess **effectiveness and organizational efficiency** of the evaluation function.

For the purposes of this evaluation, the evaluation function is made up of two main elements:

- **central evaluation function:** EVAL, the EAC, and the systems, structures and evaluations under the control of EVAL that oversee the activities of the decentralized evaluation function and represent the final stage for the approval of all ILO evaluations; and

- **decentralized evaluation function:** evaluations under the control of programme, technical and field units; and the REOs, departmental evaluation focal points or persons (DEFPs), and certified evaluation managers supporting decentralized evaluations.

Moreover, this evaluation provides actionable recommendations to improve the overall vision of the evaluation function and strengthen the culture of evaluation within the ILO in order to ensure that all entities recognize that it is a tool to improve the performance of all activities.

3.3 EVALUATION STAKEHOLDERS

As part of the preliminary analysis, a comprehensive stakeholder assessment was undertaken based on the information provided by EVAL (table 4). A complete version is included in Annex H. To further enhance the usability of the evaluation, a more focused group of primary intended users and uses has been developed. This is based on both the ToR and group discussions conducted with EVAL during the inception phase. The evaluation process was designed to maximize utilization by these primary intended users, whilst adhering to human rights principles of participation, non-discrimination and accountability.

Table 4. Primary and secondary intended users and uses of the evaluation

Primary intended users	Intended uses
GB constituents	- Inform strategic decisions related to the new ILO strategic plan 2018–2021 - Have a greater awareness and understanding of ILO's approach to evaluation so as to better engage in these
Management (DG's Office and Deputy Directors-General – DDGs)	- Inform strategic decisions related to the new ILO strategic plan 2018–2021
EVAL	- Inform strategic decisions related to the new evaluation strategy 2018–2021 - Support organizational learning and accountability on developing evaluation
Country offices, ROs	Use lessons and recommendations to enhance existing and future project implementation
Technical departments	Use lessons and recommendations to enhance existing and future coordination and supervision of project implementation, and to design new projects and strategies

Primary intended users	Intended uses
Evaluation architecture within ILO (evaluation focal persons, evaluation managers, and REOs)	■ Have a greater awareness and understanding of ILO's approach to evaluations so as to better engage with them ■ Use lessons and recommendations to strengthen and improve their role within the evaluation function
Secondary intended users	
PROGRAM and PARDEV	Use of evaluation findings as inputs to ILO's RBM framework – design, resource allocation, etc., both for programmes using RB resources (PROGRAM) and for projects (using extra-budgetary resources)
Other entities at HQ	Have a greater awareness and understanding of ILO's approach to evaluations so as to better engage with them
Constituents at the country level	Have a greater awareness and understanding of ILO's approach to evaluations so as to better engage with them

4 EVALUATION OBJECT

4.1 OBJECTIVE OF THE EVALUATION

From an organizational perspective the evaluation function is primarily led by EVAL, which is mandated to implement the evaluation policy and strategy. The main functional responsibilities of EVAL are to commission, manage and participate in high-level evaluations (HLEs), elaborate policies and guidelines, provide quality assurance, disseminate evaluation reports, and monitor and report on the implementation of evaluation recommendations. EVAL is supported by REOs in five regional offices and a network of DEFPs. In addition, ILO has an EAC, which oversees the use and implementation of evaluation recommendations and lessons learned.

The IEE assesses the above-mentioned evaluation structures and their role in the implementation of the evaluation strategy. Other ILO entities, including technical departments and ROs, are entrusted with managing decentralized evaluations and have been studied from the perspective of the quality and use of such evaluations. At the same time, the focus is on EVAL's oversight and quality assurance function.

The inception phase of the evaluation identified three main products of the evaluation function:

- governance level (policy and strategy) evaluations (centralized) commissioned by the GB;
- project evaluations (decentralized) triggered by the size of the projects' budgets and other criteria; and
- meta-studies, synthesis reviews, and thematic evaluations (and most impact evaluations), which are optional.

Independent project evaluations are carried out of centralized global and interregional projects, managed by HQ, and decentralized projects, managed by regional and field offices. The management of independent evaluations is carried out by an assigned evaluation manager under the supervision of the DEFPs for centralized projects and by REOs for decentralized projects, both overseen by EVAL, with final approval by EVAL.

Decentralized evaluations include: (1) mandatory independent project-level evaluations managed by evaluation managers with the coordination and oversight of REOs/DEFPs and an EVAL senior evaluation officer; (2) internal evaluations directly managed by offices and projects; (3) impact evaluations under the responsibility of the technical departments/field offices; and (4) other evaluative activities, such as synthesis reviews managed by departments/field offices.

The financial resources for the *evaluation function* come from multiple sources. RB covers a majority of the core evaluation positions in EVAL and the cost of HLEs. Other types of evaluations are funded from programme, TC and project budgets. REOs are financed from a mix of funds.

4.2 PERFORMANCE FRAMEWORK

The ILO evaluation policy (2005) calls for a "sustained and expanding institutional culture of accountability, transparency and quality improvement". Accordingly, the evaluation function is "expected to make an essential contribution to policy-making and decision-making within the results-based budgeting system in the Office, to optimize the allocation of resources and improve their overall management." It is, therefore, expected to contribute to the more effective delivery of the ILO Strategic Plan, and the strategy sets nine priorities for improving its effectiveness:

    i. more systematic use of self-evaluation and independent evaluation;

    ii. regular reporting to senior management and the GB on evaluation activities and their effects;

    iii. follow-up to evaluation findings and recommendations, including their use in the results-based planning, programming and budgeting process;

    iv. improved institutional learning and knowledge sharing;

    v. harmonization of evaluation practices and methods within the Office, regardless of source of funds;

    vi. decentralized evaluation responsibilities and accountabilities, as appropriate;

    vii. improved internal capacity and skills in evaluation and self-evaluation methodologies;

    viii. participatory process of ILO constituents in evaluation; and

    ix. independence of the evaluation function preserved.

Responding to the evaluation policy (2005), the evaluation strategy (2011–2015) is accompanied by a performance framework defining outcomes, indicators, baselines, end targets, and biennial milestones (table 5). In connection with the extension of the evaluation strategy, biennial milestones for 2016–17 were added. Table 5 summarizes the indicators, baselines and end targets for each of the three outcomes as presented in the evaluation strategy (2011–2015). The performance framework presented some weakness. Indeed, the end targets do not aggregate the milestone results; additional indicators were added (four additional indicators under Outcome 2) in reporting to capture important developments related to the outcomes.

Table 5. Evaluation strategy 2011–2017 performance framework

Outcomes	Indicators	Baselines	End targets
Outcome 1: Improved use of evaluation by ILO constituents and management for governance	1.1 The frequency and quality of EAC decisions and advice on evaluation programming, and the adequacy of follow-up to evaluation results	Three meetings in 2010; topics discussed for coming year only; no discussion on strategy; use of evaluation recommendations	EAC convenes meetings and forums where analysis and dialogue on evaluation topics and follow-up lead to documented plans and follow-up for strategic use
	1.2 AER synthesizes recommendations and lessons learned based on evaluations	Reporting on implementation of evaluation strategy without analysis of broader ILO effectiveness	Annual evaluation reporting based on analysis of evaluation reports
	1.3 HLEs assess the contribution of technical and country strategies to strategic policy framework/strategic plan (SPF/SP), programme and budget (P&B) outcomes	External quality rating of evaluations; 2005–09	HLEs better inform governance-level strategic and programming decisions
Outcome 2: Harmonized office-wide evaluation practice to support transparency and accountability	2.1 By 2015, 100 per cent of DWCPs and projects have mechanisms in place for regularly engaging constituents in the use of evaluation processes	–	Periodic ex-post surveys and reporting of management response and follow-up shows that 100 per cent of evaluations address constituent involvement

Outcomes	Indicators	Baselines	End targets
	2.2 Upgrade and expand the use of evaluations for management (decentralized)	Count of self, internal, thematic and impact evaluations conducted by sectors and regions	All regions and sectors have biennial evaluation plans linked to management accountability and organizational learning, which are reviewed by EAC
Additional indicator A2.2(1):	*Codification and upgrading of procedures and guidelines*	*i*-Track database adjusted "EVAL Plone" site is being upgraded Quarterly newsletter launched in 2011	Third edition of the policy guidelines for results-based evaluation made available to ILO staff
Additional indicator A2.2(2):	*Maintenance of a strong network of departmental and regional evaluation network*	An informal network of more than 90 ILO officials working in evaluation receives regular updates	Biennial evaluation network workshop, quarterly conference calls, co-hosted with UNEG "Evaluation Practice Exchange"
Additional indicator A2.2(3a):	*Use of a rigorous quality control system*	Developed draft job descriptions for P.2–P.5 evaluation officers	EVAL and Office of Legal Services (JUR) review and provide feedback on proposed evaluation arrangements at the project drafting stage, biennial quality appraisal of independent evaluations
Additional indicator A2.2(3b):	*Follow-up to evaluation recommendations*	By mid-2011, implementation of 72 per cent of 2010 evaluations' recommendations	Altogether, 89 per cent of recommendations were completed or partly completed
Outcome 3: Evaluation capability expanded through enhanced knowledge, skills and tools	3.1 Evaluation capacity and practice among ILO staff and constituents improved	Number of staff and constituents receiving technical training and hands-on support	All interested constituents can avail themselves of training in specialized evaluation skills
	3.2 Standardized roles and responsibilities applied to evaluation officers and focal points	No standard job descriptions for evaluation officers; compliance with evaluation guidelines unknown	Evaluation responsibilities standardized and specified in job descriptions for focal points EVAL participation in performance appraisals for all evaluation officers and focal points

The performance framework has informed the Evaluation Matrix (see Annex I, Vol. II) and has been used, together with the AERs, as a reference source for the analysis of the evaluation function's effectiveness. The baseline and progress data have been verified against source documentation and explanations as to why examined results have occurred (or not occurred). Once verified, the progress data have been aggregated to establish the extent of achievement of end targets and outcomes. In addition, the quality of the performance framework has been discussed and analysed.

4.3 SUMMARY OF KEY INTERNATIONAL COMMITMENTS

The IEE examined the extent to which ILO has made progress in implementing the UNEG norms and standards and the extent to which the Organisation for Economic Co-operation and Development/Development Assistance Committee (OECD/DAC) and the UN System Wide Action Plan on Gender Equality and the Empowerment of Women (UN SWAP) norms and standards have been taken into consideration. The UNEG norms and standards were first issued in 2005 and have recently (June 2016) been updated (table 6). The IEE used the 2005 norms and standards as a benchmark for evaluating ILO's past performance, and the 2016 norms and standards as a basis for future recommendations.

Table 6 Scope of the 2016 UNEG norms and standards

Norms		Standards	
General norms	**Institutional framework**	**Evaluation competencies**	**Quality**
■ International principles, goals and targets	■ Institutional framework for evaluation	■ Competencies	■ Quality assurance system
■ Utility	■ Evaluation policy	■ Ethics	■ Quality control of the evaluation design
■ Credibility	■ Evaluation plan and reporting		■ Quality control at the final stage of evaluation
■ Independence	■ Management response and follow- up	**Conduct of evaluations**	
■ Impartiality	■ Disclosure policy	■ Timeliness and intentionality	
■ Ethics		■ Evaluability assessment	
■ Transparency		■ ToR	
■ Human rights and gender equality	**Management of the evaluation function**	■ Evaluation scope and objectives	
■ National evaluation capacities		■ Methodology	
■ Professionalism	■ Head of evaluation	■ Stakeholder engagement and reference groups	
	■ Evaluation guidelines	■ Human rights-based approach and gender mainstreaming strategy	
Institutional norms	■ Responsiveness of the evaluation function	■ Selection and composition of evaluation teams	
■ Enabling environment		■ Evaluation report and products	
■ Evaluation policy		■ Recommendations	
■ Responsibility for the evaluation function		■ Communication and disseminations	
■ Evaluation use and follow-up			

OECD/DAC has adopted a set of core principles for the evaluation of development assistance focusing on the management and institutional arrangements of the evaluation system of development agencies. These principles are complemented by DAC evaluation criteria, quality standards and guidance documents. The quality standards were adopted in 2010 and include overarching considerations (transparency, ethics, participation, capacity development, quality control) as well as detailed considerations with regard to evaluation purpose, planning and design, implementation and reporting, and follow-up use and learning.

UN-SWAP constitutes the first accountability framework for gender mainstreaming in the UN system and is composed of 15 performance indicators for tracking six elements of gender mainstreaming. The oversight element of UN-SWAP is composed of three performance indicators, including one dedicated to evaluation that is linked to meeting the gender-related UNEG norms and standards, and demonstrating effective use of UNEG guidance on integrating gender equality in evaluation.

There are several other international commitments that may influence ILO's future evaluation strategy. Most importantly, the adoption of the Sustainable Development Goals (SDGs) in 2015 is likely to further strengthen the demand for development effectiveness, policy coherence, national ownership and impact orientation. This will have implications for how evaluation functions are structured and operationalized, and open up new opportunities for harmonization, joint evaluations and faster feedback loops. ILO fully recognizes that evaluation needs to be part of the overall ILO response to SDGs, and its commitment is clearly confirmed in the following note of June 2016:

> From an evaluative perspective, the nature of the required action will be both short and longer term in nature, ultimately focusing on achievements and impacts. In the coming years evaluative elements will help to ensure that the most cost-effective approach to SDG/DW implementation is being followed and that at the we will be able to provide a compelling 'Performance Story' regarding our contribution to the SDGs.[11] (EVAL, 2016)

---

[11] ILO: *Evaluation and the Sustainable Development Goals (SDGs) The implications of the SDGs on ILO's results framework – remarks from an evaluative perspective*, Evaluation Office (Geneva, June 2016).

4.4 ORGANIZATIONAL CONTEXT

The ILO's evaluation function is organized through interlinked levels and units.

The **Evaluation Office (EVAL)** plays a central role in ILO's evaluation function (figure 2). It comprises: the Director, three Senior Evaluation Officers, a Communications and Knowledge Management Officer and a Secretary/Administrative Assistant.

The Evaluation Office has its origins in the central Evaluation Unit (also known as EVAL), which was established in 2005 as part of ILO's management and administration structure. In 2011, based on the recommendations of the 2010 IEE, EVAL was placed directly under the DG's Office in ILO's organizational structure. In 2014, as part of the DG's reform agenda, it was renamed ILO's Evaluation Office. EVAL is responsible for monitoring and reporting to the DG and the GB on the implementation of HLE recommendations and related management responses. EVAL oversees independent evaluations undertaken at the decentralized level and directly manages HLEs.

Figure 2. An overview of the structure of the ILO evaluation function
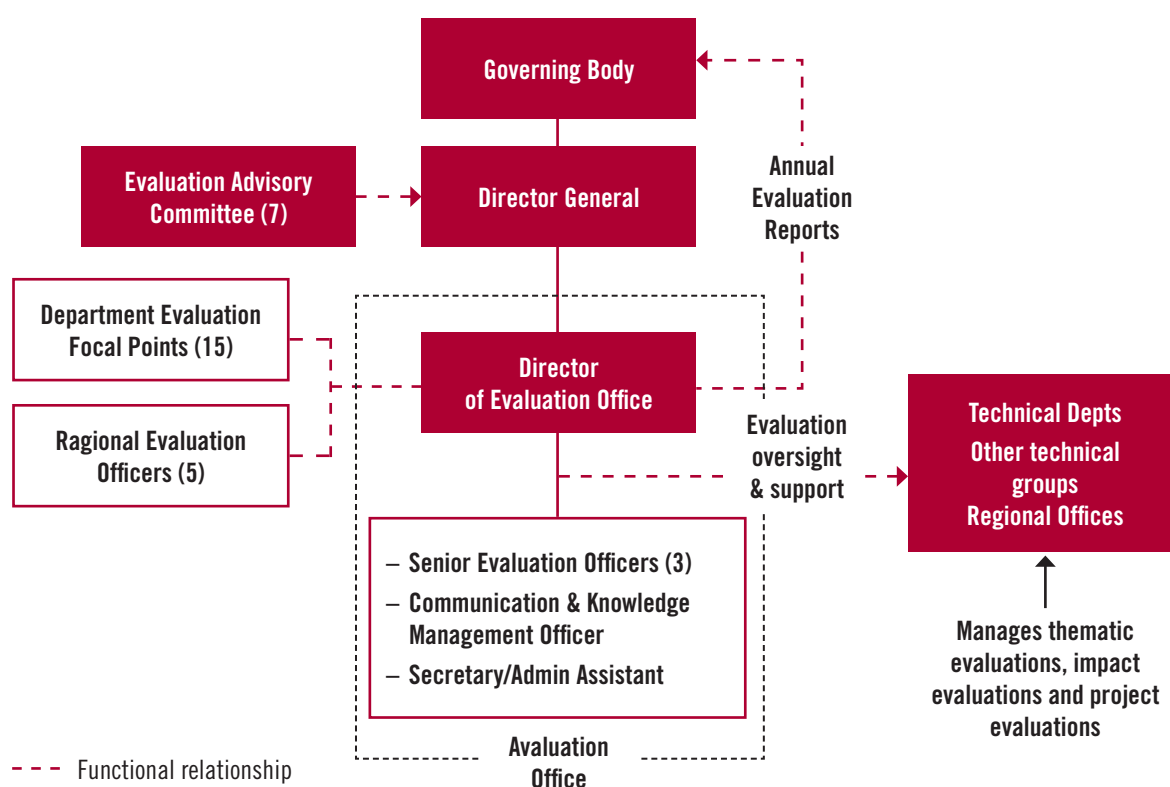
The full-time REOs[12] have dual reporting lines. Administratively, they report to regional management and functionally to EVAL. The DEFPs form part of an office-wide evaluation network (including EVAL staff and the REOs) established by EVAL to facilitate the harmonization of evaluation policies and practices. For the implementation of decentralized evaluations, the ROs identify evaluation managers within ILO.

The EAC reports directly to the DG's Office and is made up of the directors of PROGRAM and PARDEV, two directors of technical departments, two regional representatives (regional directors), and one official from the DG's Office. Its chair is the DDG for Management and Reform. The Director of EVAL serves as the secretary of the EAC.

---

[12]   In some regions, it is now officially called "Regional Monitoring and Evaluation Officer".

The users of the evaluations can be divided in two main groups (table 4):

- primary intended users: departmental, regional and country directors, technical and specialist staff and constituents; and
- secondary intended users: PARDEV and PROGRAM staff members.

4.5 OTHER CONTEXTUAL FACTORS

In line with the ToR, the IEE examined the overall organizational environment for the implementation of the evaluation function. This included an assessment of the working relationship between EVAL and other entities, including but not limited to those presented in the chart above, and the usefulness of existing evaluation products (including different types of evaluations).

To ensure an enabling environment for the evaluation function is one of the UNEG institutional norms. It entails the establishment of an organizational culture that values evaluation as a basis for accountability, learning and evidence-based decision-making, a firm commitment from the organization's leadership to use, publish and follow-up on evaluation outcomes, and recognition of evaluation as a key corporate function for achieving results and public accountability. As indicated by EVAL staff members, establishing a richer evaluation culture within the ILO is one of the main areas for consideration in implementing the evaluation strategy.

A unique contextual factor for the ILO evaluation function is the tripartite structure. This is especially relevant in terms of considering the enabling environment for evaluation. Furthermore, as a normative entity of the UN System, evaluation is expected to contribute to national implementation and UN system support for international labour standards (ILS). One emerging entry point that intersects with this is the need for ILO to respond to the SDGs and the 2030 Agenda for Sustainable Development.

EVAL has already sought to identify the major implications of this Agenda[13] for the evaluation function. These are centred on the twin role of: (1) enabling the Decent Work Agenda (DWA) within the SDGs through evaluative capacity and practice of member States and the ILO; and (2) recognizing and enhancing ILO's contribution as regards other entities in the achievement of the SDGs. In particular, this will require recognition of country-level capacity constraints and actions to address these, and support to evaluative analysis for learning during the implementation of the 2030 Agenda (adding 'how' and 'why' to the 'what' of the existing focus on tracking indicators). Related to this will be changes or adjustments in how the UN system works at the global and country level in support of the 2030 Agenda.

Finally, the performance of the evaluation function in ILO will continue to be framed by changes to norms and standards at the global level, including ongoing revision of the UN SWAP Evaluation Performance Indicator and initial dialogue on the revision to the standard OECD DAC evaluation criteria.

---

[13]   On 1 January 2016, the 17 Sustainable Development Goals (SDGs) of the 2030 Agenda for Sustainable Development – adopted by world leaders in September 2015 at an historic UN Summit – officially came into force. Over the next 15 years, with these new goals that apply universally, countries will mobilize efforts to end all forms of poverty, fight inequalities and tackle climate change, while ensuring that no one is left behind.
The SDGs build on the success of the Millennium Development Goals (MDGs) and aim to go further to end all forms of poverty. The new goals are unique in that they call for action by all countries, poor, rich and middle-income to promote prosperity while protecting the planet. *Sustainable Development Goals: 17 goals to transform our world*, UN, 2016, www.un.org/sustainabledevelopment/development-agenda/ [accessed 1 March 2017].

5 THEORETICAL BASIS FOR THE EVALUATION

The analytical framework for the evaluation is conveyed by the Evaluation Matrix (see Annex I, Vol. II), as refined during the inception phase. Based on the evaluation purpose, scope and specific issues defined in the ToRs, the Evaluation Matrix identifies four evaluation questions and indicators for each issue/sub-criterion to be studied. In overall conformity with the ToRs, the issues/sub-criteria and indicators are categorized into four main areas: independence of evaluations and evaluation systems, credibility of evaluations, utility of evaluations, and enabling environment (figure 3). Considerations regarding organizational efficiency and sustainability are regarded as crosscutting in nature.

Figure 3. Analytical framework for the evaluation (developed by the evaluators based on UNEG peer-review guidance)
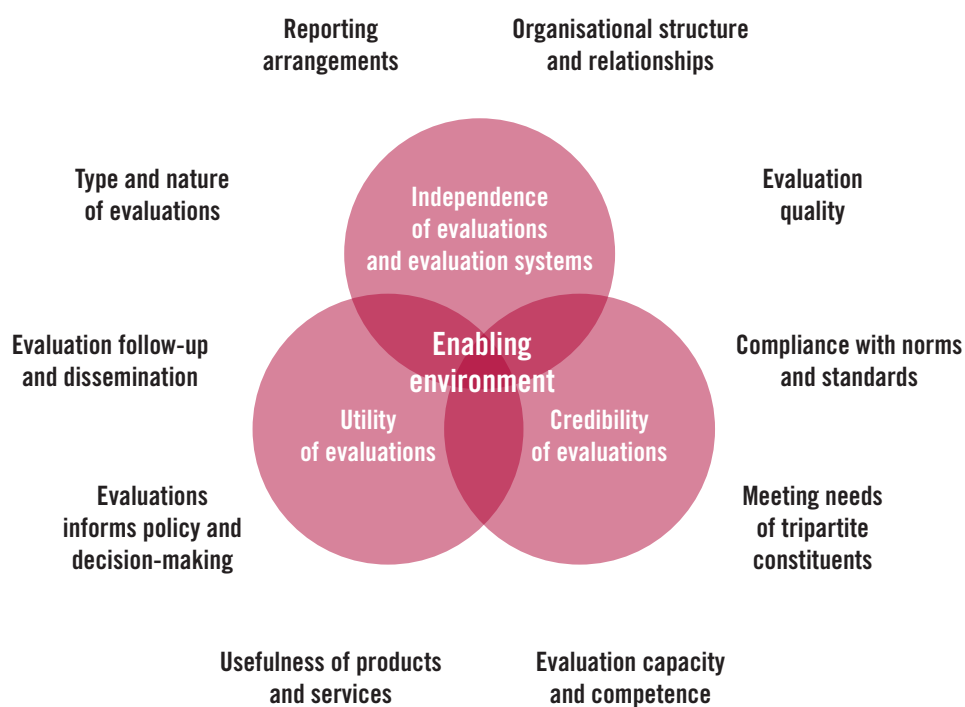
According to the ToRs and as visualized above, UNEG norms and standards have been specifically used as a basis for comparing the assessment of issues relating to the credibility of evaluations. However, given their all-encompassing nature, these norms and standards have been treated as a

general benchmark for the assessment of all four areas. The JIU's model of the components of the evaluation function, and other norms and standards, including those of OECD/DAC, has been referenced as and when relevant.

During the inception phase, EVAL mapped the main activities undertaken under the evaluation strategy to a matrix that combines the UNEG framework (independence, credibility, utility and enabling environment) with the three outcome areas of the strategy. This framework has been used as a starting point for interrogating questions of contribution (see Annex F).

6 METHODOLOGY

The evaluation design is informed by and has evolved from a number of precedents: the *Independent external evaluation of the evaluation function* (ILO, 2010); the *Analysis of the evaluation function in the United Nations system* (JIU, 2014); and the *UNEG framework for professional peer reviews of the evaluation function of UN organizations* (2011).

In accordance with these precedents, the evaluation applies three main design principles:

- **theory-based:** assessing the performance of the evaluation function based on normative frameworks established by UNEG, the ILO evaluation policy and the ILO evaluation strategy;

- **utilization-focused:** maximizing the utility of the evaluation process and products to the primarily intended users; and

- **systemic:** taking into consideration the multi-dimensional aspects of the evaluation function in ILO, including – among others – elements of strategic design and evaluation quality, as well as elements of organizational efficiency and knowledge management.

These principles were realized through a mixed-methods evaluation using parallel multiple lines and levels of evidence (qualitative and quantitative) as a means to triangulate and interrogate findings. Three main methods were used: (1) case studies of each ILO region and HQ; (2) a purposively sampled portfolio analysis of 20 evaluations;[14] and (3) a realist synthesis of secondary evidence. These were combined through the use of contribution analysis. Contribution analysis involves six main steps: setting out the attribution problem; establishing the ToCs; gathering existing evidence for them; assembling the contribution story and challenges to it; seeking additional evidence; and strengthening the contribution story.

Each finding is presented along with the main sources of data used. To aid interpretation, a qualitative rating of the levels of evidence supporting the finding is also presented:

- Level 4: Multiple lines and levels of evidence with very strong triangulation;

- Level 3: Multiple lines and levels of evidence, most of which triangulate;

- Level 2: Limited lines and levels of evidence with strong triangulation; and

- Level 1: A single line of evidence, or weak triangulation.

The main activities carried out for the collection of data and information were:

- d**esk reviews:** more than 150 documents were reviewed, the full list is provided in Annex A;

- **field visits:** three short field visits to regional offices (Abidjan, Bangkok and Geneva) and two virtual country visits to regional offices (Beirut and Lima);

---

[14] Sampling criteria and a list of the selected evaluations are provided in annex H

- **interviews:** 138 persons were interviewed both face to face or through skype/telephone, see Annex B for the full list;

- **surveys:** two surveys (architecture and users) in three languages (English, French and Spanish) were submitted to 186 persons with a response rate of 27 per cent, see Annex C for an overview of the results; and

- **portfolio reviews:** a selection of 20 evaluations were reviewed using checklists (ToRs, report and management repose), self-assessments and interviews with the evaluation consultant, evaluation manager and an intended user, see Annex D for an overview of the responses to the self-assessments and interviews.

The full method and tools, including how the evaluation applied UNEG guidance on integrating human rights and gender equality, and on ethics, is included in annexes H and J, Vol. II.

6.1 A RECONSTRUCTED TOC FOR THE EVALUATION FUNCTION

The ToC of the evaluation function is partly conveyed by the performance framework included with the strategy 2011–2017. The performance framework defines the desired outcomes of the strategy and includes indicators, biennial targets and end targets for quantifying these outcomes and measuring to what extent they have been achieved. The performance framework does not identify outputs and activities, or risks and assumptions. According to the narrative part of the strategy 2011–2015, the performance framework is based on the expectation that "ILO will continue to make progress in nurturing a culture of evaluation (enabling environment), where managers and constituents take ownership of the evaluation process as a means to improve their own areas of work". In relation to this, the evaluation strategy assumes that:

- there is a commitment to use evaluations in decision-making at all levels;

- improvements in evaluation quality will lead to greater use and demand for evaluations;

- evaluation capacity will be sufficient to ensure methodological development and learning opportunities; and

- extra-budgetary resources will be made available to train constituents on evaluation management and implementation.

The above listed assumptions are recognized as relevant to the evaluation strategy and fully support the implicit ToC which is described below (figure 4).

The chart below present the reconstructed the ToC based on the information contained in the performance framework, and the strategies, activities and outputs described in the narrative part of the evaluation strategy. The impact statement is a truncated version of the vision informally defined by EVAL in consultation with the evaluation network during a biennial workshop in 2012.

6.2 LIMITATIONS

The evaluation team identified a number of limitations, assumptions and constraints that will have implications on the primary intended uses (purpose) of the evaluation. These are:

- The **selected methods prioritize learning**. Contribution analysis enables the interrogation of the probable influence of the evaluation function on intended outcomes, but does not provide statistically significant attribution.

- The **design principles** (theory-based and utilization-focused) include an inherent focus on the perspective of the primary intended users of the evaluation (ILO). Some (but not all) unintended effects and explanations may be identified through the inclusion of interviews

Figure 4. Reconstructed ToC for the evaluation function

Impact

A more relevant, efficient and effective ILO with greater impact on the lives of people it serves

Outcomes

Improved use of evaluation by ILO constituents and management for governance	Harmonized office-wide evaluation practice to support transparency and accountability	Evaluation capability expanded through enhanced knowledge, skills and tools

Intermediate outcomes

■ High-level evaluations better inform governance-level strategic and programming decisions ■ Increased effectiveness of the Evaluation Advisory Committee (EAC) ■ Annual report based on analysis of evaluation reports	■ Organisation-wide mechanisms for engaging constituents in evaluation process ■ Type of evaluations and associated roles and responsabilities standardised ■ Use of decentralised evaluations expanded and upgraded	■ Increased evaluation capacity and practice among ILO officials and constituents ■ Standardised evaluation roles and responsabilities ■ Improved use of knowledge systems for evaluation

Outputs

■ Evaluation topics are selected according to established criteria and strategic use ■ Rolling three-year evaluation programme of work presented to the GB and updated annualy ■ Regular EAC meetings ■ EAC action plan with schedules, procedures and deliverables ■ EAC recommendations on evaluation programme of work and use of evaluation ■ Constituents have adequate opportunities to comment on avaluation plans, outcomes and follow-up ■ Regular synthesis of conclusions, recommendations and lessons learned from project and other types of evaluations ■ System for acquiring performance data on regular basis ■ Internal peer reviews conducted to ensure that high-level evaluations are of satisfactory quality	■ Two-year rolling work plans of decentralised evaluations ■ Internal and self-evaluations reviewed by the Evaluation Office and available to EAC and GB ■ Sectors and regions benefit from Evaluation Office technical advice and support ■ One independant DWCP evaluation conducted per year ■ Thematic, project and all forms of internal reviews and self-evaluation decentralised ■ Guidelines and good practices for impact evaluations and joint evaluations ■ IT-based compilation and analysis of evaluation reports	■ Constituents and ILO officials trained in specialised evaluation skills ■ Evaluation responsabilities standardised and specified in job description ■ Performance appraisals carried out of all evaluation officers and focal points ■ Evaluation network institutionalised ■ Certification procedures for evaluation managers ■ Common approach to decentralised evaluations, including monitoring and follow-up ■ I-Track online evaluation database more user-friendly and demand-responsive ■ I-Track integrated with office-wide knowledge management system

with a variety of stakeholders, but the evidence available to the evaluation is naturally limited by the perceptions and knowledge of informants.

■ The evaluation sought to include **social dialogue** with the tripartite constituents through meetings with GB representatives, and interviews with constituents at the country level. The depth of this participation was constrained by the time available to the evaluation and the limited response of some stakeholders (such as donors).

■ The **response rate** of the staff survey was 27 per cent, despite three rounds of reminders. The qualitative and quantitative data from the survey have thus only been used in triangulation with other evidence to draw out findings.

■ The **practical constraints** of time, sampling of the portfolio analysis, depth of the desk review, and different treatment of regions (as separate case studies) and HQ (as part of the overall synthesis) limited the level of triangulation possible in the final synthesis.

■ Not all **UNEG evaluation criteria** were covered by the evaluation. The evaluation assumed stability of institutional funding and did not, therefore, assess the sustainability of the evaluation function in multiple scenarios. In accordance with the ToR, the evaluation also did not assess the impact of the evaluation function in terms of its effects on the performance of ILO interventions.

7 FINDINGS

7.1 THE PERFORMANCE AND USE OF EVALUATION IN ILO

7.1.1 Finding 1: Overall quality

Overall, the quality of the evaluation function in ILO has increased significantly since the 2010 IEE, and has been recognized by several independent assessments as exceeding the performance of comparable organizations in delivering on international norms. This is especially the case regarding independence and standardizing levels of credibility. Maintaining this comparative strength requires: improving the extent to which independent evaluations meet UNEG standards for evaluation design principles, labour rights and gender equality; and renewed emphasis in the evaluation strategy on maximizing the utility of evaluations at the decentralized level.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	4/4
Constituents	X	–	–	–	
External	X	–	–	X	
– = nil.					

Overall, the evaluation finds significant and relevant improvement in the quality and systems of the evaluation function within the ILO. This is particularly noticeable with respect to the consistency of independent evaluations to UNEG standards, which has been recognized in assessments undertaken by Australia (2012), the Danish International Development Agency (DANIDA) (2014), and JIU (2014). JIU (2014, p. 19) acknowledged that the evaluation function in the ILO is transitioning from focusing on "systemizing and routinizing" evaluation to "refinement and enhancing use". The ultimate aim of the evaluation function, in the JIU maturity model, is as an agent of change within the organization and the wider system: "On average, the evaluation functions of only six organizations are assessed to be ready to address the challenges. They are ILO, UNICEF, UNDP, UNIDO, UN-Women and WFP. They are assessed to have the requisite outward-looking focus and a definite intention of addressing global changes and emerging imperatives." (JIU, 2014:45)

Under the current evaluation strategy, the evaluation function completes an average of up to 330 evaluative activities each year (an average of 55 separate activities for each full-time evaluation staff member in EVAL and in the regions (table 7). Many of these activities involve multiple tasks (such

as each evaluation requiring tracking in *i*-Track, supporting the recruitment of evaluators, multiple rounds of quality assurance, formatting, and disclosure). EVAL also supports impact evaluations, independent quality assurance of evaluation reports, and participation in UNEG.

Table 7. Average volume of work undertaken by the evaluation function each year

Activities	Description
246 project evaluation activities	Independent evaluations (45), internal evaluations (34), management responses (37), lessons learned (60), good practices (25), summaries (45)
9 governance-level activities	DWCP evaluation (cluster of 4), DWCP reviews (6), strategy evaluations (2), synthesis reviews (2), AER (1)
53 guidance updates	Policy guidelines (1), guidance documents (50), training modules (2)
22 outreach services	Newsletters (4), meta-analyses (2), think pieces (3), EAC meetings (4), EFP meetings (5), website, intranet, Plone, *i*-Track.

Mapping these activities to the outcomes of the evaluation strategy and the evaluation criteria (table 8) reveals several characteristics of the emerging focus of the central evaluation function. In particular, a substantive body of work is emerging around the utility of evaluation at the governance level (Outcome 1), whilst the focus at the decentralized level (under Outcome 2) is around harmonizing the use of credible evaluation practices.

Table 8. Level of activities of the central evaluation function mapped to the evaluation strategy performance framework and the IEE evaluation criteria

PERFORMANCE CRITERIA	OUTCOMES		
	1. Improved use for governance	2. Harmonized evaluation practice	3. Evaluation capability
Independence	**HIGHEST** Major focus in early years of strategy through restructuring High proportion of recurrent activity through quality oversight.	**HIGH** Strong focus through guidelines, evaluation management and evaluation network	**LOWEST** Covered by evaluation manager certification programme, but not a strong focus of activity
Credibility	**MEDIUM** Some activity through the introduction of strategic evaluations and meta- studies	**HIGHEST** High proportion of recurrent activity through quality assurance, evaluation management, synthesis and major investment in tools, facilities and guidelines	**LOWEST** Ad hoc activities to support organizational knowledge and initiatives.
Utility	**HIGH** Strong focus at central level, especially in recent years Many new communications and know management products launched, new initiatives piloted, and recurrent effort of *i*-Track	**MEDIUM** Some activity to support and guide utility at the decentralized level Investment in knowledge sharing platform	**MEDIUM** Covered by constituents training and ad hoc activities Emerging area of activity around impact evaluation
Enabling environment	**HIGHEST** Major focus in early years through restructuring and securing REOs.	**MEDIUM** Recurrent support to DEFPs and evaluation management	**MEDIUM** Periodic activities to support Programme & Budget (P&B) and strategic planning

Structural independence

UNEG's Norm 6 (2005) and Norm 4 (2016) elaborate a requirement for behavioural independence and organizational independence of the central evaluation function. According to JIU (2014:15), ILO is now one of only six UN entities to have achieved the highest level of structural independence of 24 assessed entities in the UN system. This is a major achievement in comparison with the baseline indicated by IEE (2010) findings.

As much as 90 per cent of respondents to a survey of the evaluation architecture expressed a belief that the current policy and strategy successfully support independence of the evaluation function. Whilst several interviewees commented that the head of evaluation still reports to the DG, and is not fully independent, none challenged the applied independence of EVAL in its work to implement the policy and strategy.

As noted previously, structural changes have substantively enhanced the level of organizational independence. There were some concerns expressed by interviewees regarding two aspects of behavioural independence. Firstly, that freedom of information exists technically, but is not yet fully supported by a culture of openness due to the perception of evaluation as a threat and the need to prove the success of projects. Secondly, that evaluators recruited primarily for their thematic knowledge can be influenced by their professional preferences rather than objectively evaluating the technical aspects of projects.

Behavioural independence

Impartiality is required under UNEG's Norm 5 (2016), defined as objectivity, integrity, and the absence of bias. External consultants are used in all independent evaluations and many internal evaluations conducted by ILO. An examination of the ToR for the portfolio analysis found that only 25 per cent of evaluations sufficiently identify the need for evaluator impartiality. However, 50 per cent of evaluation ToR do sufficiently meet requirements for specifying roles and responsibilities and referencing ILO evaluation guidance.

In general, it is reported that it is difficult to find appropriate and available consultants. Tender procedures involving companies are largely avoided for budgetary and time reasons. Individual consultants are identified through different means. In some cases, expressions of interest are received based on announcements posted by the evaluation officers on a list server maintained by EVAL, or through consultancy rosters maintained by EVAL and the REOs. Whilst evaluation managers are technically responsible for the recruitment of evaluators, they are heavily dependent on the REOs' specialist knowledge and advice.

The REO or evaluation manager prepares a shortlist and shares this with concerned technical specialists, project teams, senior line managers and EVAL staff for comments, on the basis of which the final selection is made. Although the IEE has not allowed for a more in-depth study of the procurement of evaluation consultants, observations suggest that the process could be more transparent and competitive.

There is concern that some of the same consultants are hired repeatedly, which may have implications for the overall independence and credibility of the evaluation function. Several respondents emphasized that the pool of evaluators with sufficient experience of ILO's mandate and the region is narrow, and that there is often a co-dependency in relations. For example, it was noted that evaluators are, on occasion, subsequently employed to develop the next phase of projects that they have recently evaluated.

Whilst most evaluation respondents acknowledged that the evaluation manager system practiced by ILO has value as an idea, the dominant perception is that it is a superficial form of independence

rather than the assured avoidance of bias. This view was not universally held, with some evaluation managers giving examples of where they enforced independence, but these examples were in a minority. Even in large regions many colleagues still know one another personally. The additional level of oversight provided for independent evaluations by EVAL is thus an important contribution to maintaining robust levels of independence. This hybrid system of central oversight of decentralized evaluations is increasingly being explored by other entities, including the World Food Programme (WFP) and UN Women (JIU, 2014).

Most ILO staff members consider that the evaluation manager system is primarily, therefore, a response to the Organization having insufficient M&E human resources to manage the required volume of evaluations. The level of work involved in evaluation management (which is in addition to normal duties) and the low level of incentives is resulting in low levels of retention of evaluation managers.

REOs, evaluation focal points and appointed evaluation managers are key to the independence of the evaluation function, and the case studies found that, overall, they have a clear understanding of their roles and responsibilities. The REOs were found to be committed to maintaining the independence of the evaluation function, however, they report directly to the regional management structure, and also cover monitoring. Consequently, most of them face pressure to become involved in general programming tasks, which has implications for their later independence during evaluations.

Credibility

Under UNEG's Norm 8 (2005) and Norm 3 (2016), evaluations must be credible and of good quality, including through transparency, inclusiveness, and robust quality assurance. EVAL commissions a biennial independent assessment of the quality of independent project evaluation reports measured against UNEG standards. Of the 13 UN entities that undertake similar exercises, the JIU (2014, p. 39) found that ILO is one of only seven entities to have achieved a robust level of report quality using professional standards.

Figure 5. Mean independent quality ratings of independent evaluation reports, 2011–2015
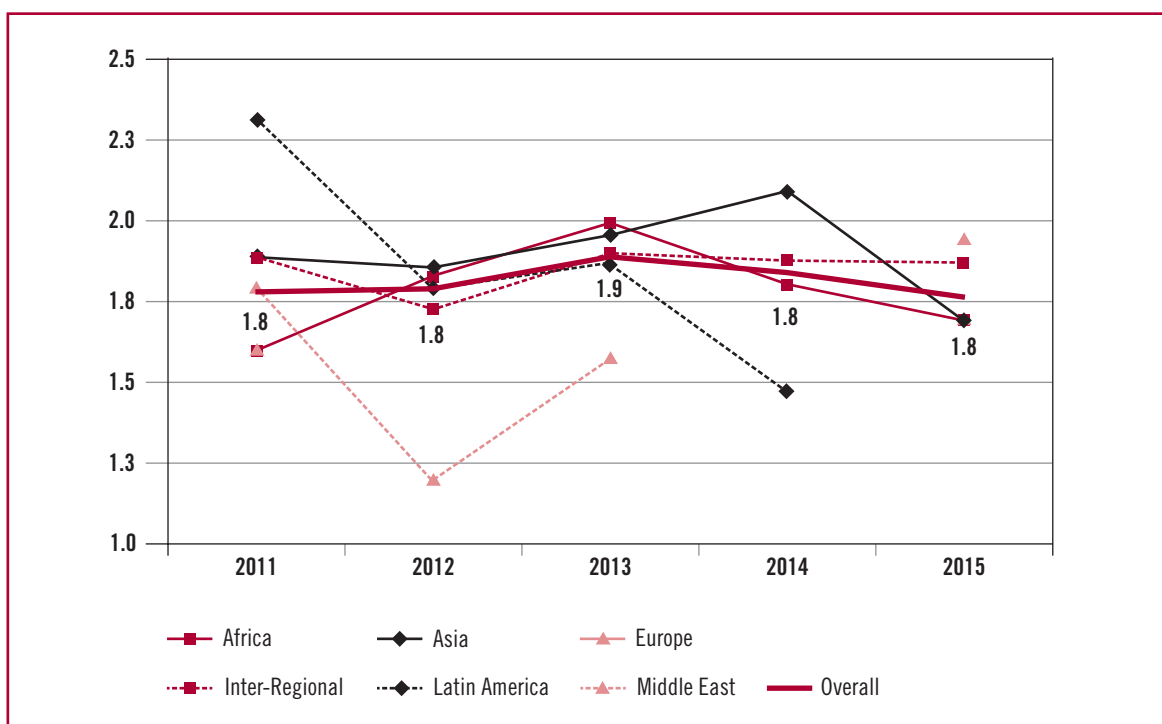
Note: Reports rated 0–3, where 2=meets UNEG standards and 3=exceeds UNEG standards.

EVAL's own independent assessments reveal that an "acceptable" level of report quality has persisted over the period 2011–2015, despite the growing complexity of evaluations. Figure 5 presents the results of ILO's independent assessments of evaluation quality using a 0–3 scale (where 2=meets UNEG standards). Within this data, however, are some trends regarding different aspects of report quality. For example, conclusions and executive summaries are consistently rated as of high quality, whereas evaluation methodologies and the integration of gender are consistently rated as of lower quality.

To answer the question of what is driving this trend, the portfolio analysis undertook a more nuanced examination of the 20 sampled evaluations (Figure 6). This rated sections of the sample of reports as exceeding, meeting, approaching or missing UNEG standards (either through being absent or because of their low quality). Regarding evaluation methodologies, it found that rather than being weak overall, evaluations are using appropriate methods for data collection and analysis. The weakness in the sampled reports was regarding evaluation design (the evaluative principles applied by the evaluation) and, most noticeably, the absence of good-quality evaluation matrices (95 per cent of reports were weak in this regard), i.e. matrices presented in the evaluation reports as part of the method section, linking evaluation question to indicator and sources of evidence.

Figure 6. Strengths and weaknesses of evaluations included in the portfolio analysis in meeting the requirements of ILO guidance (based on UNEG standards, 2005)

Analysis of evaluation ToRs in the portfolio review (covered by UNEG Standard 4.3) reveals that context, purpose, scope, criteria, gender and design are all of relatively good quality. By contrast, stakeholder analysis and the specification of evaluator competencies and impartiality were weaker (Figure 7). Overall, the portfolio suggests, therefore, that there is a challenge in translating terms of reference into a robust (and gender responsive) evaluation matrix that can support high-quality final reports.

Figure 7. Strengths and weaknesses of ToRs included in the portfolio analysis in meeting the requirements of ILO guidance
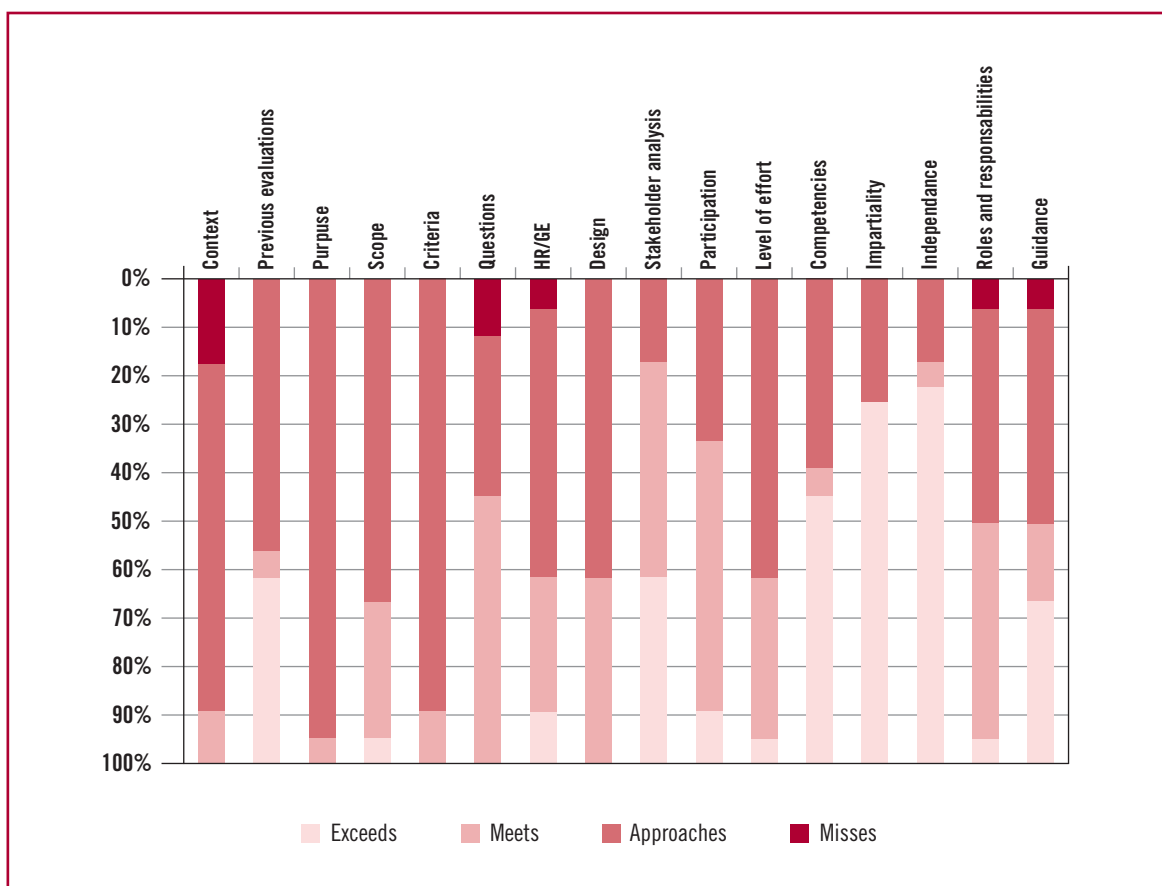
Another aspect of credibility, ethics, is covered by UNEG's Norm 11 and standards 2.5–2.7 (2005), and Norm 6 and Standard 3.2 (2016). The independent quality assessment of project evaluations (2014–2015) revealed that 55 per cent of reports discussed ethics to the required standard, whilst 60 per cent promoted participation through social dialogue (another important aspect of ethics).

In table 9, the evaluation notes strengths and opportunities for the evaluation function in ILO insofar as ethics were explicitly addressed in evaluations (noting that gaps in the system do not imply that evaluations were unethical, just that UNEG standards could be more explicitly addressed).

Table 9. Strengths of ILO evaluation systems in explicitly addressing ethics, and opportunities to improve

UNEG Standard 3.2	Strength	Current opportunities
Intentionality of use	Tripartism – interviewing constituents	Low direct involvement of rights holders
Conflicts of interest	Principle of the Independent Evaluation Manager System (when realized)	Small pool of evaluators, and some used to design subsequent projects
Interactions with participants	Consultations with social partners	Use of single evaluators prevents diversity possible in evaluation teams
Evaluation processes and product quality	Quality assessment (QA) system for independent evaluations	Low oversight of internal evaluations

There is also an issue of perception regarding evaluation credibility. The evaluation case studies found that the primary intended audience for evaluation in ILO exhibits a strong preference for particular types of analysis, often favouring quantitative data and research methods as proxies for high-quality evaluation. More broadly within ILO, there is relatively low awareness of international standards for 'good' evaluations (25 per cent of surveyed users were aware of UNEG). This highlights the challenges that have been faced in carrying out the evaluation function when balancing the needs and expectations of users with appropriate evaluation approaches and practices.

Utility

The utility of evaluations is covered under UNEG's Norm 4 (2005), where it is referred to as 'intentionality'; and Norm 2 (2016), which specifies a requirement for a clear intention to use evaluation to inform decisions, contribute to organizational learning, and ensure accountability for results. The explicit focus on utility in the revised norms and standards was partly informed by JIU (2014, p. 25) in conclusions 5 and 6 stating that the evaluation functions in the UN system have generally been driven by demands for accountability from governing bodies, rather than learning. Along with most other UN entities, ILO is rated as 'average' in terms of utility in the JIU maturity matrix (2014, p. 41).

A portfolio analysis undertaken by the evaluation of 20 sampled reports found that 94 per cent of ToR clearly stated the purpose of evaluations, and 67 per cent clearly stated the scope. Both of these elements are necessary to ensure the use of evaluation. Of the reports themselves, 85 per cent met required standards for specifying the purpose, and 75 per cent for specifying a relevant scope.

The survey of evaluation users within ILO found that all respondents were positive about the evaluation portfolio contributing to meeting the actual needs of users. The evaluation function is, however, generally seen to be more focused on ensuring that mandated evaluations be completed with less emphasis on learning aspects. It is acknowledged that the wider development environment is primarily driving this, with donors and the GB leading a demand to meet certain levels of evaluation to meet standards of accountability: "In contrast [to accountability], there has been no equivalent focus [in the UN system as a whole] on ensuring that the evaluation function is integrated into the fabric of the organization and that it plays a role in its development, and that evaluation is used to strategically position the organization to respond to a constantly changing environment." (JIU, 2014, p. 26.)

Consequently, most evaluations are being framed by the results frameworks of the evaluation object (which is most often a TC project), attenuating the extent to which evidence of wider effects and contributions to policy changes are assessed. The evaluation did not find evidence to suggest that the current portfolio of evaluations is systematically based on needs assessment for learning, accountability, or decision-making at each level of the ILO. EVAL has recognized this weakness in the AER to the GB, along with proposals for more strategic and cluster evaluations as a solution. However, this would require a change in evaluation policy which would need to be negotiated with constituents, particularly donors.

Nevertheless, an independent quality assurance of ILO evaluation reports (2014–2015) found that the executive summaries – a key element supporting utilization – were consistently of high quality across the portfolio of independent project evaluations. The communications strategy that was developed for evaluation in 2013 is another strength that ILO could build upon. It is noteworthy that prior to 2010, the annual evaluation reports constituted the only source of summary information on evaluations and evaluation recommendations. These reports have since been complemented by regular newsletters (16 in total so far) and think pieces (10), as well as meta-studies, synthesis reviews and evaluation guidelines.

7.1.2 Finding 2: Relevance of the portfolio

The current portfolio of independent evaluations is heavily weighted by the evaluation policy towards TC projects. Whilst this has been relevant to the demand of the GB and donors in the past, there is a strong and growing case for supporting the call in the AERs for greater diversity, and more strategic decentralized evaluation to meet ILO's and constituents' needs.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	–	–	X	3/4
Constituents	X	–	–	–	
External	X	–	–	–	
– = nil.					

The current evaluation portfolio

AERs carried out from 2011 to 2015 reveal that 508 evaluations were completed in this period, of which 48 per cent were independent project evaluations, 28 per cent were internal evaluations, 23 per cent were external or joint evaluations, and 3 per cent were for activities carried out under the regular budget supplementary account (RBSA). The evaluation policy estimated that up to 40 per cent of the total evaluation portfolio could be represented by internal evaluations (not all of which are submitted to EVAL). In the period under review, 17 per cent of internal evaluations were of DWCPs, and the rest were of projects.

Respondents to the evaluation survey (97 per cent) were largely of the view that the current portfolio of evaluations is relevant to the actual needs of users, although 43 per cent reported that substantial improvement is still possible. Two-thirds of respondents agreed with the view that they could influence evaluations to support their relevance.

The case studies and the survey data give credibility to the view that the current coverage of ILO's evaluation function is weighted towards TC projects. The view expressed in case studies is that this is a result of strong demand from donors, the relative ease of evaluation, and the ease of securing funds by applying the 2 per cent allocation specified in the evaluation policy. Whilst some donors are carrying out external evaluations for individual projects, the policy of providing for evaluation in TC project budgets is absolutely necessary to maintain evaluative learning and capability for both the ILO and its constituents, and is fully in line with the systems-strengthening principles of good donorship.

By comparison, the case study identified other areas of ILO's work where greater levels of evaluation would have considerable benefit. These include: (1) the work of technical specialists outside of projects (including policy advice to governments and contributing to ILS discussions at global level); (2) DWCPs, and country programme outcomes (CPOs) (including the strategic selection of CPOs); and (3) operations (including the relations between programme staff and technical specialists).

Among technical staff there is a general feeling that more strategic evaluations and less small project evaluations would be useful to better place ILO within the international scenario. Whilst the evaluation agrees that project evaluations are typically too discrete to provide fundamental insights, it also notes that the coverage of active DWCPs (and, therefore, opportunities for DWCP evaluations) is still growing (for example only 9/32 are active in the Asia Pacific region) and the project evaluation policy will continue to be required to ensure coverage of all projects.

A central question identified by one of the case studies related to the need for the evaluation policy to address the disconnect between long-term (policy-level) interventions and short-term evaluations primarily driven by a donor agenda. The question was raised as to whether a wider range of options could be considered, including independent reviews in the short term and ex-post evaluations that consider the effects of policies over a 10–15-year timeframe.

Whilst most case study informants suggested that thematic evaluations or DWCP evaluations could be more useful in this regard, a donor emphasized that these would still need to include case studies of individual projects to answer their specific corporate requirements. Any de-emphasis of independent project evaluations would, by implication, increase reliance on sufficiently strong monitoring systems to provide project-level performance data. The evaluation found that this is not an assumption that can easily be made without additional investment in RBM.

Joint evaluations

The Delivering as One and 2030 agendas emphasize the desirability of joint evaluation as a means of contributing to integrated and mutual learning and accountability. Overall, JIU (2014, p. 44) reports that the level of joint evaluation is very low in the UN system, with less than 10 joint evaluations being completed by the system each year. However, the JIU assessment also explicitly recognized ILO for contributing advisory, coordination and support to the evaluation work of other entities, with ILO data suggesting that at least 106 evaluations in the period 2011–2016 were the result of joint activities. Joint working can thus be considered a strong comparative advantage of the ILO evaluation function. It is recognized that joint evaluations can signify evaluations that are carried out together with other key stakeholders (e.g. donors and, in some cases, governments and national constituents). This is different from the concept of external evaluations, which are commissioned, managed and funded wholly by an external actor (e.g. the donor).

Impact evaluations

Since 2011, ILO has published 42 impact evaluations. In more recent years, these have been clustered in Africa and the Americas regions. Impact evaluations are the responsibility of technical departments rather than the evaluation function, and are most often used to build global knowledge on what works in advancing labour rights and decent work.

A common view within the case studies and at HQ was the need for evaluations to prioritize assessment of ILO's relevance, contribution to policy change, and long-term impact. In particular, there is an awareness of the need to demonstrate ILO's particular contribution to meeting ILS and the SDGs. Thus, there is a clear intersection between the evaluation function and the practice of impact evaluations.

Recognizing this demand, EVAL have launched a three-tier strategy, including: (1) development of an impact evaluation framework using concept mapping and based on existing work on an inventory of impact evaluations, and the ILO impact evaluation community of practice; (2) an impact evaluation review facility to support impact evaluations; and (3) a biennial quality appraisal of completed impact evaluations.

This evaluation notes that since quality assurance of impact evaluations is not a core responsibility of the central evaluation function, further expansion of impact evaluations (which is in the interests of ILO) will thus need additional resources and/or be carefully balanced with maintaining the principal mandate of performance evaluation (both at the strategic and project level). Impact evaluations also imply that interventions are designed from the ground-up with a view to being evaluated, through controlled targeting or phasing of interventions; and robust baselines, monitoring systems and endline assessments.

7.1.3 Finding 3: Use of evaluations

Whilst there is some scope to adjust the types, timing and focus of evaluations through the evaluation function, a low prevalence of specialist RBM/M&E staff within projects, field offices and departments means that ILO struggles to maximize and consistently use evaluation evidence to improve its programming.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	4/4
Constituents	X	X	–	–	
External	X	X	–	–	
– = nil.					

Timeliness and intentionality

UNEG's Standard 3.1 (2005) and Standard 4.1 (2016) specify that evaluations should be designed to ensure that they provide timely, valid and reliable information. At present, the timing of evaluations is primarily driven by the mandatory evaluations required by the evaluation policy. In general, this calls for a mid-term and a final evaluative exercise – the intensity and the independence of which is variable based on project size.

Implicit in this requirement is the need to focus on accountability, rather than responsiveness to opportunities for learning and decision-making. The case studies found that final evaluations tend only to be used for learning where there is a subsequent phase of the project, although their timing means that they often use the same data collected by project teams for the design phase. By contrast, mid-term evaluations were reported as being timelier to support learning and adjustment within the project being evaluated.

EVAL has a rolling work plan for HLEs and departmental evaluations. There is an evaluation planning process covering TC project evaluations, but this is not integrated formally into the regional RBSA M&E plan. TC project evaluations are independently monitored by EVAL (using *i*-Track) and the REOs, who inform project teams when it is time to start the evaluation process.

It could be argued that a more integrated planning system would improve sequencing, coordination and timing. Currently, as indicated by interviewees, there are signs that national constituents suffer from "evaluation fatigue". One example mentioned is Tajikistan, which this year has been subject to five different types of evaluations. This lends substantial weight to the proposal to cluster evaluations put forward through the AERs (but this is subject to agreement with donors).

Evaluability

Evaluability is not included in UNEG (2005) standards, but was introduced in Standard 4.2 (UNEG, 2016), in recognition of the importance of evaluability to the timeliness and usefulness of evaluations for decision-making. The case studies found that evaluability tends to be seen as a separate activity, rather than an integrated RBM component. This is compounded by the rarity of M&E specialists (who would normally be expected to advise on evaluability in the UN system) within technical departments, projects, and country offices.

As a result, already-stretched REOs (as M&E officers) are in high demand to support the evaluability of individual projects. As per the evaluation policy, EVAL also contributes to assessing evaluability for project proposals over $5 million through participation in the ex-ante appraisal

process. However, greater involvement of the evaluation function in ensuring evaluability is neither feasible, nor desirable from the perspective of independence. More M&E and RBM capacity in decent work programme teams and technical departments would better enable evaluability to be done throughout, thus helping in determining appropriate evaluation arrangements and contributing to better evaluations.

Whilst there is an overall underinvestment in M&E staff, good practice examples do exist. The TRIANGLE project, for example, was found to have substantially increased evaluability (and is now planning an impact evaluation) due to the recruitment of an M&E staff member following recommendations made in the mid-term review.

In the absence of sufficient M&E staff in projects and programmes, one innovative approach suggested during the case studies would be to hire independent evaluation consultants at the project start (for projects below $5 million) and to engage them at different points throughout the implementation of the project to support evaluability, assess progress, and provide iterative recommendations. This is an approach that is promoted by advocates of utilization-focused and developmental evaluations. It would also encourage a more participatory evaluation methodology whereby project stakeholders, including national constituents, would have the opportunity to contribute to evaluations in a more proactive and meaningful manner, expanding the evaluation culture beyond ILO.

Institutionalization and socialization of learning

EVAL actively collects and stores lessons and recommendations in the *i*-Track database.[15] Between 2011 and 2016, a total of 574 lessons learned from evaluations were documented and made available via the *i*-Track database. Users can search these by year, theme, country/region, strategic area, keyword, and technical or administrative office. Nearly half of all lessons learned concerned management issues. Evaluations conducted between 2011 and 2016 documented 88 emerging good practices and 1,670 recommendations, all of which are accessible in EVAL's knowledge-sharing platform.

Whilst data from the evaluation survey suggest that 88 per cent of evaluation users propose that evaluation knowledge management could be improved and mainstreamed throughout ILO, this does not account for the recent launch of *i*-eval Discovery that significantly improves the user experience of *i*-Track. In light of this improved tool, it would be interesting for EVAL to track any changes over the coming year in the 75 per cent of surveyed users reporting that they only access evaluative evidence rarely.

Evidence from the case studies suggests that more of these innovative approaches to socializing learning are required. Stakeholders in the ILO emphasized the use of dialogue (for example, brown bag lunches or the presence of EVAL staff in meetings), or more sophisticated automation (such as semantic dashboards). Evidence from country-level constituents emphasized the evaluation process itself and the need for deeper levels of participation. Most of these are already in line with EVAL's thinking and activities, but they are constrained by the level of human resourcing versus the continued quality assurance of evaluation reports.

---

[15] *i*-Track also stores all evaluation reports and good practices up to 2013, as well as all EVAL publications. Furthermore, it stores all relevant PARDEV minutes and, thereby, all evaluation planning records.

7.1.4 Finding 4: Use of recommendations

The establishment of formal systems to follow-up management responses has brought an important element of accountability to the evaluation function. This works best for HLEs, through the EAC. For decentralized evaluations, inconsistent attention from senior managers means that the use of recommendations and lessons learned often relies on there being a subsequent phase of the evaluated project

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	3/4
Constituents	X	X	–	–	
External	–	X	–	–	
– = nil.					

Evaluation recommendations

UNEG's Standard 4.16 (2005) and Standard 4.10 (2016) require that recommendations be firmly based on evidence and analysis, and be clear, results-oriented and realistic in terms of implementation. In the last biennial quality assessment, recommendations were, on average, rated as below the required standards. The portfolio analysis undertaken by this evaluation found a more positive trend with 85 per cent of reports providing relevant recommendations, and 80 per cent providing specific recommendations.

Between 2011 and 2016, ILO evaluations generated 2,410 recommendations of which 27 per cent were targeted at constituents (with a gradual increase in this proportion over the period). The case studies suggest that mid-term project evaluations are usually better followed up on than other types of evaluations since the recommendations of these evaluations tended to be more concrete and could be directly implemented through the work planning process. The recommendations of final project evaluations are harder to implement if a follow-on phase is not immediately planned. Internal evaluations are felt to be important since the experience is that many small-scale projects can generate valuable lessons learned and good practices that can be replicated.

Several interviewees highlighted the evaluation of the MAGNET project as a good example of where evaluation recommendations were directly and systematically fed into the design of a new related project, FAIRWAY. A combination of factors seems to have contributed to this result, including a high-quality evaluation consultant and report, common recognition among stakeholders of the weaknesses of the MAGNET project, and the donors' close engagement in the follow-up to the evaluation.

Strategic evaluations are usually appreciated and deemed to provide food-for-thought. For example, the recent HLE on jobs and skills for growth, which reportedly has been used as an input to the development of DCWPs in the Europe region. These evaluations' recommendations are sometimes seen by technical specialists working in the field as being at too high a level, but they directly influence the policies and strategies under which these specialists work. Interviewees in the HQ technical departments suggest that the usefulness of recommendations from HLEs is enhanced when reports coincide with the development of new strategies or other opportunities for change.

While there is a general perception that the recommendations of decentralized independent project evaluations are usually relevant and feasible, interviewees indicated that these recommendations have a tendency to be broad and relate to factors that are not within the control of the intended users. The use of recommendations appears to depend on the interest of individual staff in the purpose of the evaluation, since systems for ensuring follow-up need strengthening at regional level.

Management response and follow-up

Under UNEG's Norm 12 and Standard 1.4 (2005), and Norm 14 and Standard 1.4 (2016), the implementation of evaluation recommendations should be systematically followed up, including through a management response. An evaluation management response system for independent evaluations is in place whereby EVAL extracts the recommendations from evaluation reports and inserts them into a template that requires the primary intended users to indicate how the recommendations will be applied. The completed form is then sent back to EVAL. After six months, the primary intended user is required to up-date the form by indicating what progress has been made.

According to the AER, in 2015, 89 per cent of evaluations received a management response and, six months' later, users reported that 42 per cent of recommendations had been completed and 47 per cent had been partially completed.[16] Most interviewees could give examples of the use of evaluations. The EAC has adopted a highly systematic and rigorous process of follow-up on HLE recommendations. This is carefully documented, and has included multiple rounds of follow-up to ensure that management responses are both detailed and implemented.

Most interviewees in the case studies suggested that the management response system at the decentralized level is weaker, and 50 per cent of surveyed users were not clear on the system. Few of the senior managers interviewed through the portfolio analysis could recall to what extent the templates are filled out or what happens to them afterwards.

The impression is that evaluation recommendations are not yet being systematically integrated into the programming and project cycle. This speaks to the evaluation culture in the Organization, and suggests that there is a case for more systematic and meaningful follow-up of management responses by senior managers. Based on the contrast between the follow-up of centralized and decentralized evaluations, EVAL may consider establishing regional versions of the EAC to support this.

Meta-studies and meta-syntheses

In addition to standalone DWCP, project, policy and strategic evaluations, EVAL has commissioned eight themed meta-studies and three meta-syntheses of findings and lessons from selected bodies of evaluations. These have been shared as information resources, and have also been direct inputs into HLEs and informed recommendations presented to the GB in the AER. In this way, EVAL has sought to ensure that the large body of decentralized evaluations being produced in ILO is used to support improved decision-making at the level of central governance – a direct contribution to achieving Outcome 1 of the evaluation strategy.

In addition to supporting utility, the meta-studies and meta-syntheses have also significant potential to contribute to enhanced credibility under Outcome 2 of the evaluation strategy. The studies triangulate the findings of multiple evaluations, and find discrepancies (and gaps in evidence) in addition to patterns and correlations. These insights guide the work of EVAL in assuring the quality of subsequent evaluations. It was further noted that some of these meta-evaluations (e.g. the 2014 meta-analysis of 15 DWCPs) also include comparative analysis with preceding meta-studies, adding an additional layer of triangulation.

The main challenge, identified in the case studies, is the use of meta-analyses for enhanced programming at the decentralized level. This in part reflects the centralized focus of Outcome 1 of the evaluation strategy. Decent work teams and technical specialists were universally found to consider the current meta-analyses as being too global in scope to significantly inform their specific requirements (a perception that is not fully supported by close reading of the substance of meta-analysis reports).

---

[16]   The reliability of the data may be questioned since it is based on the information reported by the relevant organizational unit and does not include a qualitative assessment of the relevance and change brought about through the agreed action.

Addressing this challenge of 'downwards' use of global evidence (in terms of both perception and substance) could be a key tenet of revisions to Outcome 1 of a future evaluation strategy.

7.2 VALUE OF EVALUATION

7.2.1 Finding 5: Evaluation products and services

EVAL has demonstrated a commitment to progressively improving the design and communication of evaluation products and services to increase utility. Many new product types have been released since the IEE (2010). Considerable potential exists for evaluation communication to contribute to strengthening ILO as a learning Organization.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	2/4
Constituents	X	–	–	–	
External	–	–	–	–	
– = nil.					

Evaluation function documentation

Since 2011, EVAL has developed a comprehensive set of tools (e.g. checklists, templates, protocols), guidelines, newsletters (16 in total), think pieces (10 in total), fact sheets, videos and on-line resources (e.g. EVAL website and *i*-Track) to serve the evaluation function. When surveyed, 98 per cent of respondents within the evaluation function find these to be helpful (55 per cent) or very helpful (43 per cent). In addition to these tools, meta-studies and synthesis reports have been published to make evaluation findings available to a wider audience. These are seen by interviewees as being useful products to support emerging programming.

At the level of the GB, EVAL has steadily enhanced the AER to address key strategic themes and present lessons learned in addition to accountability for the evaluation strategy. Constituents from the GB requested that comments given to the AER might be followed-up and communicated in detail at subsequent meetings to demonstrate the impact of evaluation as a function. There was also a request for a stronger mechanism to feed political and policy issues raised by evaluations into International Labour Conference (ILC) discussions.

At the country level, constituents requested a higher level of participation in ILO evaluations – an action that could also increase their knowledge and understanding of the evaluation function when they subsequently attend the GB.

Evaluation reports and products

UNEG's Standard 3.16 (2005) and Standard 4.9 (2016) establish the requirement for logical and useful evaluation reports. However, it is important to frame analysis of the usefulness of evaluation products by two caveats identified during the case studies: (1) evaluations can only be as useful as the host culture is 'open' to inform them; and (2) existing ILO evaluations tend to be self-referencing, consulting the same stakeholders that were involved in project design (or with ILO's broader portfolio), and lacking in a wider set of voices.

It was emphasized that addressing these issues requires both ILO leadership and individual evaluation managers to set a tone of learning, by establishing evaluations as a safe space to be frank. It was proposed that addressing the range of voices included in the evaluation process is required at two

levels: (1) supporting constituents in the use of new technologies to harvest evidence from a much wider public; and (2) the inclusion of external members in the EAC to add a wider perspective and to bring in knowledge from outside the Organization.

EVAL has made a significant effort to standardize the structure and quality of reports, including for example, providing model guidance on how to prepare adequate lessons learned and good practices. Apart from meeting UNEG standards, however, the case studies found that evaluation reports tended to make assumptions that could affect their use. For example, that additional human resources could be made available to the projects, or calls for action by national constituents, who also tend to have limited capacities and resources.

Some of the regional case studies identified language as a barrier to wider participation of the social partners in evaluation processes. Draft reports in English, French or Spanish are often a second or third language for them, and evaluation reports tend to be long and unwieldy for non-evaluators. Whilst EVAL has attempted to diversify the range of materials available to communicate evaluations reports, there remains significant scope to address multilingual documentation during evaluation processes.

For other ILO staff and donors, the main challenge identified with TC project evaluations was not the communication of knowledge, but the absence of genuinely new information. In most cases, it was felt that evaluations synthesized what was already known. The greatest use, therefore, came from evaluations that considered specific questions of interest to the primary intended users, such as examining a particular innovation, or using new evidence to propose future intervention models for ILO.

The SDGs are seen by some ILO staff as a potential entry point for this discussion, with the Office acutely aware of the need to delineate ILO's contribution to the DWA now that it is mainstreamed across multiple SDGs and targets. To this end, recent work by EVAL in an *i*-eval Think Piece (December 2016) emphasizes the need to promote evaluative thinking throughout the ILO, support the paradigm of national evaluation capacity development, embed the SDGs into clear ToCs for each ILO intervention, and build a comprehensive picture of ILO's performance story through enhanced M&E activities.

Communications and knowledge management

Communication and dissemination were not explicitly included in UNEG norms and standards (2005), but were introduced in UNEG's Standard 4.11 (2016). This specifies that communication and dissemination are integral and essential parts of evaluations. Recognizing the importance of communications to the use of evaluation, EVAL had already adopted a communications strategy for evaluation in 2013, based on a user survey and more effective utilization of existing resources. In addition, the *i*-Track system, which has continuously been upgraded, is far more comprehensive than the report repositories used by other UN entities (for example, by tracking all TC projects and linking to PROGRAM databases).

In comparison with the baseline established by the IEE (2010) and the evaluation strategy (2011), there has been a major increase in the number and scope of communications activities, despite having no additional budget from that envisaged in the 2011 evaluation strategy. Communications products (figure 8) now include an *i*-eval Resource Kit, *i*-eval Flash News (triennial), Think Pieces (10 so far), cartoons, meta-studies, synthesis reviews, EVAL website, a new dashboard (*i*-eval Discovery), and direct engagement (seminars, trainings, webinars). Survey respondents (75 per cent) receive or access these communications products and knowledge-sharing platforms.

Figure 8. Communications and knowledge-management activities since 2013
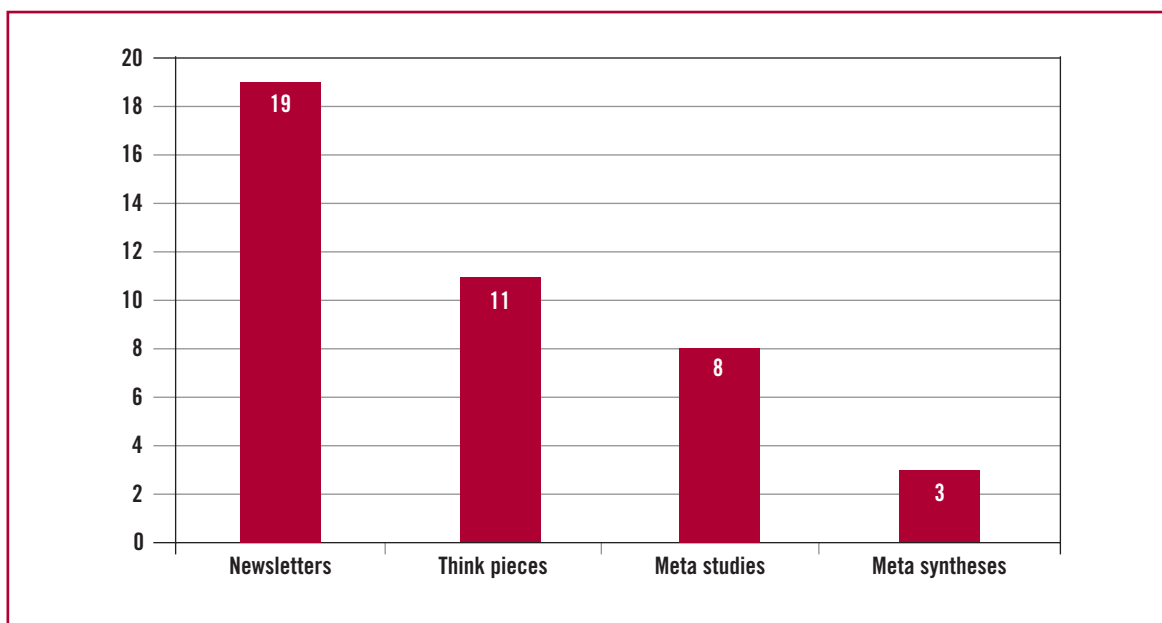
Based on the case studies and HQ interviews, the evaluation finds that the 2013 evaluation strategy was an important step in the development of communications and knowledge management. Given the experience gained since 2013, integrating knowledge management into a future evaluation strategy, however, will require that some of the assumptions on which it was based will need revisiting. For example, the strategy emphasizes broadcast platforms (websites, publications) when interviewees emphasize the importance of dialogue around their needs. A positive example of this is the inclusion of EVAL staff as resource persons in some of the recent discussions to develop flagship programmes.

The evaluation also finds that the underlying analysis of satisfaction of EVAL's services and products is probably influenced by misunderstandings of the evaluation function. For example, at the global level, the evaluation observed that many stakeholders could not fully differentiate evaluation, audit, monitoring and other accountability layers. In the absence of a clearly worded and commonly understood mission statement for the evaluation function, stakeholders are projecting their own vision of evaluation when defining the boundaries of EVAL's role.

Case studies suggest that visibility and understanding of evaluation (including greater prioritization of evaluations) can be effectively enhanced through the presence of EVAL staff at events such as regional directors' meetings and DWCP committees. Given the importance of enhancing communications to address the understanding of the evaluation function (including, potentially, through travel), the evaluation disagrees with the explicit assertion in the communications strategy that no additional budget is needed. There is, however, a return on investment given that at least one case study suggested that increasing the visibility of evaluation is leading to that regional office strongly emphasising evaluability in the next generation of DWCPs.

Human resources

The use of evaluation evidence, over any period, is hampered in regions by a low level of human resources in key positions. Few of the TC projects or country offices have dedicated M&E staff, meaning that there is a little capacity to distil evaluation evidence into key points for use by management, or to ensure the evaluability of projects under development. Since some of the EVAL communications products are explicitly intended to help address these challenges (such as the new *i*-eval

Discovery interface), there is a case to be made for updating the communications strategy with more precise audience disaggregation.

In the absence of human resources to support the socialization of evaluation insights, ideas shared by participants for packaging evaluation information to improve its use include: (1) providing standard texts (with links to evaluation evidence) for the section of the Document, Publications Production, Printing and Distribution Branch (PRODOC) template that requires information on ILO's experience and unique value proposition; (2) improving connection to local and regional evaluation networks and think tanks; (3) using standardized key indicators and meta-data across all evaluations (linked to DWCP/CPOs) to allow for an automated analytics dashboard so that managers can have more informed conversations with constituents: "For example, I could generate a dashboard for all evaluation reports from [a specific donor] funded programmes around the world with an infographic of what we are doing well, less well, and where we should concentrate." (ILO staff member.)

The evaluation notes that EVAL has been responsive and proactive in identifying and addressing weaknesses identified in existing knowledge tools, such *i*-Track being based upon superceded technology that limits user-friendly interactions. Survey respondents highlighted the value they see in having access to such dashboards of synthesized evaluative knowledge, as well as to direct dialogue with EVAL staff in order to socialize knowledge.

7.2.2 Finding 6: Tripartite needs

Evaluation activities systematically consult ILO constituents, but social partners' demand-for, participation-in and ownership-of an evaluation is relatively low. The design and conduct of evaluations insufficiently models social dialogue, and project evaluations infrequently address strategic questions that are of most interest to national constituents.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	3/4
Constituents	X	–	–	–	
External	–	–	–	–	
– = nil.					

Tripartite engagement

Tripartite constituents engage with the evaluation function at three main levels: (1) through the GB; (2) through the DWCP process; and (3) in individual evaluations as informants. From a prospective perspective, the importance of a responsive evaluation function is established under UNEG's Standard 2.3 (2016), and national evaluation capacity is a major priority under the 2030 Agenda. In this regard, JIU (2014, p. 49) found that only two entities, the United Nations Development Programme (UNDP) and the United Nations International Children's Emergency Fund (UNICEF), have so far integrated national evaluation capacity into the operations of their central evaluation functions.

In ILO, it was found that the evaluation function contributes significantly to accountability to constituents at the GB (through both the AERs and HLEs), but much less so at the country level. Constituents were frequently consulted as part of evaluation processes, but not involved in the commissioning, purposing, design, implementation, interpretation, or communication of evaluations. However, follow-up action was still taken in 94 per cent of cases where evaluation recommendations were directed to constituents.

In the case studies in particular, it was noted that the participation of constituents in project monitoring and subsequent evaluations is weak, partly because it is not integrated into national systems.

The ILO evaluation function survey data contrast with this finding because primary intended users within the function are principally thought of by most respondents as being restricted to ILO staff.

As a result, national constituents indicated that they do not always receive the reports of the evaluations to which they are party as interviewees. They would like to be involved to a greater extent in the evaluation process and to receive training in order to improve their internal evaluation functions. Whilst the UN system is increasingly supporting national evaluation capacity, it focuses on governments and civil society organizations (CSOs). Thus, there is a need for ILO to advocate for supporting employers and workers through its position in UN country teams.

Dialogue and participation

In addition to social dialogue being a core value of ILO, UNEG's standards 3.11 and 3.12 (2005), and Standard 4.6 (2016) require that evaluations have inclusive and diverse stakeholder engagement. According to the 2014–2015 independent quality assessment, 60 per cent of project evaluation reports reflect that commitment.

The case studies found that the social partners' demand for ILO evaluations is quite low, and the use of evaluations as an opportunity to promote dialogue with constituents and government counterparts is not being systematically encouraged. Furthermore, evaluations are not being used to assess questions of interest to constituents, such as the contribution of all social partners, barriers to realizing ILS, and whether ILO is focusing in the right places.

The reportedly low level of integration of human rights-based approaches into either project design or evaluation characterizes evaluations as being consultative rather than participatory. Both tend to focus on institutions, and large aggregated groups of people, rather than on the identification of the disaggregated groups of ultimate beneficiaries (duty bearers, rights holders). Of the 2014–2015 reports, as much as 86 per cent did not disaggregate data by beneficiary groups. However, the evaluation notes that its sample was oriented towards project evaluations, and that with more focus on strategic and impact evaluations, greater participation would become more feasible.

ToRs and evaluation reports are circulated to relevant ILO departments and staff members (technical specialists, project team members, line managers, etc.), and often also to donors and social partners, for comments and suggestions. However, this type of feedback mechanism is not participatory and does not always generate the desired response, especially since many people have very limited time availability.

In addition, when the draft evaluation report has been prepared, it may already be too late to come back with substantive comments and suggestions. Evaluation report findings are often shared with national constituents at workshops organized for this purpose or during DWCP steering committee meetings and/or through the circulation of executive summaries of the reports. Generally, however, the involvement of national constituents and beneficiaries is limited and passive (as interviewees).

There are few examples of national constituents or beneficiaries being involved in the planning and implementation of evaluations, or attempts by ILO to build corresponding national evaluation capacity. As such, there are few opportunities for constituents to influence and oversee evaluations and follow-up actions. This observation leads on to a more general issue of the need to develop more utilization-focused evaluation approaches, as further discussed below.

Despite this overall trend, one case study did identify good practices of participation, including the evaluation of TACKLE in Fiji, which embedded the inclusion of social partners in the ToR and was reported by ILO regional staff to have led to significant learning at both country and global levels.

7.3 ENABLING ENVIRONMENT FOR EVALUATION IN ILO

7.3.1 Finding 7: Enabling environment

The enabling environment for evaluation in ILO is improving, but remains fragile. Support for evaluation within the GB and by the senior management has strengthened EVAL's mandate, as has its repositioning in the organizational structure as an independent function. However, outside of EVAL, the incentives remain somewhat weak for ILO professional staff to engage in evaluative activities and maximize the force-multiplication these offer.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	4/4
Constituents	X	–	–	–	
External	X	–	–	–	
– = nil.					

Evaluation culture

UNEG established a new Norm 11 (2016) that requires an enabling environment for evaluation, defined as including: (1) organizational culture; (2) firm commitment from leadership; (3) recognition of evaluation as a key corporate function; and (4) predictable and adequate resources. Within the UN system more broadly, there are patterns of strengths and weaknesses (JIU, 2014, p. 24) that this evaluation finds frequently correspond with the situation in ILO. For example, significant progress has been made in the mandate for evaluation from the GB and the codification of UNEG norms and standards into ILO policy, and in the continuing challenges with the organizational culture for results, and securing adequate resources. In comparison with the other entities assessed by JIU, however, this evaluation found strong support for the evaluation function amongst ILO senior managers.

While many people see improvements in recent years, there is consistent agreement that the evaluation culture remains relatively fragile across ILO. The evaluation explored the organizational culture both directly (through interviews) and indirectly (through the portfolio analysis and the review of documents). Overall, despite a broad range of views, the ILO evaluation culture was observed to exhibit the following traits:

- The intellectual value as a technical exercise and the administrative necessity of evaluation is largely recognized at all levels of ILO and is not resisted.
- The contribution of evaluation to learning and as a comparative advantage for a knowledge-based organization is not sufficiently recognized throughout the Organization.
- Substantial progress has been made in normalizing independent evaluations as a regular corporate function, including securing RB and other resources for REOs at the regional level in a zero-growth budget context.
- All of the survey respondents (100 per cent) are positive about the commitment of ILO's leadership to supporting useful evaluation, and 75 per cent are positive that evaluation is recognized as of organizational value.
- The senior leadership of ILO and the GB have both demonstrated the political will to implement the recommendations of the IEE (2010) and to support the independence of the evaluation function.
- Some constituents remain unclear as to how evaluation is differentiated from the other accountability functions within ILO. The GB generally considers evaluation in terms of accountability, whereas national-level constituents emphasize learning.

- Across the technical departments there is a general recognition of the value of evaluation, with pockets of innovation and greater openness to learning opportunities.

- At the decentralized level there is a more mixed perception of evaluation, with a predominant view of it as a bureaucratic accountability function that can impact future funding of projects and programmes.

- Technical specialists – a major part of ILO's knowledge function – can be disconnected from the current evaluation culture since their inputs are contributions to multiple lines of activity where accountability, learning and decision-making are diffused (however, some technical specialists have qualified as part of the evaluation manager certification programme and have voluntarily engaged in the evaluation function).

According to interviews, many stakeholders equate evaluations to financial audits that are mainly used for accountability purposes rather than as management tools for promoting learning. This perception is reinforced when the value of evaluations is not clear, i.e. when stakeholders are treated mainly as informants and when the follow-up is limited or not widely known. The use of evaluation frameworks that primarily examine the achievement of the original project results statements and indicators is another factor that contributes to the 'accountability' narrative.

Institutional incentives

Organizationally, incentive structures for staff are not tilted in favour of the evaluation function. This is most clearly reflected in limited recognition of the work carried out by evaluation managers. Evaluation management responsibilities are not acknowledged in job descriptions, and only on an ad hoc basis in the performance appraisal process. It is therefore commonly seen as an 'optional' contribution by both managers and staff. An estimate of the time required for the ILO evaluation management of a single evaluation (IOD PARC, 2013) indicates that this optional contribution can be 5–15 per cent of an individual's annual work plan.

The case studies found examples of senior managers that understand and promote the importance of evaluations, and encourage staff to become involved. In these cases, however, the incentives are created by individuals rather than by the Organization. The same is true for how evaluations are received. Interviews indicate that when evaluation findings and recommendations have been properly responded to and used for project improvements and development, it has been mainly due to the commitment of the individuals involved (e.g. technical specialists, programme staff, chief technical advisers – CTAs).

Compounding the silos of positive evaluation culture in ILO is the weakness of the internal evaluation network. Whilst biennial workshops bringing all members of the evaluation network together for a week, and quarterly video-conferences with REOs alternatingly include departmental evaluation focal points (i.e. twice a year), the network is used primarily as a broadcast mechanism. The evaluation found no evidence of an active peer-to-peer community among the network's members. Facilitating such a network requires substantial time and skills from EVAL, something that is inhibited by the current focus of senior evaluation officers' time on quality assurance of reports. There is also scope for greater inclusion of REOs in global EVAL activities.

RO representatives consider the lack of funding and human capacity as key constraint to expanding the evaluation function. At the same time, it is noted that improving the use of evaluations requires that greater attention be paid to M&E during the project design phase. Many projects have inadequate baselines, indicators, targets and data collection tools and, when this is realized during implementation, it is often too late or, for other reasons, not possible to make changes to the project design.

A common comment is that ILO should move away from inflexible logical frameworks to a ToC approach that provides for a more dynamic M&E function. ToCs allow a more holistic examination of how ILO is contributing to the realization of ILS. This is in line with the recommendation of AERs since 2013–14 for ILO to make having a strong ToC a requirement at all levels of the RBM system. EVAL has made an important contribution in following up with concerned units each year to track the status of follow-up in support of GB governance.

The evaluation found that flagship programmes provide an important opportunity to innovate with ToC approaches, and EVAL is already working with several flagship programmes, such as on social protection floors, to apply them. This will include tracking outcomes at the level of rights holders. EVAL is also planning to develop specific guidance material and tools on ToC, targeted at M&E staff of departments, programmes, offices and projects.

7.3.2 Finding 8: Evaluation structures

Substantial efforts to strengthen the structures of the evaluation function have embedded systematic evaluation in ILO. A noticeable predisposition towards emphasizing accountability over learning has been driven by both the mandatory requirements of the evaluation policy and an organizational reliance on evaluation as substitute for more comprehensive RBM. Much of this imbalance could be addressed by enforcing the recommended 3 per cent budget allocation to monitoring.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	–	–	X	2/4
Constituents	X	–	–	–	
External	X	–	–	–	
– = nil.					

The evaluation function

UNEG's Norm 2 (2005) and Norm 13 (2016) establish that an organization's governing body and its executive head are responsible for ensuring the establishment of a duly independent, competent and adequately resourced evaluation function. To implement this, Standard 1.1 determines that the organization should have an adequate institutional framework for the effective management of its evaluation function.

Overall, analysis of ILO's institutional framework for evaluation by the JIU (2014) rated ILO the second-highest in terms of its maturity, second to the UNDP evaluation function and equal to the WFP evaluation function. Other entities rated as above average were UN Women, United Nations Industrial Development Organization (UNIDO), UNICEF, United Nations Office of Internal Oversight Services (OIOS), and the United Nations Educational, Scientific and Cultural Organization (UNESCO).

The main structural elements of the evaluation function in ILO are: (1) the EAC; (2) EVAL; (3) the head of evaluation; (4) evaluation focal persons; (5) REOs; (6) evaluation managers; (7) M&E officers; and (8) the ILO training centre.

All respondents to the evaluation survey of users consider the organization of the ILO evaluation function to be reasonably clear, with 38 per cent rating it as very clear. However, a third of respondents (all from the regions) were not clear to whom they should refer for clarifications on evaluations. Respondents from the evaluation structure suggest that, from their perspective, scope exists to

improve the availability of evaluation knowledge to support their day-to-day activities (64 per cent) and evaluative knowledge management within ILO overall (71 per cent). The evaluation network is a key structure for further elaborating the details of these requirements.

Within the evaluation structure itself, 29 per cent of survey respondents suggest that their actual activities are not fully in line with their official role. Half of respondents carry out evaluation activities outside of their formal responsibilities. Nevertheless, 93 per cent are positive (and 36 per cent strongly positive) that the evaluation structure supports their professional development and learning.

Evaluation Advisory Committee (EAC)

The EAC is a high level committee that follows up on the management response to HLEs, advises on the selection of topics for these evaluations, and supports organizational learning. The evaluation found that the EAC has been convened regularly and consistently applies its mandate to ensure the implementation of HLE recommendations.

Whilst the EAC functions as intended, and demonstrates a strong organizational commitment to evaluation, committee members and constituents on the GB did identify some opportunities to strengthen its activities. The first of these ideas is to include external members from outside of the ILO to bring new perspectives to EAC discussions and to encourage a more strategic debate. The second proposal, from the GB, was to explicitly rank the topics for HLEs when presenting the list in the AER so as to be transparent regarding the demand that was expressed for each topic (members of the GB are particularly interested in the fate of topics that they propose).

Evaluation Office (EVAL)

There is a general perception among stakeholders that ILO's evaluation function has been strengthened – in terms of capacity and structures – over the last few years. The Evaluation Office is at the centre of this. It is noted that EVAL has been professionalized, achieved a more independent status and increased the standardization of the evaluation process. EVAL is structurally equivalent to the Ethics Office, ILO Administration Tribunal, and Internal Audit and Oversight Office.

Centrally, EVAL has six members of staff including the head, three senior evaluation officers, an administrative assistant and learning coordinator, and a knowledge management specialist. Based on ILO's 2015 staff complement of 2,490 people, this gives a staff ratio of 0.28 evaluation specialists for each ILO staff member. Based on JIU (2014) calculations, this is in line with wider UN practices, which the JIU report assesses to be "under-resourced and over-stretched". The REOs are not staff members of EVAL, although they certainly contribute to the evaluation function.

UNEG's Standard 1.5 (2005) and Standard 2.1 (2016) specify that the head of evaluation, has the primary responsibility for ensuring that UNEG norms and standards for evaluation are upheld, that the evaluation function is fully operational and duly independent, and that evaluation work is conducted according to the highest professional standards. From a future perspective, UNEG (2016) defines the responsibilities of the central evaluation function as follows.

Head of evaluation

As part of the JIU assessment of evaluation in the UN system, a recommendation was made to all entities that heads of evaluation offices should be recruited based on a single non-renewable term limited to between five and seven years with no possibility for re-entry into the organization, and with qualifications and substantial experience in evaluation. ILO is one of eight organizations that

Responsibilities	Current strengths	Current opportunities
An evaluation policy is implemented that adheres to UNEG norms and standards and applies the latest evaluation practices	ILO Evaluation Policy 2005 and Evaluation Strategy 2011	The policy reflects 2005 UNEG norms, not 2016
The governing body/executive head of the organization is adequately informed and advised on the need to review or update the evaluation policy	AER	Insufficient indicators to track the impact of the evaluation function
An evaluation plan is appropriately developed and implemented	Mandated evaluations over $500,000 tracked in *i*-Track	Purpose and timing of evaluations could be more linked to IPC or policy changes
The evaluation budget is efficiently managed	Highly efficient operation	None
Robust and appropriate evaluation methodologies that reflect the highest professional standards are adopted, developed and updated frequently	Standardized measurement of effectiveness linked to RBM	Need for methods specific to ILO normative and tripartite work
Evaluations are conducted in a timely manner and with a focus on intended use for key stakeholders/users	Effective mechanism for enforcing policy using *i*-Track	Credibility has been prioritized over use
Timely and appropriate communication of evaluation results support organizational learning, including publishing evaluation products on the organization's website	High degree of transparency through *i*-Track and other knowledge management tools	Need to socialize learning in future programming
An adequate follow-up mechanism on the implementation of actions committed to within the management response is in place and supported	EAC and management response tracking	Fragile evaluation culture
Evaluation findings, lessons and recommendations are distilled and disseminated as appropriate to enhance learning and organizational improvement	Communications products and meta- syntheses	Under-use of direct dialogue and artificial intelligence

JIU (2014, p. 36) specifically recommends to review the structural independence of how the head of evaluation is appointed.

The aspect of this recommendation relating to independence was referenced by a small number of interviewees from both the GB and ILO staff. Given the evidence available to the evaluation, it finds that whilst there is no theoretical reason to disagree with the proposal, no substantive evidence exists to suggest that the current leadership of EVAL is not fully independent. Indeed, the changes that have been achieved over the course of the evaluation strategy (2011) may not have been so successful without an internal understanding of ILO's systems and culture.

This evaluation notes that the current arrangement for the head of evaluation is the same as that for the head of internal audit. It also notes a low level of empirical evidence from other UN entities on the effectiveness and implications of implementing this recommendation, and the necessity for a new administrative option to implement the suggested conditions. Given these observations, and considering the other findings, the evaluation proposes that an incremental approach to such structural changes would be the most expeditious approach, with the appoint of the next head of evaluation being subject to input from the GB, as is the case for the United Nations Office on Drugs and Crime (UNODC), the World Meteorological Organization (WMO), and UN Women.

Evaluation focal persons (DEFPs)

Each HQ department has a person assigned as DEFP, responsible for coordinating mandated evaluation activities with EVAL. A DEFP may have additional M&E activities within their main role, in which case they may be more familiar with concepts of evaluation. However, this is not necessarily the case and some DEFPs acknowledge that they lack knowledge of the basic tenets of evaluation.

A network of evaluation focal points (including REOs) established by EVAL meets on a biennial basis, and an impact evaluation network of interested colleagues has recently be convened to support high-quality impact evaluations. The main evaluation network is appreciated by focal persons, but is not deemed to be active enough to ensure cross-regional and cross-thematic learning. The network was largely conceived as a channel to distribute evaluation information, and execute roles and responsibilities with regard to the evaluation policy.

The evaluation found a lot of scope for the evaluation network to be more actively facilitated – putting the emphasis on more regular meetings, and nurturing a stronger sense of identity (belonging to an evaluation community), and exchange between DEFPs and REOs. These aspects are required to support self-led professional development on evaluation and to avoid reliance on only formal training.

Regional evaluation officers (REOs)

Administratively, REOs are M&E specialists employed directly by regional offices. Whilst they are not members of EVAL staff, they work full-time on M&E and related issues under the technical guidance of EVAL. REOs manage, coordinate and support the implementation of mandatory evaluations in the regions and are often the first point of contact for evaluation issues.

Their work is highly praised in ROs. Often being the only dedicated M&E person at the regional level they play a critical role in each step of the evaluation process and deal with all evaluation-related procedures. These include: preparing evaluation ToRs; supporting the process of identifying and selecting evaluation consultants; arranging missions and meeting schedules; reviewing, circulating and commenting on inception reports and evaluation reports; consolidating feedback; and organizing debriefings. Additionally, in some regions, they are involved in appraising new projects, monitoring DWCPs, and developing new regional strategies.

REOs' professional level and type of contract varies across the regions, in some cases with low levels of job security. There is also a single REO per region, despite a large variation in the number of evaluations being undertaken. Africa region previously had four M&E officers at the subregional level to support the REO role, but this is no longer the case. The Asia Pacific region is planning to recruit a national M&E officer in 2017 to support the REO.

There is broad interpretation of the job description among REOs. Some focus primarily on coordinating or managing mandated evaluations, recruiting and overseeing evaluation managers, and briefing evaluators. Others are more engaged in developing proposals or evaluation capacity.

Since EVAL views the REOs as its extended arm in the regions, it is important that they have the opportunity to take part in and be regularly informed about developments at HQ, including in relation to partnerships being forged between departments, new tools being developed, and discussions about the SDGs and evaluation. A natural progression would be to formalize the network by developing a ToR outlining its purpose and roles, and making it accountable for achieving some results.

The evaluation compared the REO system to other agencies of the UN system at a similar level of maturity in the JIU rankings. It finds that the current arrangement is insufficient to support the needs of the evaluation function:

■ Whilst it is recognized as important for the evaluation function to have a presence in the regions, the mixed M&E mandates of the REOs mean that they are not independent from management. This is a key weakness in the overall independence of the function. There is a precedent in UN Women – a smaller entity than ILO – for independent regional evaluation specialists under the direct management (and budget) of the evaluation office.

- The inconsistency in the level and type of contracts for REOs is a barrier to ensuring consistent application of evaluation policy across regions. It is also a barrier to the development of a community of practice between evaluation staff in the regions and HQ.

- The mixed M&E responsibilities of REOs are insufficient to address the overall shortfalls in availability of monitoring expertise, whilst having a significant negative impact on the time available for evaluation activities.

- The exclusion of REOs from the management responsibilities of EVAL means that accountability for evaluation report quality (and updating *i*-Track) cannot be adequately decentralized. Consequently, accountability is placed on senior evaluation officers and other EVAL staff, leading to inefficient use of their time and expertise in enforcing EVAL's checklists and guidance.

Evaluation managers

The evaluation manager system is somewhat unique to ILO and is one of the differentiating factors (along with quality assurance by EVAL) of 'independent' evaluations (as compared to internal evaluations). Whereas evaluations might be managed in other UN agencies by M&E staff in the same office as the programme being evaluated, the ILO evaluation manager system requires that a member of staff be appointed from a separate part of the Organization. One of the main responsibilities of REOs and DEFPs is to find and support the evaluation managers in decentralized independent evaluations.

Whilst the independence of evaluation managers is the aspect that is most emphasized in reports, the evaluation interviewees suggest that the primary need for the evaluation manager system is an inadequate number of M&E specialists in country and regional offices to manage the volume of evaluations mandated by the evaluation policy.

To support the performance of evaluation managers, EVAL and the International Training Centre of the ILO (ITCILO) in Turin developed and offer the Evaluation Manager Certification Programme (EMCP). This is a unique innovation within the UN system.

The evaluation manager is an optional position (REOs have no power to force the appointment of an evaluation manager). Interviews suggest that it can be difficult to convince staff members to enrol in the EMCP and, especially, to take on the role of evaluation manager beyond that which is required for certification.

The evaluation found that there are no sustained incentives for assuming the role of evaluation manager. The responsibility is not included in job descriptions and not always documented in the performance appraisal process. Similarly, the time and effort staff members spend on collecting information for evaluations and servicing evaluation consultants (estimated at 10–30 days per evaluation) is not formally recognized.

Trained evaluation managers acknowledged that the achievement of a certification and the experience of managing one or two evaluations are sufficient to meet their professional objective of including evaluation on their CV. Participants indicated that the lack of continuous evaluation practice led to much of the training course being forgotten.

Evaluation managers also report that they are often missing the very basics of evaluation theory and need refresher training because of infrequent practice. Interviewees in the case studies suggest that more training on evaluation methodologies and tools would be more welcome than training on evaluation management. However, this contradicts the low uptake by ILO staff of pilot training on conducting internal evaluations that has also been offered by EVAL and the ITCILO.

The evaluation finds that EVAL has strongly emphasized relevant evaluation capacity development activities in accordance with the reconstructed ToC developed for this evaluation. Despite these

efforts, an independent assessment of the impact of the EMCP in 2015 found no statistically significant difference between the average quality of evaluations produced under managers who had completed the EMCP versus those who had not completed the EMCP (this assessment included the practicum that was undertaken as part of the training).

Given that the EMCP is innovative, this evaluation does not find this assessment to be negative, but it is an opportunity for learning. It would, therefore, be relevant to expect to see the approach to developing evaluation management capacity evolve in a future evaluation strategy. The case studies indicate that the infrequent practice of evaluation management – rather than the EMCP itself – may be a more important barrier to improved evaluation quality.

As noted above, the level of practice of evaluation manager is driven by organizational incentives, most of which are not under the control of EVAL. For instance, the inclusion of evaluation management in ToRs, work plans, and appraisals requires a corporate response. It was the view of interviewees in the case studies that the current arrangements – and incentive structures – mean that the existing approach to evaluation management is unlikely to be sustainable.

Despite these challenges, the evaluation notes that EVAL and REOs have identified some more promising incentives for encouraging greater participation in the evaluation manager system. These include negotiating with projects and programmes to allow for evaluation managers to travel to evaluation case studies as an opportunity for peer exchange (something that would be facilitated by early advertising of upcoming opportunities so that staff can identify evaluations relevant to their own work and writing of proposals). There is also greater scope to integrate evaluation managers into the EVAL network/community-of-practice to promote peer-to-peer support.

The wider evaluation system: M&E officers

As noted throughout the evaluation, despite ILO having a formal RBM system and the DWCP process in place, M&E officers in departments, field offices and projects are few in number and work on isolated areas of ILO's portfolio. However, a strong RBM system underpins both the quality of evaluation, and the absence of good monitoring data is driving many of the demands for ILO to intervene in areas that are not central to its mandate. M&E officers are, therefore, a central consideration in the overall limits of the evaluation function.

The wider evaluation system: International Training Centre of the ILO (ITCILO)

The evaluation notes that there is an extensive collaboration between EVAL and the ITCILO. This encompasses work on the production of evaluation documents, two training courses and certifications, and the development of guidelines for evaluation consultants. Whilst the evaluation did not examine these activities in depth, it acknowledges EVAL's commitment to supporting synergies with ILO's other functions.

The wider evaluation system: Gender, Equality and Diversity Branch (GED)

EVAL has collaborated closely with GED in both coordinating independent evaluations of implementation of the ILO Action Plan for Gender Equality, and in mainstreaming gender further into EVAL evaluation guidance in accordance with the requirements of the ILO Action Plan for Gender Equality, which is aligned with the UN SWAP. Interviewees indicate that EVAL is considered to be an important and valuable champion for gender mainstreaming within the ILO, and GED's collaboration with EVAL is considered to also be an important contribution to the evaluation function.

7.3.3 Finding 9: Planning and reporting arrangements

EVAL has established and systematically maintained relevant and effective reporting arrangements on the outputs of the evaluation function to the Director General (DG), Evaluation Advisory Committee (EAC), and the Governing Body. These have, however, been insufficient to fully address strategic considerations that have arisen since the last IEE, including the structural under-resourcing of evaluation.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	–	–	X	2/4
Constituents	X	–	–	–	
External	–	–	–	–	
– = nil.					

Reporting to the Governing Body

Reporting requirements for central evaluation functions are more explicitly covered by JIU recommendations 3 and 6 (2014), rather than by UNEG norms and standards (2005 or 2016), which both offer Standard 1.3: "evaluations should have a mechanism to inform the governing body and/or management on the evaluation plan and on the progress made in plan implementation."

- JIU Recommendation 3: The legislative bodies should request the executive heads of United Nations system organizations to develop comprehensive budget frameworks and resource allocation plans for their respective evaluation functions, based on the cost of maintaining an effective and sustainable evaluation function that adds value to the organization. The plans should be submitted for consideration to the legislative bodies within existing budgetary and reporting mechanisms and processes.

- JIU Recommendation 6: The executive heads of United Nations system organizations should make the use of evaluation an organizational priority based on a well-defined vision, strategy and results framework for the evaluation function, and report to their legislative bodies on the level, nature and impact of use of evaluation.

Regarding JIU recommendation 6, the evaluation finds that the AER prepared by EVAL and presented to the GB is a major contribution to fulfilling the specifications and the intention of this requirement. The AER certainly fulfils the requirement to systematically report on the level and nature of evaluation (including the acceptance of recommendations in management responses); and on the evaluation strategy and results framework.

The interviews did identify scope to further enhance the contribution of the AER in support of the vision of a learning organization. These include: (1) establishing and tracking, within the results framework of the next evaluation strategy, indicators of the impact of the evaluation function beyond the monitoring of management responses; (2) reporting back to the GB on changes made due to their comments in subsequent reports; and (3) publishing a clear vision and mission for the evaluation function.

It is noted that these are future challenges for EVAL, since they are universally considered weak aspects of UN evaluation systems (JIU, 2014, p. 42), and were not considered in the 2005 UNEG standards. ILO will, in several cases, need to pioneer new approaches to reporting on the impact of evaluation.

Evaluation planning

Evaluation planning takes place at several levels in ILO:

- The GB approves the topics for three HLEs per year based on recommendations from the EAC presented in the AER. The three-year rolling work plan includes one geographic (DWCP) and one outcome evaluation on a rotational basis, and one institutional topic selected through consultations. EVAL manages this work plan. The budget for each evaluation is based on guidelines set by the GB.

- All mandated decentralized evaluations (independent and internal) with project budgets over $500,000 are tracked in *i*-Track by EVAL's knowledge management officer based on PAR-DEV minutes. They are coordinated by REOs and DEFPs and overseen by EVAL's senior evaluation officers (in the case of independent evaluations). EVAL enforces a mandatory 2 per cent budget allocation for evaluation during project formulation.

- Non-mandated evaluations of DWCP and RBSA activities are included in RBSA M&E plans, REO's own management tools (such as Excel sheets), and communicated to EVAL for inclusion in *i*-Track. The budget for M&E is negotiated as part of the planning process.

There is not a single consolidated planning process for evaluation at regional or global levels since the majority of evaluations are mandatory, and there are only a few additional evaluations. Information on global evaluations with components in the regions is not shared with ROs in a systematic manner. This exacerbates 'over-evaluation' in countries subject to multiple project evaluations, DWCP evaluations, and global evaluations in a concentrated period. Generally, there appears to be a need for better coordination between centralized and decentralized evaluations, scope to negotiate with donors that TC projects should be included as case studies in larger evaluations, and more opportunities for sharing experience between REOs and DEFPs.

Evaluation resourcing

A set of estimates agreed by the evaluation team and EVAL regarding the central evaluation function (see figure 6) are based on RB (staff assigned to EVAL at HQ, HLEs, meta-studies, meetings, training), RBSA (a top-up mostly for non-project related evaluations in the regions[17] and contribution to the salaries of the REOs), and programme support income (allocated by the D-G to support the effective delivery of extra-budgetary projects and programmes). These estimates reveal that annual expenditure was $2.4 million in 2011, increasing to $2.8 million in 2016. Only around half of this ($1.5 million) is part of RB, the rest fluctuates based on regional work and TC.

The majority of evaluations are mandatory, and are thus funded from the 2 per cent allocation set aside in TC project budgets. This is extra-budgetary resource, and does not appear in ILO's P&B plan. It is known from the interviews that the 2 per cent allocation within projects is used flexibly (sometimes redistributed for large projects), most often with the consent of EVAL. ILO's finance system does not have a line item for evaluation, and thus it is not possible to track the exact amount spent on decentralized evaluation by the Organization. However, based on the average annual number of decentralized independent (45) and internal (34) evaluations (see Finding 1), and estimated average costs (based on interviews) of $30,000 per independent and $10,000 per internal evaluation, the evaluation finds that annual expenditure on TC project evaluations is likely to be up to $1.69 million.

The cumulative estimate of expenditure on evaluation over the period 2011–2016 is $26 million, equivalent to $4.3 million per year (figure 8). Based on the biennial budget for ILO 2014–2015

---

[17] Such as internal evaluations, baseline surveys, country programme reviews, monitoring missions, diagnostic tool development, thematic evaluations, M&E capacity building support, and support to implementation of HLEs.

($400.6 million per year), evaluation represents 0.8 per cent of total budgeted expenditure (i.e. excluding extra budgetary resources such as TC projects). Based on the recommendations of JIU (2014), and given that ILO is a medium-sized UN agency in JIU classifications, this evaluation would expect to see expenditure in the range of 1.5–2 per cent of the total organizational budget being allocated to evaluation. ILO is currently committing only half of this amount. This finding explains the general perception among interviewees that there are inadequate resources and staff capacity to manage evaluations and follow-up on their outcomes.[18]

Figure 9. Tracking of central evaluation funds allocations 2006–2016
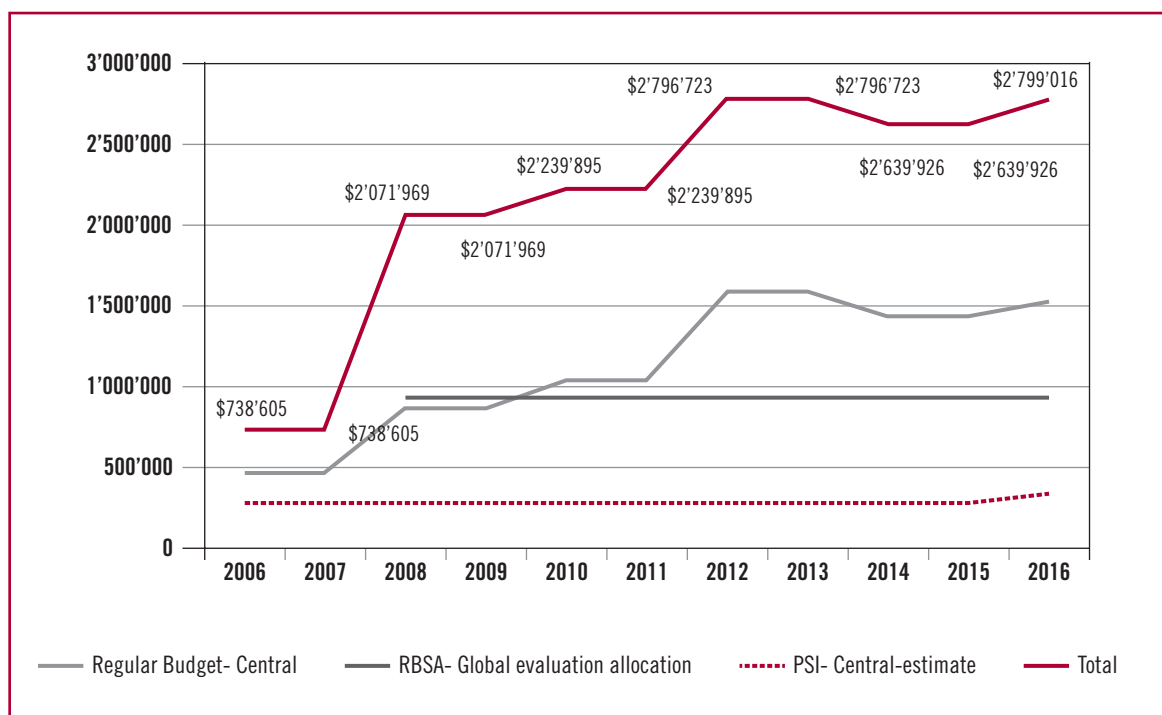
Evaluation budgeting

At the project level, the case studies and portfolio analyses indicate that the budgets for evaluations have not been linked to specific appraisals of the intended purpose and required quality. In the period covered by this evaluation, EVAL and PARDEV have worked closely to implement the evaluation policy that requires a fixed allocation within project proposals of 2 per cent (the evaluation notes that 2 per cent is an appropriate amount in global terms given JIU's 2014 recommended allocations, and that provision for real-cost estimates is included in the latest version of the ILO's internal development cooperation governance manual).

However, the evaluation also found that use of a fixed percentage rate has led to many evaluations that are considered by users to be either under- or over-resourced given their purpose and objectives. For projects of $1 million, a budget of $20,000 is insufficient to cover the required independent and internal evaluations (mid-term and final); for projects of $20 million, $400,000 is seen as an excessive amount to spend on two independent evaluations (mid-term and final).

---

[18] This observation runs counter to the results of a survey of evaluation managers undertaken as part of the independent quality assessment of ILO reports (2015), in which evaluation managers were of the view that financial resources were sufficient, but time was constrained. The views reported in this evaluation are drawn from a much wider range of stakeholders.

Evaluation users suggest that the scope of ILO's evaluation practices is not only constrained by limited resources, but also by the requirement to spend the funds within the lifespan of a project or programme. Some evaluations would benefit from being performed after the formal close of the relevant projects and programmes, such as impact evaluations. Currently, there is no pooled funding mechanism for selected evaluation exercises to assess the sustainability or impact of the results.

Suggestions from ILO staff on how to address these anomalies include:

- allocating a range of fixed costs for project evaluations rather than a percentage (for example $30,000–100,000 per evaluation depending on project size);

- using some of the excess budget in large projects to pay for more thematic meta-analyses and the communication of evaluation insights to management;

- establishing a pooled funding mechanism for TC projects to contribute to thematic, joint or DWCP evaluations based on project case studies (an arrangement that would require donors to accept that resources could be transferred to the pooled fund as committed expenditure); and

- combining evaluation budgets from TC projects to fund additional M&E officers to support REOs and deliver evaluability, RBM and evaluation management across multiple projects.

Whilst it would be useful for EVAL to provide guidance on the minimum viable costs of project evaluations, this evaluation finds reasons to be cautious and iterative in replacing the current 2 per cent allocation with more appropriate results-based budgeting approaches (despite the high demand from management to do so quickly). Given that the evaluation function is structurally underfunded and that TC projects account for at least half of evaluation expenditure, any action that reduces the overall resourcing of evaluation will have significant negative implications on overall expenditure. Current RB and RBSA allocations are not sufficient to cover the evaluation needs of different organizational entities, including for DWCP evaluations and quality assurance impact evaluations.[19]

The evaluation has identified three necessary structures that are not currently present in ILO's evaluation planning and budgeting system: (1) a mechanism to pool evaluation resources from multiple projects and programmes (either through a pooled fund, or through advanced planning systems for commissioning multi-project evaluations); (2) a comprehensive monitoring, evaluation, research and learning plan at each level of the Organization that is linked to organizational budgeting and financial tracking systems; and (3) a means of ensuring sufficient TC project resources for monitoring so as to better meet donor accountability requirements and, thus, enable more flexibility in evaluation policy.

At present, the evaluation finds that the EAC and the AER are focusing primarily on reporting on the delivery of evaluation outputs and the EVAL work plan. They are insufficient mechanisms for reporting to the GB and the DG on strategic aspects of the evaluation function. A strategic scope was, however, granted to this and the previous independent evaluations of the evaluation function. Thus, this evaluation confirms the conclusion of the previous IEE (2010) that periodic evaluations of the evaluation function are, therefore, necessary. It notes, however, that most UNEG members achieve this through an established peer review function. Given the commitment of ILO to Delivering as One, the evaluation has not found any evidence to refute the view that a future IEE could be achieved through the UNEG peer review mechanism.

---

[19]   Impact evaluations are normally funded as substantive activities in projects under "knowledge building" components of projects and by impact evaluations of specific projects. Including such evaluations – particularly with an ex-post and ILO intervention focus – as part of integrated evaluation, research and learning planning might be a useful consideration.

7.4 TRENDS IN EVALUATION CULTURE AND PRACTICE

7.4.1 Finding 10: Institutional framework

ILO has successfully established the technical practice of evaluation through a systematic institutional framework that is aligned with UNEG norms and standards. Within the context of the 2030 Agenda, significant scope exists for the evaluation function to be more flexible, learning-oriented and outward-looking in the future.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	3/4
Constituents	X	X	–	–	
External	X	X	–	X	
– = nil.					

At the global level, EVAL participates in UNEG and other professional evaluation networks. The evaluation surveyed existing members of the evaluation function in ILO, including DEFPs, REOs, evaluation managers, and EVAL staff. Forty-two people responded. Of these, 29 per cent were insufficiently knowledgeable of the UNEG norms and standards to take a view on the extent to which the current evaluation function is implemented in accordance with these requirements. This highlights one implication of the reliance of the evaluation function in ILO on non-specialists.

Survey respondents noted the absence of a position in field offices that has the responsibility for evaluation (in comparison with other operational UN organizations in which M&E officer positions are an integral part of decentralized office structures) as a barrier to delivering the UNEG standards. This creates challenges when ILO is evaluating its component of a joint programme (as mandated by the EVAL guidance note). The evaluation noted a complex relationship between the current evaluation policy, evaluation guidance and joint programmes, with country offices looking for more explicit guidance on how to integrate their responsibilities with jointly commissioned evaluations.

2030 Agenda

Since the evaluation policy (2005) and evaluation strategy (2011) were established, a major change in the normative framework for ILO has been the agreement of the SDGs and the commitment to 2030 Agenda. The revised UNEG Norm 1 (2016) explicitly requires UN evaluation functions to respect, promote and contribute to the 2030 Agenda. A new Norm 9, meanwhile, emphasizes that national evaluation capacities should be supported when requested by member States. Finally, JIU's Recommendation 9 (2014) requires that the institutional framework for evaluations (especially decentralized evaluations) should support the UN system to respond to the 2030 Agenda in an integrated way.

It is commonly acknowledged among interviewees that ILO will increasingly have to consider the impact of its work in the context of the SDGs. It is the view of several ILO staff that the 10 policy outcome areas found in the ILO's results framework for 2016–17 are not currently sufficient to demonstrate ILO's contribution to the SDG indicators through changes for ultimate beneficiaries. EVAL has already responded to this concern through the preparation of an *i*-eval Think Piece on *Applying evaluative thinking towards an effective ILO contribution to the implementation of the SDGs* (December 2016), and an internal paper on *The implications of the SDGs on ILO's results framework* (June 2016).

It is felt by interviewees at regional level that M&E systems overall are not geared towards capturing impact and are generally weak in terms of indicators, targets and data collection methods. M&E tends to be looked at only towards the end of the project design process, leaving little time for baseline studies to be conducted. Furthermore, the short time duration of TC projects (commonly 12, 18 or 24 months) is also seen as an inhibiting factor. In view of this, evaluation of outcomes over longer timeframes – through the DWCPs and thematic evaluations – takes on an increasingly important role.

National constituents in the case studies were found to be largely disengaged from the M&E of ILO projects and programmes, not always seeing the value of spending time and resources on evaluations with a comparatively narrow scope. In most evaluations, their participation is limited to making themselves available for, or helping to organize, interviews or presentations. The evaluation did find evidence of offices attempting to address this. For example, as part of the process of developing new DWCPs, the Regional Office for Arab States is currently planning to organize capacity building workshops on M&E for national constituents in Jordan, Lebanon, the Occupied Palestinian Territory and Oman.

The evaluation agrees with the EVAL analysis (June 2016) that some adaptations to the existing practice in ILO will be necessary in view of the 2030 Agenda, and emphasizes the following three priorities:

- At global and national levels, ILO will need to advocate for the inclusion of its social partners in the UN system concept of evaluation capacity development, since it is the only UN entity to work through a tripartite mechanism.

- Ensuring that no one is left behind requires more inclusive evaluations, which implies directly supporting national constituents in collecting, retrieving and analysing data from a wider range of people.

- Evaluating ILO's contribution to the SDGs implies longer term evaluations. This fits with the demand from social partners for evaluations that give a broader view of what is working and what barriers are being faced in advancing the realization of ILS. Project evaluation alone cannot deliver on either of these demands.

Quality assurance of evaluation

Whereas UNEG's standards 4.1–4.18 (2005) specified requirements for individual reports, UNEG introduced Standard 5.1 (2016), which calls for the establishment of a quality assurance system, and Standards 5.2 and 5.3, which respectively call for quality control at the design phase and final stage. EVAL has established a quality assurance system for independent evaluations that is multi-layered, methodical, rigorous, and consistent with UNEG requirements.

Evaluation managers supported by REOs and evaluation focal persons are the first level of quality assurance. Senior evaluation officers in EVAL, however, are also accountable for the quality of decentralized independent evaluations under the areas that they cover (partly because REOs are not full EVAL staff). This results in multiple exchanges for each stage of the evaluation between the different actors in the evaluation function to ensure that ILO evaluation guidance is fully applied. Finally, the published evaluation reports are subject to an independent quality assessment on a biennial basis.

These quality assurance systems are based on UNEG norms and standards for evaluation processes and products. However, this evaluation found that very often the intended users of the evaluations, especially the ILO technical specialists, misunderstand evaluation standards and can develop expectations that are too high in terms of the depth of the technical analysis and the specificity of related recommendations. Generally, there seems to be a lack of common understanding on the purpose

and limitations of evaluations, and the distinction between evaluations, research and programming missions. Whilst this evaluation heard opinions about evaluation quality that were sometimes based on such misunderstandings, the quality assessment undertaken in the portfolio analysis generally confirmed EVAL's own analysis of compliance with UNEG quality standards.

Apart from the expectations of some users, the evaluation found that the quality assurance system has technical strengths, but also several opportunities for improvement (Table 10).

Table 10. Strengths of ILO's evaluation quality assurance and opportunities for further improvement

Strength	Opportunities
Biennial independent quality assessment of independent evaluations	Other UN system entities (UNFPA, UNICEF, UN Women, WFP) have long-term agreements for continuous external assessments, allowing for annual reporting to their executives and on UN SWAP
Senior evaluation officers bring expertise to the quality control role	The large number of independent evaluations creates a tendency for implementing standards to become mechanical, and may not be the best use of senior expertise
High level of consistency in the application of ILO and UNEG standards to independent evaluations	No similar quality assurance or independent quality assessment of internal and self-evaluations

Human rights and gender equality

In 2016, UNEG introduced a new Norm 8 that requires the values and principles of human rights and gender equality be respected, addressed and promoted by the evaluation function. Meanwhile, Standard 4.7 requires that evaluations examine the application of a human rights-based approach and gender mainstreaming strategy in the object of the evaluation. Finally, ILO is required under the UN System Wide Action Plan on Gender Equality and the Empowerment of Women to report on the four criteria established by UNEG for the Evaluation Performance Indicator (UN-SWAP EPI).

As noted earlier, based on the ILO Action Plan for Gender Equality, EVAL has collaborated closely with GED to further mainstream gender into its evaluation products, guidelines and requirements. GED considers EVAL as a visible champion for gender mainstreaming and for introducing good practices that are in some ways ahead of its UN evaluation peers. As a result of this work, EVAL now requires that all evaluation ToR include, as a minimum, a gender-mainstreaming clause to support this crosscutting theme. In addition, evaluation managers must now confirm, in a formal submission, that the quality checklist for evaluations has been applied. In the survey of users, 88 per cent of respondents consider that human rights and gender equality are now mainstreamed into the evaluation process.

Despite these initiatives, the independent quality assessment 2014–2015 found that only 21 per cent of reports sufficiently disaggregate data by sex, and only 14 per cent disaggregate by other target groups. Discussion of human rights and ethics was rated more highly, with 55 per cent of reports meeting the required standard. The IEE observed significant interest from some of the flagships to extend their evaluative analysis to the level of individual beneficiaries.

The portfolio analysis for this evaluation found that 62 per cent of ToR adequately included reference to ILO's gender-mainstreaming commitments (11 per cent did not mention gender). The main areas of weakness in the reports were insufficient inclusion of gender in criteria, questions and indicators (85 per cent of reports did not meet UN SWAP standards) and in methods that were not gender-responsive (80 per cent did not meet UN SWAP standards). By comparison, the inclusion of gender analysis in findings and conclusions was a little higher (30 per cent of reports did meet UN SWAP standards).

These data suggest that, despite specific guidance notes on the topic, the strong commitment to gender equality in EVAL is not fully translating into ToR, and gender mainstreaming in terms of reference is not fully translating into evaluation practice (and reports). Specific reasons for this trend were not identified. However, sufficient evidence was gained from interviewees to make two hypotheses that could be tested by EVAL.

- Hypothesis 1: The link between human rights, gender equality and ILO's core value of social dialogue has not been sufficiently made in the practice of evaluations. Methods used in evaluations (partly because of time and resource constraints) are not well suited to social dialogue, and are, therefore, not maximizing ILO's inherent potential to be responsive to gender and human rights.

- Hypothesis 2: Projects and programmes, and their evaluation, focus on institutions rather than people. Whilst some areas – such as IPEC and the social protection flagship – have included indicator changes in the lives of ultimate beneficiaries in their results frameworks,[20] most results frameworks are at policy and institutional levels. Since evaluations are using results frameworks as their departure point, they also restrict themselves to an institutional perspective in these cases.

Networks

EVAL has an internal evaluation network made up of regional M&E officers and DEFPs. As previously noted, this network meets on a biennial basis, and in between these meetings receives updates through the EVAL newsletter and quarterly conference calls. Whilst EVAL is piloting new communications strategies with the network (such graphical guides to evaluation processes), the evaluation found that there is strong demand for EVAL facilitation to promote peer-to-peer dialogue between colleagues, develop a stronger sense of evaluation identity, and to transform the network into a community of practice.

Externally, EVAL is a member of UNEG, the International Development Evaluation Association (IDEAS), and the Geneva Evaluation Network. Through these memberships, ILO participates in UNEG practice exchanges and has served on the management committee for the pilot UN Independent System-Wide Evaluation that examined efforts by the UN system to build national statistical capacities. The head of EVAL was co-chair for the Working Group on National Evaluation Capacity Development in 2012–2013, which led to a UNEG guideline on this topic.

Whilst the evaluation acknowledges the strong efforts of a small team to engage in these processes, it also notes that EVAL does not regularly participate in international evaluation conferences or networks of the Voluntary Organizations for Professional Evaluation (VOPE). This limits its ability to nurture a community of evaluators with interest, knowledge and experience of evaluating decent work, normative interventions and social dialogue.

Furthermore, given ILO's substantive understanding of 'professionalization' as well as unique experience in social dialogue – both of which are processes that are underway within the development evaluation community itself – this evaluation finds that there is a strong case for EVAL to be resourced to undertake more outward looking activities. Advocating for greater understanding and capacity for decent work and social dialogue within the development evaluation community (as highlighted in EVAL's June 2016 paper on the SDGs) could potentially be conceived and supported as a programmatic intervention that is complementary to evaluation of ILO's own work on which the evaluation strategy is currently focused.

---

[20]   The evaluation notes that some programmes in the past, e.g. IPEC, extensively reported on direct recipients and beneficiaries, including impact studies (retrospective longitudinal tracer studies) to document changes in the lives of ultimate beneficiaries.

7.4.2 Finding 11: Competencies

Structural under-resourcing of the evaluation function is the main limit on the extent to which capacities, competencies and methods for evaluation can be expanded beyond the significant achievements made since 2011.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	–	X	2/4
Constituents	–	X	–	–	
External	–	X	–	X	
– = nil.					

UNEG's Norm 9 (2005) and Norm 10 (2016) require that members of the evaluation function conduct their work with professionalism and integrity. UNEG's Standard 2.1 (2005) and Standard 3.1 (2016) establish detailed lists of competencies for evaluation. These are in addition to wider UNEG guidance on human rights and ethics that also outlines the expected conduct of the evaluation function. UNEG guidance is complemented by the work of other networks to which EVAL belongs, including the IDEAS evaluator competencies and ongoing working group on professionalization.

The survey of the evaluation architecture found that 90 per cent of respondents believe that the current evaluation capacities, methods and tools applied in ILO are adequate in meeting needs. However, 66 per cent were also of the view that there remains scope for improvement, particularly with regard to: the SDGs, coping with large numbers of evaluations in Africa and the Asia-Pacific regions; strengthening the internal evaluation network; mainstreaming evaluation and communications in field offices; improving knowledge of quantitative methods; simplifying project evaluation processes; and managing follow-up to management responses. The most frequently mentioned issue was the need for simplification of mandatory evaluation processes.

Regarding training in evaluation, 50 per cent of survey respondents have received evaluation training (mostly through the EMCP) within the past year. Of those who have not received training, half would be interested in participating in future opportunities. Since 2013, 84 ILO officers have participated and 43 have successfully graduated from the programme. In addition, in the period 2011–2016, 430 ILO staff and 838 constituents were supported with evaluation trainings.

There is a general perception among stakeholders in the regional case studies that ILO's evaluation function has been strengthened – in terms of capacity and structures – over the last few years and that a remarkable change has occurred from 2010. It is noted that EVAL has been professionalized and that the capacity for managing evaluations has been strengthened across the Organization through the EMCP. There is also widespread recognition of the increasing standardization and systematization of the evaluation process, which has been achieved through the development and implementation of guidelines for each step of the evaluation process.

However, there is also a general feeling that internal capacities still need to be improved and that the organization of the evaluation function within the region needs stronger support in terms of human resources. REOs are currently the only staff members overseeing the evaluation function in the regions, supported by project focal points for the implementation of evaluations (organization of the field visits, identification of stakeholders, collection of documents, etc). M&E specialists appointed within specific projects do not contribute to wider regional evaluation activities, and probably should not beyond being part of a community of practice. Constraints in human resources capacity ultimately limit the impact that training can have on improving the evaluation function.

Evaluation teams

UNEG's standards 3.13–3.15 (2005) and Standard 4.8 (2016) require that an evaluation team should be selected through an open and transparent process, taking into account the required competencies, diversity in perspectives and accessibility to the local population. The core members of the team should be experienced evaluators.

The case studies found that evaluation consultants are generally considered by ILO staff to possess the required competences and experience. However, a remark repeated by several interviewees is that, occasionally, consultants lack understanding of the ILO, including its unique tripartite constituency structure, and that this unfamiliarity may affect the quality of evaluation findings and recommendations. EVAL is in the process of launching a consultant orientation course in an attempt to address this.

Generally, as with evaluation reports, there are high expectations on evaluation consultants in terms of technical expertise, regional/country-level knowledge, evaluation experience, and language skills. The number of consultants who possess this combination of competences and experience, and are willing to accept ILO's fee rates, are seen by REOs and DEFPs to be limited and, as noted previously, tend to be re-hired. Globally, EVAL currently has a roster of 350 registered consultants who have previously undertaken an evaluation for ILO. However, not all of these are still active and available, and the need for specific technical experience narrows the field.

Whilst this situation creates the risk of a symbiotic relationship between ILO and evaluation consultants (potentially affecting the overall independence of the evaluation function as well as the quality of reports) the evaluation found no clear evidence of such bias in ILO's current portfolio of independent evaluations (internal evaluations were not assessed). Indeed, according to JIU (2014, p. 78) data, ILO used 85 different evaluation consultants in a sample biennium (2012/13), at least double the number of consultants in other UN agencies (such as UNFPA).

The mitigation of potential for co-dependence between ILO and its evaluation consultants is a notable achievement. This is not always the case in the independent reviews of other UN agencies, suggesting that the quality-assurance process established by EVAL is acting as a guarantee of impartiality in decentralized evaluations.

An individual evaluator is currently undertaking most evaluations. In the experience of ILO staff, there are few cases where a single person has both the evaluation and technical knowledge that is required to prepare recommendations that are pragmatic and feasible within the human resources available to ILO. This is a significant limitation associated with not using evaluation teams recruited either as a group of individuals (which is ILO's current practice) or through the procurement of services from specialist firms.

Responsibility for the selection of consultants rests with the evaluation manager. The independent quality assessment of evaluations (2014–2015) also undertook a survey of ILO evaluation managers. Survey respondents indicated that: "they frequently or always refer to the ILO evaluation standards and share them with evaluation consultants. When selecting evaluators for projects, the dominant criteria selected by managers are competency, knowledge of ILO, and familiarity with the country and context."

This recruitment process may include the posting of the evaluation ToR on regional and global sites and/or a search of the consultancy roster maintained by the REOs or EVAL. Typically, a shortlist of candidates is prepared from which the selection is made against certain criteria, but this is not always the case. The CV of the selected candidate is shared with the chief of the programme unit and with EVAL and, if there are no objections, an offer is made to the consultant who is then contracted by the RO. By comparison with other UN entitles, this process is considered to be appropriate and sufficient.

One case study also examined the rates being offered by ILO in that region. It found there is a $400 a day ceiling on rates before the approval of the regional director is required, and there were no cases where it was recalled that an evaluator had been paid over $700 a day. This implies that ILO is seeking niche expertise whilst recruiting from the lower middle range of the consulting market – a challenging proposition to sustain. Apart from raising the thresholds for daily rates, other UN entities seek to address the question of fees by: (1) using fixed rate individual contracts and decentralizing authorization up to a total threshold, such as $30,000; (2) engaging in long-term framework agreements with firms that include fee tables based on levels of expertise; and (3) more frequent use of evaluation service procurement from firms than is currently practiced by ILO.

Designs and methods used by individual evaluations

UNEG's Standard 3.7 (2005) and Standard 4.5 (2016) specify that evaluation methodologies must be sufficiently rigorous. Most independent project evaluations in ILO apply similar qualitative designs, with mixed methods being adopted where a source of quantitative data is available within the time and resource limitations of the evaluation. This is characteristic of evaluation within the UN system more broadly. It also reflects the normative nature of ILO's work and emphasis on consultation with social partners.

The choice and appropriateness of evaluation methodology was an issue raised by several interviewees. Among some users there is an impression that evaluations are unduly based on interviews and similar qualitative methods, which may result in findings being based on a narrow group of informants. Some evaluation consultants also see this as a potential problem and have the impression that national stakeholders often tell them what they think will please the ILO and the constituents.

Evaluation managers do not always challenge these concerns regarding methods. Several interviews show that to respect the independency of the evaluation, evaluation managers refrained from commenting on the methodology proposed by the consultant during the inception report. This results in few comments on the methodology. An example is the stakeholders' weak analysis of the evaluation Promotion of Decent Work in Southern African Ports (Phase II), which according to the evaluation manager was not satisfactory (although the quality assessment of 2015 provided a good rating for this evaluation). This demonstrates a need to further clarify with evaluation managers how they should interpret the value of independence.

At the same time, this evaluation acknowledges that the scope for consultants to propose alternative evaluation methods is limited since, usually, the methodology has already been defined by the ToR, and the timeframe provided is often too short to develop and administer more complex data collection methods. In addition, ILO staff members and consultants express concern about limited project M&E systems, including lack of explicit ToCs, inadequate baselines and unreliable data collection systems.

The view from case study participants on credibility echoed the wider finding that evaluations are delivering on the basic requirements for accountability, but are not systematically going beyond them. Within TC project evaluations, users see a need for ILO to place greater emphasis on evaluating its value proposition, including policy influence and thematic expertise.

This broadening of the evidence base is not applicable only to the evaluation function, but requires more sophisticated RBM tools to track changes in discourse and policies over time. Participants suggested that to achieve this requires the adoption of approaches such as: (1) balanced scorecards; (2) including standard key indicators in all evaluations to allow for aggregation across the portfolio; and (3) using new technologies to capture a wider range of views.

Feedback from internal stakeholders during the evaluation process emphasized that, given the proximity to centralized capacity, there is a case for initiating innovations in both evaluation and RBM

in the decentralized evaluations that are undertaken by technical departments at HQ, especially in terms of supporting the use of evaluations for knowledge building and organizational learning.

However, at present, overall evaluation practices are found to be justified but conservative. As suggested by case study interviewees, and confirmed by the portfolio analysis, evaluation designs are consultative rather than participatory, and stick closely to an assessment of performance based on the original results framework. Consequently, evaluations are seen to struggle when dealing with poorly documented innovations and evolutions in programming, however relevant these may be in the wider context.

7.4.3 Finding 12: Disclosure

The transparency and accountability of the evaluation function is an organizational strength in ILO, and *i*-Track is a major organizational asset and world-leading system in supporting this. Continuing to invest in this comparative advantage and modernizing *i*-Track will help maintain this leadership.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	–	X	3/4
Constituents	–	–		–	
External	–	–	–	X	
– = nil.					

UNEG's Norm 10 (2005) and Norm 7 (2016) require that evaluation products be publicly accessible, and Standard 1.4 (2005) and Standard 1.5 (2005) require organizations to have an explicit disclosure policy for evaluations. Except for management responses, ILO is in full compliance with these norms and standards for independent evaluations (and most internal evaluations) through the *i*-Track database available on the ILO public website. The proportion of recommendations implemented by management is available through the AER, but this evaluation finds that the next evaluation policy will need to ensure that management responses themselves are published to be in full compliance with UNEG.

ILO's *i*-Track system was ahead of its time as an electronic repository of evaluations. Its use to track mandatory evaluations, link to project documents and finance data, and to ensure transparency is still relatively unusual, including outside of the UN system. *i*-Track is, therefore, an important organizational asset and the effort of EVAL to both maintain and extend *i*-Track is exemplarily. However, it is also based on older technology that means it is time intensive to keep up-to-date and unable to support decentralization of roles and responsibilities.

EVAL is aware of these limitations and has made some efforts to improve the user experience through a new interface, *i*-eval Discovery. This is a good step. It would also be advantageous, however, to improve the underlying technologies used by *i*-Track, and the evaluation finds that EVAL proposals to do so in 2017 are relevant, timely, and promise a high return on investment if undertaken using contemporary user-centred design.

7.5 FIT FOR THE FUTURE: EVALUATION FUNCTION MOVING FORWARD

7.5.1 Finding 13: Evaluation policy

The ILO evaluation policy has served the Organization well, but requires updating to meet the revised UNEG norms and standards and to better fit with the changing nature of ILO's external context and internal capacity.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	–	X	X	3/4
Constituents	X	–	–	–	
External	X	–	–	X	
– = nil.					

UNEG's Norm 3 (2005) and Norm 12 (2016) require that all organizations should have an explicit evaluation policy, and Standard 1.2 (2016) sets out the requirements of this policy. The standard also references JIU benchmark of 0.5–3.0 per cent of organizational expenditure being allocated to evaluation.

ILO's evaluation policy was established in response to the UNEG norms and standards (2005), which have since been updated (2016). For this reason alone, it is, therefore, relevant to review the current evaluation policy. In doing so, however, due consideration of some of the critiques of the UNEG norms and standards must be considered, including that they insufficiently account for the specific nature of decentralized evaluations.

ILO is one of only five entities assessed by JIU (2014, p. 56) to have high levels of institutionalization of policies for decentralized evaluation (the others being UNDP, UNFPA, UNICEF and UN Women). Revision of the evaluation policy is an opportunity to further consolidate this lead, and extend institutionalization to the portfolio of decentralized internal and self-evaluations. In this regard, JIU (2014, p. 57) draws attention to an emerging cluster of entities, including the Food and Agriculture Organization of the UN (FAO), UNEP, UNIDO, WFP and UN Women, in which decentralized evaluations are controlled to a degree by the central evaluation function. The institutionalization of EVAL since 2011 now makes such an approach an option for ILO, where before it was not.

Of the three main norms assessed by this evaluation (independence, credibility, utility), the current evaluation policy is strongly oriented towards independence and credibility. At the time of its creation, there was a common assumption that good quality evaluations would lead to increased use. More recent discourse in the development evaluation community has questioned this assumption, giving additional emphasis to utility as a consideration of equal weight (rather than a natural outcome of) independence and credibility.

Reflecting this, JIU's Recommendation 2 (2014) states: "…organizations should adopt a balanced approach in addressing the purpose of evaluation for accountability, and for the development of a learning organization that has the appropriate incentive systems for innovation, risk-taking and the use of multidisciplinary perspectives." Given the current practice observed by this evaluation, a revised evaluation policy could thus be better balanced between accountability, learning and innovation.

The mandatory evaluation arrangements in the current evaluation policy made sense in the context of the 2005 policy and the 2011 evaluation strategy. However, they are increasingly seen as rigid and not always responding to local needs, developing national evaluation capacity, addressing contributions to 2030 Agenda, or operating within the context of Delivering as One. The specific challenge faced by EVAL in adopting a more flexible approach is the dearth of specialist human resources for

evaluation. JIU (2014, p. 59) notes that the top-tier evaluation functions in the UN system, except for ILO, have been increasing the number of evaluation specialists at global, regional and national levels to meet the demands of a high-quality learning and accountability function.

It was observed in the case studies that greater flexibility in the criteria used to commission evaluations could also support greater ownership and use, whilst maintaining appropriate levels of accountability. In particular, the evaluation policy should provide guidance for increasingly common scenarios encountered at the decentralized level.

- Addressing anomalies created in regions with many high-budget TC projects, where multiple mandatory independent evaluations consult with the same stakeholders and contribute to evaluation fatigue. This is leading to the perception that evaluations are serving the policy, and not accountability, learning, or management needs.
- Addressing the duplication of evaluations where donors or joint programme partners insist on their own external evaluation of programming, through either clear guidance on joint evaluations (outside of joint programmes) or allowing repurposing of ILO evaluations to address questions not addressed by external evaluators.
- Establishing alternative evaluative options for mandatory final evaluations (for example to explore specific innovations or lessons) where a mid-term evaluation has only recently been completed or RBM data are found to be sufficient to report on the results framework.

One option from the evaluation policies of other UN entities that EVAL could further explore is to complement evaluation requirements based on project budgets with other targeting criteria, such as: (1) risky or innovative interventions; (2) demands from social partners; (3) potential for continuation, replication or scaling; (4) potential for joint UN country team evaluations; (5) knowledge gaps and opportunities to influence policies; and (6) scope to use innovative evaluation methods.

Purpose of evaluation

The evaluation policy could also more clearly define the purpose (vision and mission) of the ILO evaluation function. The case studies found that there are many diverse expectations of the evaluation function and, by extension EVAL, which are not always compatible with the scale and capacity of the human resources allocated to evaluation. These demands range from including evaluation evidence and evaluability assessments in the design of projects, to producing final project reports and impact evaluations of completed work. Establishing more clearly understood boundaries for EVAL's responsibilities will help to maintain support for the function among ILO colleagues.

From the perspective of accountability, interviewees suggest that a more robust system of follow-up is required to management responses. The current process is seen to be mechanistic, and lacking enforcement. Implementation is particularly problematic when there are no further phases to a project, with the participants emphasizing that more effective means need to be found to transmit evaluation insights to the technical specialists and project managers who design new projects and programmes.

Of the various purposes of evaluation, learning was found to be perhaps the more important for final evaluations given that the timing in most cases is both too early to see impact, and too late to influence decisions about future phases (the case studies found little evidence to suggest that donors are systematically using final evaluations for accountability purposes beyond the needs served by good quality monitoring data).

The socializing of evaluative insights into ILO performance is seen as a significant challenge in the regions. Whilst databases are a useful asset, published information alone is insufficient. Most case study interviewees agree that knowledge is only embedded into programming through dialogue. ILO's social partners, in particular, emphasize the importance that they place on being participants in evaluation processes rather than just receiving the outputs.

7.5.2 Finding 14: Evaluation strategy

ILO's evaluation strategy has been successfully implemented and has played an important role in maturing the evaluation function. It now requires updating to more accurately reflect the priorities, needs and objectives of the function's ToC.

Source	Cases	Portfolio	Survey	Documents	Level of evidence
ILO	X	X	X	X	3/4
Constituents	X	X	–	–	
External	X	X	–	X	
– = nil.					

Since the previous IEE in 2010, EVAL has successfully transitioned to a functionally independent unit within ILO's structure. The head of evaluation reports administratively to the DG and provides AERs to the GB on EVAL's activities with recommendations for its further work. This hybrid system is echoed in several other UN entities, including those assessed by JIU as being at the same level of evaluation maturity as ILO, such as UN Women.

Other significant work since 2010 has included in the adoption of an evaluation strategy, the AER to the GB, the production of a complete suite of evaluation guidelines and tools, the development of the EMCP, the development of a wide range of communication products (e.g. think-pieces, synthesis reviews and meta-studies), and ongoing work on the *i*-Track global database of evaluations. In essence, the focus of work has necessarily been on establishing robust systems and structures of evaluation within ILO, with less of a focus on strategic issues, more advanced methods, or utilization.

The evaluation strategy 2011–2016 included a results framework (with three outcomes and seven indicators) that has been reported on systematically each year through the AER. This was based firmly on the report of the IEE 2010. AERs have also included four additional indicators under Outcome 2: "harmonized Office-wide evaluation practice to support transparency and accountability". EVAL sees these additional indicators as more relevant than the original indicators.

Given the nature of most of the indicators and the reporting data for the evaluation strategy, the following table and figure 10 represent a qualitative assessment of the performance of the current evaluation strategy (see table 11 for criteria) based on data from the AERs triangulated with interview data. The analysis was carried out by comparing biennial milestones. The detailed data on the performance by outcome and per year are provided in Annex F.

Table 11. The criteria used for assessing the performance against the moving biennial milestones

Rating	Score	Description
Exceeding	3	Documented evidence substantially surpasses the stated biennial milestone in quantitative or qualitative terms
Met	2	Documented evidence supports the main focus of the biennial milestone as having been achieved
Approaching	1	Documented evidence demonstrates that substantive work is underway on this milestone, but the intended result has not yet manifested
Missing	0	Little documented evidence exists to suggest that this milestone will be met within the biennium

The following graph shows that capacity development of ILO staff and constituents has consistently exceeded the biennial milestones set for this output. The AER has consistently met its targets, whilst

the EAC has demonstrated the greatest improvement in performance. Work on the participation of constituents and the use of evaluation has represented a challenge for the evaluation function, whilst the institutionalization of roles and responsibilities has been attenuated by the long institutional processes involved (outside the control of the evaluation function). It should be noted that progress over time is generally rated against increasing biennial targets, suggesting that the absolute performance across all indicators was significantly higher in 2016 than it was in 2011.

Figure 10. Achievement of the biennial milestones for the evaluation function
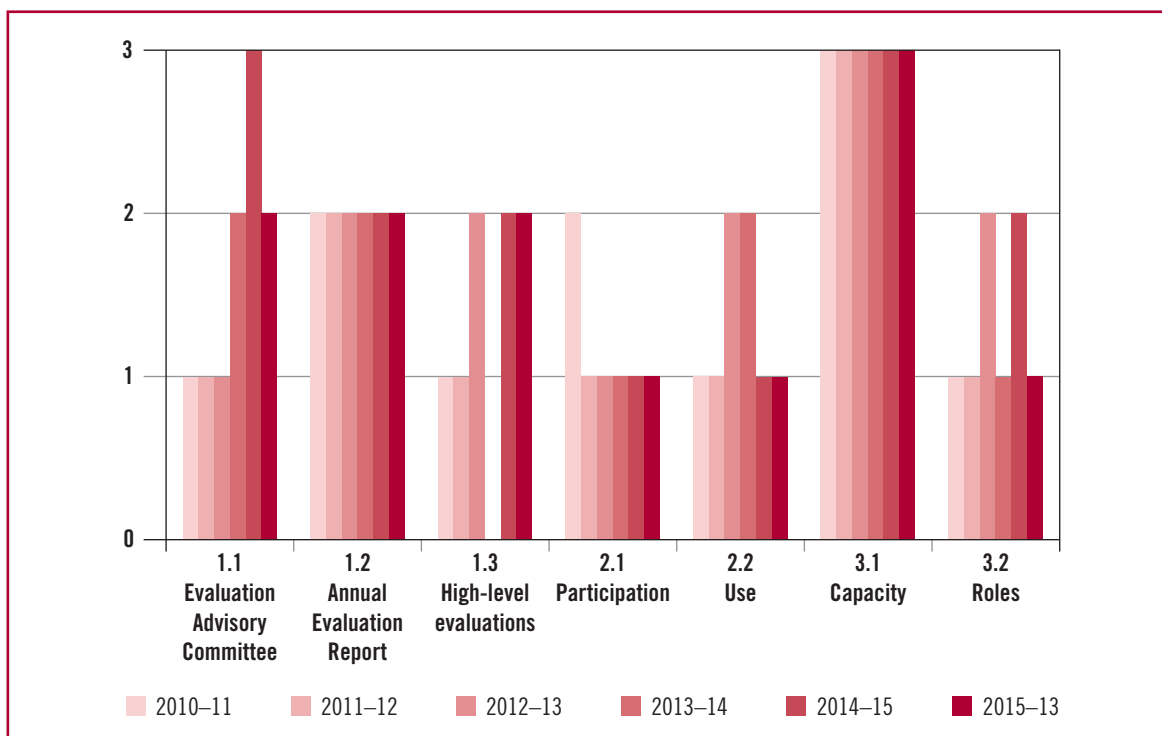(1 = approaching, 2 = met, 3 = exceeded)

Table 12 highlights the overall progress towards the outcomes of the evaluation strategy since 2011. Based on 2016 data, the strategy will have met its first outcome (improved use of evaluation for management), is in line to meet the second outcome (harmonized evaluation practice), and has exceeded its third outcome (developing capability).

Table 12. Progress towards the end targets of the 2011–2017 evaluation strategy results framework based on 2016 milestones

Outcome	Indicator	Progress (2016 milestones)
1: Improved use of evaluation by ILO constituents and management for governance	1.1: The frequency and quality of EAC decisions and advice on relevance of evaluation programme of work to GB policy decisions and strategic objectives of the Office; adequacy of follow-up to evaluation results	Met
	1.2: AER synthesizes recommendations and lessons learned based on evaluations	Met
	1.3: HLEs assess the contributions of technical and decent work country strategies to the SPF and P&B outcomes	Met

Outcome	Indicator	Progress (2016 milestones)
2: Harmonized Office-wide evaluation practice to support transparency and accountability	2.1: By 2015, 100 per cent of DWCPs and projects have mechanisms in place for regularly engaging constituents in the use of evaluation processes	Approaching (but indicator replaced by additional indicators – below)
	2.2: Upgrade and expand the use of evaluations for management (decentralized)	Approaching (but indicator replaced by additional indicators – below)
	Additional indicator A2.2.(1): Codification and upgrading of procedures and guidelines	Met
	Additional indicator A2.2(2): Maintenance of a strong departmental and regional evaluation network	Approaching. *(Network and workshops have taken place. However, the use of such network needs improvement)*
	Additional indicator A2.2(3a): Use of a rigorous quality control system	Met *(The system is rigorous but internal evaluations are excluded)*
	Additional indicator A2.2(3): Follow-up to evaluation recommendations	Met *(The follow-up system is improving and in 2015–2016, 89 per cent of recommendations were completed or partly completed)*
3: Evaluation capability expanded through enhanced knowledge, skills and tools	3.1: Evaluation capacity and practice among ILO staff and constituents improved	Exceeding *(Training throughout the Organization and among constituents has been successful)*
	3.2: For evaluation network, standardized roles and responsibilities applied throughout the ILO	Approaching *(Standardization is in progress, the approach and processes to be followed are currently under discussion)*

The evaluation strategy has thus largely been highly successful within its own terms. However, interviews and case studies found that there is relatively low knowledge of the evaluation strategy among stakeholders outside of EVAL. Furthermore, a mapping of EVAL activities since 2011 reveals that a lot of work has been done in accordance with the strategy, but other important work has also taken place outside of the scope of the strategy. The results framework, in particular, does not sufficiently capture efforts in the area of communication (for example, the use of decentralized evaluations is excluded from Outcome 1 indicators), meta-analyses and knowledge generation, joint working, or support to impact evaluation.

The JIU (2014) maturity matrix provides an alternative set of indicators for measuring the quality of outputs from the evaluation function (see table 13). These provide a more holistic view that the evaluation strategy of the work of the central evaluation function in ILO, and more aligned to the future direction of the evaluation function in the UN system.

Table 13. Qualitative performance of ILO evaluation function based on JIU (2014) output indicators

JIU output indicator	Current performance of ILO evaluation function
Coverage: Comprehensive, relevant, and adequate to support strategic decision-making, and development of policies, programmes, projects	▪ HLE themes selected through consultative process and aligned to decision-making needs of management and GB ▪ TC projects comprehensively covered by independent evaluations ▪ RB/RBSA requires greater and more consistent coverage, potentially by expanding DWCP and thematic evaluations ▪ Meta-syntheses valued at global level, but limited in number and not being used for decision-making at country level ▪ Potential for more explicit alignment with ILC themes

JIU output indicator	Current performance of ILO evaluation function
High quality and credible evaluation reports (impartial, valid, reliable, inclusive perspectives)	■ Independent evaluations have strong oversight and consistently meeting UNEG standards for reports – exceeding performance of most UN entities and with no evidence of evaluator partiality ■ Scope for greater integration of social dialogue into evaluation designs, including through an adapted ILO version of the OECD-DAC framework ■ Low oversight and quality assurance of internal and self-evaluations. Could undermine perception of the evaluation function
High-quality methodologies to advance development evaluation	■ Independent evaluations rate strongly in terms of data collection and analysis methods ■ Main weakness in terms of the evaluation principles and inadequate designs. Limited in most cases by the resources available ■ Increasing use of specialist evaluators rather than technical specialists is likely to support stronger methodologies. ILO orientation training for evaluators will be an important contribution. Strong case for greater use of teams ■ EVAL contributing to development of a high-quality impact evaluation function
Enhanced knowledge management: mining evaluations, synthesis, lessons learned and sharing internally and externally	■ EVAL publishes global meta-syntheses and meta-analyses, and increasingly integrates meta-syntheses into preparatory stages of HLEs ■ Ongoing improvements to website, dashboard, and communications products (newsletters/quick facts) ■ Demand for greater packaging/availability of evaluative evidence for decent work teams in formats other than reports/*i*-Track.
Strong learning organization and evaluation capacities: Enhanced capacity and culture for results, learning and improvement and knowledge development and use	■ Evaluation function has been institutionalized and structures and systems established ■ Need to continue to shift culture of evaluation from administrative to strategic, especially at decentralized levels ■ Greater need to nurture the evaluation network as a community of practice
Readiness and engagement for global challenges	■ EVAL has established important assets for the evaluation function (including *i*-Track and the ITCILO training courses), and has been recognized as one of the most mature evaluation functions in the UN system ■ Need for advocacy for inclusion of constituents in national evaluation capacity development by UN system ■ Flexibility and innovation required of the central evaluation function will prove challenging with existing levels of staff, ILO not keeping-up with comparable entities in this regard

As can be seen from the above table, a large body of the achievements of the central evaluation function under the current evaluation strategy are not sufficiently captured or reflected in its results framework. Interviews suggest that the current evaluation strategy results framework is mainly used for reporting purposes – the AERs include an update on developments and progress made with the implementation of the three outcomes identified in the strategy.

Evaluation function theory of change (ToC)

Section 6.1 of the report presents a reconstructed ToC for the evaluation function in ILO, along with a set of underlying assumptions, including: (1) a commitment to use evaluations in decision-making at all levels; (2) improvements in evaluation quality leading to greater use and demand for evaluations; (3) evaluation capacity will be sufficient to ensure methodological development and learning opportunities; and (4) extra-budgetary resources will be made available for training of constituents on evaluation management and implementation.

The evidence from the case studies suggests that, whilst there is an appreciation of the role of evaluation across the ILO, assuming a full commitment to the use of evaluation in decision-making does not fully account for the political economy surrounding decision–makers. Answering the right

question for the right person at the right time has been key to the examples of the use of evaluations so far.

Closely linked to this is how the quality of evaluation is being defined and how it would benefit from being revisited. Ensuring the quality of final evaluation reports to UNEG standards does not fully capture the broader aspects of timing, process, and social dialogue that all associate with driving use. How quality is defined is at the root of all other aspects of the evaluation strategy, including what is included in ILO's evaluation training programmes. Therefore, there is a case for the next evaluation strategy to define quality in ways that go beyond standardized checklists and are more attuned to social dialogue, ILS, gender, human rights processes, and the environment than the current UNEG standards.

Additional assumptions were also revealed during the evaluation case studies and analysis:

- Outcome 1 (increased use) implicitly focuses on the central evaluation function, and the use of evaluations by the GB and senior management. There is a need to revisit this assumption to give greater visibility to the use of evaluation at decentralized levels (regional and country), including by both ILO and its constituents.

- Outcome 2 (harmonized practice) has, until now, implicitly focused on the mechanisms and systems of the evaluation function. Now that these have been established, there is a case for revisiting this assumption and giving greater emphasis to building the evaluation network and culture within ILO and its constituents.

- Outcome 3 (evaluation capability) implicitly emphasized support to the achievement of the other two outcomes. Whilst many constituents have been trained through the courses developed under Outcome 3, the focus of those courses has been based on meeting the internal needs and requirements of ILO. Additional considerations for a future strategy include being more outward looking (especially given the 2030 Agenda) and advocating within the UN system for building the national evaluation capacity of all ILO's social partners.

Revisiting these assumptions in the ToC also provides an opportunity to build stronger ownership through more inclusive and participatory design processes for the future evaluation strategy. This needs to be cognisant of the enabling environment, key assumptions, and pre-conditions for the success of the evaluation function. Some of the pre-conditions for its further development identified by the IEE are listed below in table 14.

Table 14. Examples of assumptions and pre-conditions likely to impact a future evaluation strategy

Pre-condition/assumption	Enabling contributors
Strengthened RBM and M&E systems and results reporting	P&B, TC project design, individual donors, GB, PARDEV, PROGRAM
Responsibility for management response at the decentralized level	Regional directors, technical department heads, DWP team leaders, country representatives
Agreement that evaluation should be more flexible and allow for strategic pooling of resources	GB, donors, finance, auditors
Institutional recognition and incentives for evaluation managers	Human resources
National evaluation capacity development for constituents	UN country teams, large UN Development Group (UNDG) agencies (UNDP, UNICEF, etc.)
Expansion of decent work evaluators	UNEG working groups, evaluation networks, VOPEs

There is a demand from constituents to better articulate, measure and report on the impact of the evaluation function itself. Whilst the current indicators track the number of people trained, or undertake

quantitative coverage of projects and programmes, there is little analysis of the implications of these achievements in advancing ILO's mandate.

This is an area where, according to JIU (2014), all UN entities are weak and, thus, there is an opportunity for ILO to pioneer such an approach. Whilst there is little guidance available on how to measure the impact of the evaluation function, the JIU (2014, p. 65) assessment did propose a high-level ToC that could be used to inform the development of the next strategy. In particular, this covers the relationship between the outcomes of the evaluation strategy and the intended impact of "a more relevant, efficient and effective ILO, with greater impact on the lives of the people it serves" (table 15).

Table 15. Gaps in the tracking of evaluation function impact and higher level outcomes in the current evaluation strategy (based on JIU ToC)

JIU evaluation function impact indicator	Presence in current ILO evaluation strategy
Achievement of development and normative goals	Evaluation function included in strategic plan DRF
Enhanced credibility [of ILO]	IEE assesses credibility, and Outcome 2 is a proxy for quality, but not fully tracked formally
Effective oversight and governance	Captured under Outcome 1 at central level, not captured at national level
Enhanced quality of strategic plans, policies, programmes, projects and institutional reform	Proxy of implementation of recommendations tracked, but no assessment of the quality of work over time and correlation with evaluation
Enhanced global leadership role [for ILO]	Activities undertaken by EVAL, but not tracked in evaluation strategy results framework
Enhanced accountability for results	Proxy of implementation of recommendations tracked
High-level use of recommendations	Tracked under Outcome 1, EAC follow-up HLEs Decentralized use of evaluations not fully tracked
Enhanced contribution to advancing development evaluation and to global knowledge	Activities undertaken by EVAL, but not tracked in evaluation strategy results framework
Increased collaboration, efficiency and alignments in UN system evaluation	Activities undertaken by EVAL, but not tracked in evaluation strategy results framework

8 CONCLUSIONS

Overall conclusion: In the course of the evaluation strategy 2011–2017, the evaluation function in ILO has been transformed in terms of its structural independence, institutionalization of evaluation practice, and development of material. It is highly regarded for having achieved this with limited resources, including becoming a more visible champion for gender mainstreaming and for introducing good practices that are in some ways ahead of UN evaluation peers.

The independent evaluation of the evaluation function (2010) strongly emphasized its structural independence. As a result of the changes made following that evaluation, and the subsequent work of EVAL and evaluation focal points, the ILO is now recognized as having one of the three most mature evaluation functions in the UN system. This is a significant achievement.

EVAL has understandably emphasized its independence during the course of this transformation. As a small centralized office within ILO, it has progressed by establishing mandatory requirements and highly structured systems and processes to deliver the evaluation function. These include some common tools, such as guidance and manuals, and some innovations, such as the EMCP and linking the *i*-Track database to other corporate IT systems.

The implication of this success is a systemized evaluation function that is strongly independent and consistently delivers its technical requirements to UNEG standards. This has attenuated the need for an integrated evaluation planning system, since nearly all current evaluations are mandatory and are triggered through an 'automatic' process. Looking forward, a more flexible, integrated and utilization-focused approach will be increasingly demanded of EVAL. Meeting this challenge with the same small team and without sacrificing current coverage and quality will be the central challenge for the next evaluation strategy.

8.1 INDEPENDENCE

Conclusion 1: Priorities for enhancing the independence of the evaluation function are the integration of REOs as full staff members of EVAL, strengthening the capacity of evaluation managers and expanding quality assurance systems to internal evaluations.

Whilst the independence of the evaluation function has increased considerably, several limitations remain. The evaluation acknowledges that the small number of interviewees who expressed concern about independence did so regarding the accountability of the head of evaluation to the DG, or to the conditions attached to the position of the head of evaluation. Having considered these perspectives, and triangulated them with other data, the evaluation is of the view that the main priorities regarding independence are, however, not these issues.

The highest priority regarding independence is to transition the REOs into being full EVAL staff members located in the regions. These regional evaluation specialists may be administratively supported by the ROs, but independence requires that they would report directly, and only, to the head of evaluation.

The main barrier to such an arrangement is the structural underinvestment by ILO field offices and projects in M&E specialists to support the RBM requirements of projects and programmes. This is manifested in substantial demands being placed on the REOs to support evaluability and monitoring of programmes (which is included in their ToRs). The implication of this arrangement is insufficient time to support either monitoring or evaluation functions adequately, and a critical missing link in the 'independence chain'. It is recognized that not all projects, especially smaller ones, would require dedicated M&E capacity (in terms of staffing) but could instead make use of external consultants for specific inputs.

A second priority regarding independence is to continue exploring ways of supporting evaluation managers' knowledge skills and understanding. The evaluation concludes that the current evaluation manager system is also primarily a response the shortage of M&E specialists (rather than purely about independence), but that it, nevertheless, has real potential to support across-organizational learning. Fulfilling this potential requires that evaluation managers have a more nuanced understanding of how to achieve independence, and sufficient confidence to critically challenge evaluators where they see gaps in their methods or analysis.

8.2 CREDIBILITY

Conclusion 2: Whilst independent evaluation reports largely meet UNEG standards, there is much to gain from increasing the diversity of evaluations, mainstreaming social dialogue, and deepening the participation of constituents.

One of the key questions asked of this evaluation was to examine the persistent level of quality ratings for evaluation reports despite the wide range of systems strengthening activities being undertaken by EVAL. The main observations of the evaluation in this regard are that the current portfolio is relatively homogeneous in terms of the size, methods, evaluators and objectives. The dominant evaluation approaches in ILO are somewhat conservative, focusing on examining the achievement of results frameworks rather than examining the underlying ToCs.

There is also the case that quality standards themselves are somewhat reductionist, and only measure certain aspects of quality. This gives emphasis to standardization, rather than appropriateness or fit-for-purpose, which are different hallmarks of credibility. Examination of a broader range of evidence leads this evaluation to conclude that the two main facets of quality that are challenging ILO are:

- the need to expand the use of participatory methods that both model social dialogue and engage social partners in examining the impact of normative changes; and
- the flexibility to ensure credibility by commissioning the right evaluation with the right purpose at the right time.

The existence of these challenges is driven by the real and perceived demand for set levels of accountability from donors and the GB which (in the absence of a mandate from donors to commission more strategic clustered evaluations and/or sufficient staff with the expertise to make informed professional judgements on commissioning evaluations) has been achieved through large numbers of small-budget mandatory project evaluations (restricting the amount of time required to be highly participatory).

The implication of continuing this model, however, is that opportunities are being missed to more comprehensively address the questions of constituents, build the ownership of interventions through

participation, and transfer evaluation capacity to the social partners at national level. Given the constraints observed during this evaluation, EVAL's work in support of some of the flagship programmes seems to be the most promising entry point to pilot greater flexible and more participatory evaluation approaches.

Conclusion 3: Enhancing evaluation budgets, rosters, networks and procurement processes is required to consistently secure evaluators with the combination of evaluation skills and technical knowledge that ILO requires.

The combination of technical areas, normative work, and tripartite working means that ILO occupies a niche space in the international evaluation arena. Evaluators that combine experience of these dimensions with geographic, lingual and evaluative experience are relatively rare. This is why many ILO interviewees spoke to the evaluation about challenges in finding evaluators, and having to continuously work with the same individuals. The relatively low rates being offered by ILO, and the easier administrative procedures for contracting individuals rather than firms, narrows the field still further.

At the same time, the evaluation found that ILO has limited direct links with professional evaluation associations and national networks to expand the community of evaluators aware of and learning about decent work. To some extent, EVAL is hampered in doing this through its centralized location in Geneva and already high workload. However, the implication of not having a specific strategy to actively support evaluators to move into ILO's fields of interest is to continue being reliant on a restricted pool of expertise.

8.3 UTILITY

Conclusion 4: There is significant value to be realized from strengthening communications and knowledge management to enhance the utility and use of evaluation in the field.

The focus of the evaluation strategy (under Outcome 1) has been on the use of evaluations for governance at central level. An underlying assumption of Outcome 2 is that increasing evaluation quality will also lead to an increase in use in the field. Whilst the evaluation agrees that this may be partially true, it concludes that the primary driver of utility is not quality. A strong factor that emerged from the portfolio analysis was timing (the opportunity to implement recommendations immediately). However, communications and knowledge management were also highlighted as key ingredients in enhancing use outside of the context in which the evaluation was completed.

The importance of communication is recognized by EVAL and inter alia reflected in the efforts made to produce meta/synthesis studies and other knowledge products and the development of the *i*-Track platform. At the same time, evidence suggests that more could be done by ILO to support EVAL in ensuring that evaluations are shared with different audiences and with the right people, in an appropriate form, through relevant channels, and in a timely manner.

The development of a communications strategy and the appointment of a knowledge manager have laid important groundwork. At the same time, a more tailored communications strategy, disaggregating audiences at both global level and in the context of each evaluation process, would have significant benefits. The institutional link to the ITCILO is already an example of EVAL supplementing its communications capacity, and a similar strategic collaboration could be explored with DCOMM. The evaluation recognizes that EVAL already collaborates with DCOMM on a regular basis, but mainly on a case-by-case basis. By involving DCOMM in the development of the next evaluation strategy, more strategic and systematic linkages could be explored.

The evaluation highlighted the value of direct dialogue between EVAL staff, colleagues and constituents. In this sense, the informal relationships held by EVAL staff and the official internal evaluation network could both be a valuable bi-directional communications channel. At the other

end of the spectrum, EVAL's comprehensive system for documenting and storing evaluation materials indicates that there is potential to explore the use of machine learning or artificial intelligence to help match existing evaluation evidence to specific user needs.

Conclusion 5: At the decentralized level, there is a need to replicate the same success that the EAC has had in ensuring management responds effectively to recommendations from HLEs.

In its work following up on management responses, the EAC plays two important roles: firstly, it helps to distribute learning among senior managers; and secondly, it demonstrates that accountability through evaluation is a whole-of-Organization (and not just EVAL) responsibility. The system for following up management responses at the decentralized level is far inferior to the EAC arrangement and needs to be more strongly represented in the indicators for a future evaluation strategy.

The evaluation concludes that strengthening the field system for following up on management responses would make a significant contribution to use, and would also bolster EVAL's ability to more systematically track the outcomes of evaluation processes to meet the GB's interest in hearing about the impact of evaluation beyond the percentage of recommendations addressed.

8.4 ENABLING ENVIRONMENT

Conclusion 6: Despite high-level support for evaluation in the ILO, the emergence of an evaluation culture is inhibited by underinvestment in M&E specialists and too few institutional incentives for programme and technical staff to engage with evaluation.

Many of the requests and suggestions encountered during interviews for this evaluation were essentially to address gaps in the wider RBM system (especially regarding evaluable indicators or the incorporation of evaluation recommendations). Whilst these are relevant and necessary to programming needs, they are not typically the mandate of a central evaluation function.

Thus, it is important to recognize that the boundaries of what one can expect from the evaluation function are strongly influenced by the environment. The need for more M&E capacity within projects and programmes limits the results data available to evaluations, and thus the range of methods that can be used or sophistication of the questions that can be addressed. Similarly, the weak organizational incentives for programme and project staff to engage in evaluations – as witnessed for evaluation management – limits the capacity of the function.

Despite the independence of EVAL, the evaluation function in ILO does not exist in isolation. It is embedded within a culture framework and set of institutional drivers that will also need to be adjust if the performance of the function is to improve. At all levels of the Organization, there is a need for the leadership to nurture a tangible sense of evaluation as an organizational comparative advantage (rather than an issue of compliance) in order to drive willingness to invest in RBM, evaluations, and evaluation management.

Conclusion 7: The evaluation policy (2005) and the evaluation strategy (2011) have served their purpose well, but both now need updating to meet the challenges of a changing context.

The evaluation policy (2005) was an immediate response to the UNEG norms and standards that were agreed in that year, and it has provided a strong framework for evaluation to be institutionalized in ILO since that time. Similarly, the evaluation strategy (2011) was an immediate response to the IEE (2010). However, the results framework of the current strategy has become insufficient as a management tool and requires a fundamental revision. This evaluation sees three other main challenges facing the design of the next evaluation strategy:

- determining a means to strengthen the focus of the evaluation function as regards utility and national capacity in a structurally underfunded function and in a way that does not jeopardize the gains already made in systematizing an independent evaluation function within ILO;

- expanding the use of strategic, joint, thematic and DWCP evaluations in a way that does not leave TC projects with insufficient evaluative coverage;
- progressively integrating the large body of internal and self-evaluations into ILO's quality assurance and system-strengthening process, and dealing with the implications of this on both human resources and the probable near-term drop in average report quality that will result from this integration.

Missing from the current strategy is a clear ToC for the evaluation function, and indicators of the outcome of the evaluation itself (beyond the proportion of recommendations adopted). Also missing from the current policy is a clear mission for the evaluation function, and guided flexibility to maximize the value of each evaluation (for example, by avoiding duplication with external evaluations).

The evidence from this evaluation indicates that the one of the main contributing factors to this situation was the top-down development of the current strategy, based primarily on the recommendations (and framework) of the IEE (2010). The ToR for this evaluation can be interpreted to imply that a similar process is intended for the next evaluation strategy. However, whilst this evaluation should serve as a useful input to that process (along with the UNEG norms and standards, 2016), it concludes that – to be owned and understood – the next strategy (and policy) needs to be developed through a participatory process facilitated by EVAL.

9 RECOMMENDATIONS

The following 13 recommendations have been developed by the evaluation team based on the conclusions, findings, and consultation with the technical committee for the IEE.

Recommendation 1: Recognizing ILO's comparative leadership in evaluation in the UN system, the necessity for evaluative thinking to meet the 2030 Agenda, and the need to consolidate an emerging evaluation culture, it is recommended that the GB and the senior management of ILO strongly reassert organizational commitment to the evaluation function. Given the current trajectory and the recommended level of budgetary and political support, it would be possible to set an ambitious target of ILO beginning the transition to the highest level of the JIU maturity matrix.

Starting in 2010, support for evaluation within both the GB and senior management has strengthened EVAL's mandate and, within this, EVAL has established the most mature evaluation system among comparable UN entities. In a resource-constrained environment with ambitious global targets in the form of the 2030 Agenda, this represents an important comparative advantage.

Based on the current trajectory and plans, maintaining the current level of support to the evaluation function could see it fully achieve Level 4 of the JIU maturity matrix in the next five years. Level 4 is defined by JIU (2014) as: "…integration and institutionalization of the elements of the components of the function has mostly been achieved and the focus for improvement has shifted to enhancing the value added effect for the organization, as well as for the United Nations system as a whole. Increased attention is given to partnerships and alternative ways of enhancing effectiveness, efficiency, sustainability, and impact."

Maintaining, consolidating *and advancing* the emerging facets of an evaluation culture and enabling environment for evaluation in ILO requires four key commitments in the management response to the IEE:

1. alignment of organizational incentives for professional staff to participate in evaluation, practice evaluative skills, and adopt a learning culture;

2. good budgetary support to the evaluation function (ILO is currently spending up to 0.8 per cent of its budget on evaluation – JIU recommends up to 3 per cent);

3. increasing the number of professional-level evaluation staff (for example, ILO has seven professional staff, whilst UN Women has 12); and

4. a clearly articulated vision of evaluation's role and its added value and support of the organization.

With growing attention on evaluation in the UN system and its centrality in the delivery of the 2030 Agenda, many other UN entities – particularly, the development agencies, but also the specialized agencies – are working to improve their evaluation functions. Maintaining ILO's comparative advantage, and ensuring that it can fully capture the contributions made to the SDGs, thus requires additional support to these four commitments.

With an increase in the organizational commitment to the evaluation function, the GB and senior management can also afford to be ambitious in the expected outcomes from the evaluation strategy (including indicators of the impact of the evaluation function on ILO's effectiveness). Under the right conditions, the IEE proposes that a stretch target could even be to see the ILO evaluation function transitioning to the highest level of the JIU maturity matrix within a 5–7 year timeframe.

> "Level 5. At this stage, the focus is on the function as a major agent of change at organizational, United Nations system, national, regional or global level. Focus is also on making a contribution to cutting edge methods for evaluation and further enhancing the value of evaluation at all levels in an integrated fashion (organization, United Nations System, globally and nationally)." (JIU, 2014, p. 19.)

Responsible units	Priority	Timing	Resource implication
DG, GB	High	Short-term	Medium

9.1 INDEPENDENCE

Recommendation 2: Transition REOs into full staff members of EVAL.

Convert REOs into evaluation officer positions at a universal grade level. These positions should be directly subordinated to the director of EVAL and funded through regular budgetary sources reallocated to EVAL for this purpose. Subsequently, the responsibility for quality assurance of decentralized independent evaluations (region-specific) should be fully delegated to REOs. Senior evaluation officers should be re-tasked to support the strengthening of evaluation systems, communications, communities of practice, and to manager a greater diversity of strategic evaluations.

Responsible units	Priority	Timing	Resource implication
DG, PROGRAM	High	Medium-term	Low (if funds reallocated)

Recommendation 3: Incentivize and strengthen the evaluation manager and focal person system.

Build an incentive structure whereby the evaluation manager's function is recognized in job descriptions and the annual performance appraisal process. Further incentives can be created by providing opportunities for certified evaluation managers to participate in evaluation planning and debriefing sessions at country level, and strengthen EVAL's network to provide access to further competence development. To strengthen the independence of the system, publish forthcoming evaluation management opportunities internally and interregionally.

Facilitate the internal network of evaluation focal points and REOs into a community of practice. Support shared action through planning, and undertake well-defined activities aimed at strengthening the evaluation culture in ILO. More regular meetings of the network should be organized and "brown bag" lunches or similar could be organized to share evaluation findings and new evaluation thinking.

Responsible units	Priority	Timing	Resource implication
DDG, HRD, EVAL	High	Medium-term	Medium

Recommendation 4: Establish an integrated evaluation planning system.

Develop a consolidated, formal evaluation planning mechanism to ensure better sequencing and coordination of high-level and decentralized evaluations; and directly link budgetary control of TC project evaluation allocations to the central evaluation function to allow for more clustered and strategic evaluations. The system should generate an annual evaluation plan stating the title and type of evaluation, region/country, timing, budget and funding sources, responsible unit, etc., to be reviewed and updated at regular intervals. The timing of centralized HLEs should accordingly coincide with overall programming and budgeting cycles to ensure that resources are available for follow-up. The selection of HLE case-study countries and decisions to cluster project evaluations should be made in a coordinated manner to avoid certain countries becoming 'over-sampled' at particular points in time.

Responsible units	9.6 Priority	9.7 Timing	9.8 Resource implication
EVAL, PROGRAM, PARDEV	Medium	Medium-term	Low

9.2 CREDIBILITY

Recommendation 5: Further develop collaboration with other agencies of the UN system to advocate for and support a diverse community of evaluators and national constituents with expertise in evaluating decent work and promoting social dialogue.

Advertising the opportunities more widely, and increasing the maximum daily rates to drive demand could expand the pool of prospective consultants. In addition, the process of hiring consultants should become more transparent and competitive to avoid situations where the same consultants are re-hired and a co-dependency emerges between these consultants and the ILO. Working with UN entities, VOPEs, and the ITCILO to support awareness of and skills in evaluating decent work could also drive supply. Prioritize promoting joint evaluations to give evaluators working with other UN entities a chance to build experience in mainstreaming the DWA into evaluation.

In line with EVAL's published position on the SDGs and the 2030 Agenda, engage national constituents more systematically and closely in the development of M&E systems, continuous monitoring, and evaluation processes. This would involve a mix of different initiatives including training and other forms of capacity building for national constituents, the development and targeted dissemination of tailored communication products, and mechanisms for self-assessment and regular feedback during project implementation. Specific project interventions could also be designed and supported that contribute to the development of statistics on employment and other socio-economic indicators, allowing for better impact measurement.

Improved communications and the dissemination of evaluation products among constituents could be promoted by managers through the development of a formal process identifying a focal person that engages in the formal communication of lessons learned, recommendations and follow-up with the constituents.

Responsible units	Priority	Timing	Resource implication
EVAL, technical departments and ROs, ITCILO, MULTILATERALS, PROGRAM	Medium	Medium-term	Medium

Recommendation 6: Enhance evaluation value added and relevance by promoting participatory, gender-responsive and mixed-methods evaluation.

Develop a model evaluation framework for evaluating decent work (including normative interventions and taking account of UN SWAP) through social dialogue. Encourage evaluation consultants and evaluation managers to combine this with ToC-based approaches and allow flexibility and time for them to develop and administer their own corresponding gender-responsive evaluation methods and data collection tools. Evaluation consultants should also have sufficient time and opportunity to debrief stakeholders, and test and validate tentative findings and recommendations prior to report drafting. This would promote process-oriented learning and ownership of evaluation results. Prioritize the strengthening of evaluability and monitoring systems, including through the inclusion of M&E specialists in decent work teams and TC projects. Possibilities may also be explored to embed the evaluation function in projects by contracting evaluation consultants to implement developmental evaluation approaches.

Responsible units	Priority	Timing	Resource implication
EVAL, PARDEV, PROGRAM, technical departments and ROs	Medium	Short-term	Medium

Recommendation 7: Expand the quality assurance system to include internal evaluations, and switch to an annual or real-time independent quality assurance system.

The current quality assurance system is comprehensive, but does not cover the full range of evaluations undertaken in ILO. This carries risks in terms of both the quality of evidence being used to make decisions and reputational issues regarding the quality of some evaluations. To address this, expand the quality assurance system to provide feedback to commissioning offices on the quality of internal evaluation reports under a long-term agreement for services with an external contractor. This would also allow EVAL to report annually on evaluation quality to the GB and UN SWAP.

Responsible units	Priority	Timing	Resource implication
EVAL	Medium	Medium-term	Medium

9.3 UTILITY

Recommendation 8: Diversify and elevate the overall portfolio of evaluations to include more DWCP evaluations and thematic evaluations.

Introduce the flexibility and mandate within the evaluation policy and strategy for EVAL to adjust the overall focus of the evaluation function from project evaluations to DWCP evaluations, thematic agendas and flagship programmes (including with case studies of projects where required by donors). Work with like-minded donors to pilot the use of utility-focused criteria for the selection

of evaluations. For example, if mid-term evaluation is considered more important for any given project, the final evaluation could be replaced by a cluster evaluation of several projects within the same area. Introduce the modality of joint evaluations with donors within the evaluation policy, and/or the flexibility to reallocate resources to thematic/DWCP evaluations when external donors enforce evaluations. Consider allowing regional adjustment of the financial thresholds for mandatory evaluations.

Responsible units	Priority	Timing	Resource implication
PARDEV, EVAL, PROGRAM	High	Medium-term	Low

Recommendation 9: Strengthen the decentralized evaluation management response mechanism.

Continue to advocate and support managers' sharing and proper discussion of evaluation recommendations with intended users, and ensure that joint agreement is reached on follow-up actions. REOs and DEFPs should be made responsible for ensuring that management responses are completed adequately and are carried out in a timely manner with the required level of detail to allow for tracking the implementation of agreed actions. This would involve a qualitative assessment conducted through a participatory process involving constituents. REOs and DEFPs can also ensure that management responses to relevant evaluations are extracted, consolidated and circulated during the design and appraisal of new projects in the same area or of a similar nature.

Responsible units	Priority	Timing	Resource implication
ROs, technical departments	High	Medium-term	Low

Recommendation 10: Prioritize EVAL's communications capacity and coaching function.

Revisit EVAL's communications strategy and explore interfaces with ILO's DCOMM with a view to developing targeted communications products. Specific emphasis should be placed on reaching out to senior ILO managers and national constituents. If senior evaluation officers were relieved of their quality assurance function (see above), more time would be available for them to undertake meta-analyses and synthesis reviews, and to support the knowledge management/learning function through direct dialogue.

Responsible units	9.12 Priority	9.13 Timing	9.14 Resource implication
EVAL, DCOMM	Medium	Medium-term	Medium

9.4 ENABLING ENVIRONMENT

Recommendation 11: Strengthen RBM and M&E systems to promote DWCP, programme and project evaluability.

EVAL, PARDEV and PROGRAM should cooperate to ensure that sufficient resources are allocated from TC projects, RBSA and in DWCPs to build M&E systems capable of capturing contributions to the SDGs and policy changes. This can include flexible enforcement of the 3 per cent allocation in budgets for RBM. As a minimum, ROs should be encouraged to support national M&E officers covering a subregional area. EVAL should commit to supporting the technical M&E capacity of

these officers, which should include encouraging national professional staff to take the EMCP and provide opportunities for internships in EVAL.

Responsible units	Priority	Timing	Resource implication
PARDEV, PROGRAM, MULTILATER- ALS, technical departments, ROs	Medium	Medium-term	High

Recommendation 12: Update and align the evaluation policy to IEE recommendations and current organizational structure and processes.

The ILO evaluation policy should be updated to reflect changes in the ILO organizational structure, strategy and programming that have taken place since 2005, the evolution of international standards (focusing on the 2016 UNEG norms and standards), and external developments of relevance to the evaluation agenda, specifically the SDGs and 2030 Agenda. The policy should also reflect the progress made to date by ILO in terms of strengthening evaluation independence, credibility and use, and outline general priority areas for future capacity development (in line with the evaluation strategy). It is important that the evaluation policy reflect the increasing focus on utility and an enabling environment and the findings of the IEE. Specific recommendations for the policy are:

- Align with UNEG norms and standards 2016.

- Commit to sufficient resourcing of the evaluation function in the policy (achieved by including a funding floor in the evaluation strategy recommended at 1.5 per cent of total organizational expenditure on evaluation – equivalent to $6 million per year).

- Establish an accounting code to track evaluation expenditure across all budgets.

- In the short term, qualify the existing 2 per cent allocation for evaluation in TC projects over $500,000 with lower and upper limits (for example, $40,000–200,000). Based on the experience from implementing these, negotiate a replacement organizational policy based on needs analysis and real-cost budgeting.

- Progressively increase the budget allocated to HLEs to twice the current level to allow for more advanced methods and more diverse teams.

- Codify the requirement that heads of evaluation should have qualifications and substantial experience in evaluation, complemented by experience in the related fields of strategic planning, basic and operational research, and knowledge management, and should have excellent management and leadership attributes.

- Elaborate a ToC and mission statement for the evaluation function in ILO.

- Clarify the boundaries of the evaluation function as regards RBM and M&E functions.

- Require that the EAC investigate and publish a considered opinion on the ILO response to term limits and no re-entry for the head of evaluation.

- Include provision in the strategy for a further IEE or UNEG peer review at the end of five years.

Responsible units	Priority	Timing	Resource implication
GB, DG, EVAL	Medium	Short-term	Low

Recommendation 13: Develop the new evaluation strategy in a participatory manner to promote ownership and visibility.

A new evaluation strategy should be developed based on the findings and recommendations of the IEE. The process should be participatory, involving REOs and DEFPs, to ensure ownership and commitment to its implementation, and be centred on an explicit ToC that defines and visualizes the logical chain of strategic priorities, intermediate outcomes, outcomes and impacts, and associated processes, and driving factors and assumptions. In line with the findings and recommendations of the IEE, the focus of the evaluation strategy should be on improving utility and strengthening the evaluation culture in ILO. It should be implemented with the help of an annual action plan linked to the overall programme and budgeting cycle and processes, and identify responsibilities across the Organization for the implementation of specific actions and achievement of deliverables (outputs). It is recommended that the strategy should include sufficient outcome indicators to report on the impact on the evaluation function; and that it should give renewed priority to learning, communication and national evaluation capacity.

Responsible units	Priority	Timing	Resource implication
EVAL, technical departments, DG, GB	High	Short-term	Low

9.5 OFFICE RESPONSE

Discussed and agreed upon during the 329th Session of the Governing Body (GB.329/PFA/8).

The office response below addresses the recommendations, the implementation of which depends on a number of pre-conditions to be met linked to institutional approaches and enabling environment as identified in the full report. Further refinement of the recommendations will be reflected in the new policy and strategy.

Recommendations	Office response	EVAL observations
Overall recommendation	The Office welcomes the finding that substantial progress has been made in terms of structural independence, evaluation practices, capacity building and guidance material within a limited budget. It takes note of the need to consolidate an emerging evaluation culture and reasserts it organizational commitment to evaluation and the specific challenges linked to the 2030 Agenda.	
Recommendations 1-3 (independence)	The Office recognizes that constant improvement is required to enhance the high-level of independence the function has already achieved. The Office will review reporting lines, incentive structures, and integration of existing work planning tools at the global and regional level as well as funding arrangement combining both regular and extra-budgetary resources with the aim to optimize evaluations, from more strategic coverage to enhanced use.	Zero-growth budget limitations and domestic accountability requirements of donors may have an impact on the ability of the Office to implement these recommendations.
Recommendations 4-6 (credibility)	The Office welcomes a focus on credibility from the perspective of evaluation methods that both model social dialogue and engage social partners. The Office will work on capturing these issues in appropriate evaluation frameworks while recognizing that it will be an Office-wide effort to bring in institutional strategies and investment in capacity building, data collection and reporting to improve overall evaluability. The existing quality assurance mechanism for project evaluations will be expanded and the frequency increased to include internal evaluations as part of ongoing efforts to enhance organization learning.	In terms of priorities EVAL questions whether limited evaluation resources should be invested in more oversight on internal evaluations. Rather the focus should be on strengthening a culture of (self) learning among managers as part of RBM.

Recommendations	Office response	EVAL observations
Recommendations 7-9 (utility)	The need for a more clustered and strategic evaluation portfolio concurs with earlier evaluation recommendations and is already an important consideration in the flagship programmes but will be subject to agreements and approaches in development cooperation funding that are conducive to pooling of resources. The decentralized evaluation management response will be strengthened building on the existing *i-eval discovery database* complemented with a dynamic *recommendation follow-up system* and *a regional advisory body* modeled on the EAC. The Office will also request EVAL to step-up collaboration with technical departments to conduct more meta and synthesis studies as part of knowledge management and learning but without compromising its independence. The existing evaluation communication strategy will be revised in collaboration with, and building on the capacity of the communication department.	A strong Office-wide coordination system will be needed including the regions to strengthen integrated and more strategic evaluation planning with decentralized follow-up systems.
Recommendations 10-12 (enabling environment)	The Office is continuously strengthening RBM and M&E systems and will continue efforts to ensure that established guidelines and approaches on RBM and M&E are adhered to and expanded. A revised evaluation policy will be presented to the Governing Body in November 2017 reflecting the findings and recommendations of the IEE. The new Evaluation Strategy 2018-21 to implement the revised policy will be submitted in March 2018 to allow for a participatory and consultative approach in its formulation.	Support from senior management will be needed to foster Office-wide involvement and ownership in the new policy and strategy.

10 LESSONS LEARNED

The following six lessons learned are put forward for further reflection and consideration.

Lesson 1: Evaluation as an integrated function

Evaluative thinking is an integral part of RBM and organizational learning. In ILO, this has partly been achieved by establishing a direct reporting line between EVAL and the GB, by establishing of an EAC, and by developing the management response mechanism and a related formal requirement to take evaluation findings and lessons learned into account in the process of project appraisal and proposal writing. Yet, evidence suggests that the evaluation culture can be further improved. Evaluation findings and recommendations are not sufficiently transmitted to intended users, project M&E systems are generally weak, and there is a lack of institutional incentives for staff to participate in, and use and learn from evaluations. This lesson learned refers back to the conclusions about utility and an enabling environment and is applicable across ILO units, including the DG, GB, PARDEV, PROGRAM and MULTILATERALS, Technical departments and ROs, as well as EVAL. It is captured in the overall recommendation, i.e. in the four key commitments required to advance the evaluation culture and create an enabling environment for evaluation, and in the specific recommendation (Recommendation 10) on the further strengthening of RBM and M&E systems.

Lesson 2: Human and financial resources for evaluation

Evaluations can be expensive in terms of consultancy costs and staff inputs. At the same time, evaluation can help to save money by increasing efficiency and impact. ILO has in recent years invested significant resources in different types of evaluations, meta-analyses, other knowledge products as well as the EMCP. However, the expansion of project portfolios and volumes, and the demand for more complex evaluations (such as impact evaluations) and rigorous follow-up, will impose major strains on the already limited evaluation capacity, especially at regional level. The evaluation manager system is fragile due to its optional nature and trained staff members often lack knowledge on evaluation theory. This lesson learned is linked to conclusions 1, 3 and 6, especially with regard to the status of the REOs, the underinvestment in M&E specialists in the field, the need for further capacity building of evaluation managers, and requirements on enhancing evaluation budgets, rosters, networks and procurement processes. Applicable across ILO units, the lessons learned can be operationalized through the implementation of the overall recommendations, as well as recommendations 1, 2, 10, 11 and 12. The resourcing of the evaluation function would require particular attention in the updating of both the evaluation policy and evaluation strategy.

Lesson 3: Evaluations that are fit for purpose

The purpose and intended use of an evaluation should be carefully considered in evaluation design. Flexibility in choosing evaluation approaches is required to identify the "best fit" methods and ensure that credible evidence is produced and evaluations questions and criteria covered. As revealed by the IEE, there are sometimes unrealistic expectations of evaluation consultants' expertise on a given subject, the depth of technical analysis in evaluation reports, and the ability of evaluations to bring new knowledge and identify impact, especially given the short timeframe of evaluations, and weaknesses in project M&E systems. This calls for a more participatory evaluation process whereby evaluation managers and consultants have opportunities to discuss the evaluations' purpose, use, priorities, methods and limitations at an early stage, and where adequate time and resources are provided for method development and data collection. As such, this lessons learned is directly related to Conclusion 2 emphasizing the benefits to be gained from increasing the diversity of evaluations and expanding participatory methods, and Conclusion 7, which emphasizes the need for promoting the use of strategic, joint, thematic and DWCP evaluations. This is captured in recommendations 5 and 7, and applies to EVAL, PARDEV, PROGRAM, and technical departments and ROs in particular.

Lesson 4: Involving partners in evaluations

Improving the evaluation and accountability culture warrants close engagement with and capacity building of governments and social partners. With the international aid effectiveness agenda and the SDGs, there is an imperative to strengthen the capacity to supply and demand evaluation at national level and align evaluations with national policy cycles. ILO partners are key informants and intended users of evaluations, but their demand for evaluations and capacity to implement recommendations generally remain low. Involving partners more in the design, implementation and follow-up of evaluations can help to strengthen their interest in evaluation and facilitate the use of evaluations as a tool for social dialogue. This requires that learning is an explicit objective of the evaluation process, and that time and resources are set aside for capacity development. Mainstreaming social dialogue and deepening constituents' participation is important for the credibility of evaluations (as highlighted in Conclusion 2) as well as their use (Conclusion 4). The lessons learned is reflected in Recommendation 4 with regard to the need to support capacity building of national constituents, Recommendation 5 with regard to the promotion of participatory evaluation methods, and has implications for the updated evaluation policy and new evaluation strategy (recommendations 11 and 12). Consequently, it is broadly targeted at EVAL, technical departments and ROs, MULTILATERALS, PROGRAM, PARDEV, DG and GB.

Lesson 5: Evaluation management response follow-up

A formal management response system can help promote an appropriate and timely use of evaluations, and encourage Organization-wide learning. Reporting on and monitoring the implementation of management responses is crucial in ensuring adequate follow-up, identifying actual changes made, and preventing mistakes from being repeated. While ILO's management response system has been improved in recent years, it is not clear to what extent this has contributed to promoting the use of evaluations. Management responses and reporting on follow-up tend to be vague and/or incomplete, and there is no system to trace to what extent recommendations are actually used and integrated into policies, strategies, programmes and projects. More time and resources could be spent on discussing findings and recommendations with intended users and assessing the qualitative impact of evaluations. This lesson learned is associated with Conclusion 5, which calls for the strengthening of the system in following up management responses at the decentralized level, and the practical implications, which are captured in Recommendation 8. This lesson learned is applicable to ROs and technical departments in particular.

Lesson 6: Effective communication

Good communications is critical to the effective use of evaluations and the strengthening of the evaluation culture. Adequate investments should be made in developing and delivering messages and presenting evidence in a clear, easily understood way that is immediately accessible to all stakeholders. The importance of communications is recognized by ILO and inter alia reflected in the efforts made to produce meta/synthesis studies and other knowledge products, and the development of the *i*-Track database. At the same time, evidence suggests that more could be done to ensure that evaluations are shared with different audiences and with the right people, in an appropriate form, and through relevant channels. When an evaluation is being planned, a dissemination and communications strategy should be devised based on the needs and interests of intended users. As recommended in this report, institutional links could also be created between EVAL and DCOMM to facilitate the sharing of information. This lesson learned is directly linked to Conclusion 4 and Recommendation 9, which call for a more targeted communications strategy and communications products developed especially for senior ILO managers and national constituents. This lesson learned is especially applicable to EVAL and DCOMM.

ANNEX A: TERMS OF REFERENCE

April 2016

**Terms of Reference
For the**

**INDEPENDENT EVALUATION
OF ILO'S EVALUATION FUNCTION – 2016**

INTRODUCTION

1. As mandated by the Governing Body, the ILO will launch an Independent Evaluation of its Evaluation function (IEE) in 2016. Evaluation function refers to evaluation structures, processes and activities carried out to implement ILO's evaluation policy and strategy. This will be carried out five years after the GB approved results-based strategy 2011–15: *Evaluation Strategy – Strengthening the Use of Evaluations*.[1]

2. The mandate and proposed activities of evaluation in ILO are set out in the GB document: *Evaluation: a new policy and strategic framework for evaluation in the ILO*.[2] This document proposes that *"the new evaluation policy and strategy be evaluated after five years to assess its impact on the function and performance of the Office"*.[3]

3. An Independent External Evaluation was therefore undertaken in 2009 with findings, conclusions and recommendations presented in 2010 to the GB,[4] including the Office's response and plan of action for the implementation of recommendations. This has guided the evaluation strategy and subsequent work.

4. Reporting on the implementation of the strategy and other activities of the evaluation function are documented in the Annual Evaluation Reports of 2010–11, 2011–12, 2013–14, 2014–15 and related documentation.

---

[1] GB.310/PFA/4/1(Rev.), Geneva March 2011
[2] GB.294/PFA/8/4, Geneva, November 2005
[3] ILO: A new policy and strategic framework for evaluation in the ILO, GB.294/PFA/8/4; paragraph 46.
[4] GB.309/PFA/5/5, Geneva, November 2010

5. An external system-wide assessment conducted by the Joint Inspection Unit (JIU) in 2014, placed ILO among the top three UN agencies with a demonstrably relevant and effective evaluation function.

6. The GB in November 2015 approved Recommendation 1 in the Annual Evaluation Report (AER) as: "The Office should ensure that the independent evaluation of the evaluation function is launched by June 2016 at the latest so that it is ready for reporting to the Governing Body in March 2017 using the structure described in paragraph 13 of this report to ensure its independence, credibility and utility".

7. The IEE will focus on the evaluation function within the Organization, taking the 2005 evaluation policy and strategic framework and the 2011–15 Evaluation strategy, which was extended to in November 2015 to cover 2016 and 2017, as the starting point and working towards understanding how the evaluation system operates at the various levels, in order to review the quality and effectiveness in light of the objectives of the Organisation and the appropriate international standards.

8. It will be a follow-up to the 2010 Independent External Evaluation of the Evaluation Function (IEEF 2010) and will use that approach as the basis.

9. The most important rationale for and objective of the IEE is to "look ahead" and provide recommendations for a new evaluation strategy aligned with ILO's Strategic Policy Framework (2018-2021) under preparation and the Sustainable Development Goals approved as part of the 2030 development agenda. This is particularly relevant in the context of ILO's continuing commitment to RBM, which relies on links between monitoring and evaluation, policy formation and budgeting.

10. Recent developments in international development resulting from enhanced focus on development effectiveness, policy coherence, national ownership, impact orientation and in particular the 2030 Agenda - with inclusiveness, equality, mutual accountability, emphasis on evidence based progress reporting, country led evaluation and national evaluation capacity building at its centre - also call for a reflection on the role of evaluation function and implications on its way of working.

BACKGROUND TO THE EVALUATION FUNCTION IN ILO

11. Working to internationally accepted norms and standards inside and outside the United Nations system (being principally those of the UN Evaluation Group and of the Development Assistance Committee of the OECD), the evaluation policy and its 2011-15 strategy aimed to improve and strengthen the practice of independent evaluation in the ILO and systematic self-evaluation of programme performance to together provide comprehensive coverage of all ILO activities laid out in the 2010-15 Strategic Policy Framework (SPF) and the Programme and Budget documents. The evaluation policy; the evaluation strategy; the evaluation operational framework with guidance, procedures and set up by EVAL and involving relevant other parts of ILO; and the implementation of evaluation across ILO constitutes the evaluation function.

12. The objectives of the ILO Evaluation Policy are:

■ Improve Office-wide transparency and accountability for impact of ILO actions to support its constituents;

■ Strengthen the decision making process by the policy organs and senior management based on sound assessment of effectiveness, efficiency, relevance, impact and sustainability of ILO actions; and

■ Contribute feedback for learning and on-going improvement of the ILO's work.

13. The 2011-15 evaluation strategy aggregates under three main outcomes:

    a. Improved use of evaluation by ILO constituents and management for governance;

    b. Harmonized Office-wide evaluation practice to support transparency and accountability; and

    c. Evaluation capability expanded in the form of knowledge, skills and tools.

14. Work on evaluation in ILO is also contributing to the Programme and Budget outcomes on effective and efficient governance of the ILO. These outcomes included in the Programme and Budget are reported on in ILO's Programme Implementation Report and are consistent with the 2011-15 evaluation strategy.

15. Core to the evaluation policy was the creation of a central evaluation unit, which was tasked to instil a culture of accountability and learning through evaluation. EVAL was created in March 2005 within the Management and Administration Sector. Since 2011 it has been shifted to report directly to the Director General and the Governing Body in line with overall responsibility for implementing the ILO's evaluation policy. Since 2014 as part of the DG's reform agenda it was renamed "the Evaluation Office" to reflect its enhanced independence and status. Since the policy's adoption, the ILO's evaluation function has been further strengthened through the release of an updated Office Directive specifying the authority and responsibilities of EVAL (IGDS Number 74 – Version 2, of 27 June 2014). It now has a Director, three evaluation officers and a knowledge management officer.

16. The responsibilities of EVAL were to establish an Office-wide evaluation network and to facilitate progress towards harmonizing evaluation policies and practices, including monitoring of adherence; developing rules and guidance for ILO evaluations; improving coherence and complementarity between evaluations; facilitating the generation and use of evaluation information; and developing networks, visibility and credibility for the ILO in the area of evaluation.

17. The operational framework of the policy is intended to ensure transparency and independence of the evaluation function. The framework serves different needs and is aimed at different levels to align with major programming and resourcing processes in the Office. Responsibility for implementation of some of the evaluation types was to lie within line management structures (self-evaluation), while others were to be managed by evaluation focal persons in sectors and regions, with oversight provided by an independent central evaluation unit.

18. The evaluation function is characterised by four different types of evaluations:

    - *First,* evaluations of ILO strategies focussing on particular outcomes of major strategies or policies established in the Programme and Budget.[5] Since 2011 ten strategy evaluations at the outcome or institutional level have been carried out to assess effectiveness, efficiency, impact and continued strategic relevance.[6]

    - *Second,* independent country level evaluations in the form of evaluations of Decent Work Country Programmes, carried out as means of systematically reviewing progress and relevance of country level work to national constituents and partners. Since 2011 five DWCP evaluations have been completed with the last three evaluations clustered by sub-region to ensure broader coverage.

    - *Third,* thematic evaluations provided a means for ILO technical programmes to explore in depth the effectiveness and impact of major means of actions. Largely, EVAL has provided advisory support to ILO technical programmes for conducting and resourcing thematic evaluations. In addition to high-level evaluations and thematic evaluations, EVAL has

---

[5] Programme and Budget covered: 2006–2007, 2008–2009, 2010–11. 2012–2013 and 2014–15.

[6] See http://www.ilo.org/eval/Evaluationreports/lang--en/index.htm for strategic and country programme evaluations

conducted meta-evaluations and synthesis reviews which have been presented as reference material for recurrent discussions in the International Labour Conference.

■ *Fourth,* EVAL has provided support and oversight as required for ILO independent project evaluations (averaging a total of 40 per year) with support from departments and regional / country offices for evaluation management, in particular through an Evaluation Manager Certification Programme designed to build evaluation management capacity separate from line and project management. All these project evaluations have been externally appraised for quality and credibility.[7]

19. The work of the evaluation function in the ILO also includes impact evaluations by technical departments with technical support from EVAL as requested, the publication of Think Pieces of evaluation related issues, evaluability assessment of large projects and inputs to the ex-ante appraisal process of development projects and programmes, including Decent Work Country Programmes.

20. In addition, the evaluation function supports the Office in carrying out self-evaluations including country programme reviews, organizational reviews and project-level self-evaluations.

21. Some other aspects of the ILO's evaluation function deserve mention:

   a. Submission of an Annual Evaluation Report to the Governing Body, which reports on the implementation of the 2011-15 evaluation strategy and provides an overview of effectiveness and performance issues within ILO and recommendations for follow-up. These recommendations are tracked through a rolling Plan of Acton shared with the GB.

   b. An internal Evaluation Advisory Committee to oversee and promote institutional follow-up to evaluation recommendations has been established.

   c. The Evaluation Office has developed tools and guidance for evaluation as well as conducting targeted training, as a means of improving the learning function of evaluation activities.

   d. A close connection exists between the evaluation function and the results-based management framework, with the former providing lessons learned information for the latter.

   e. Evaluability assessments have been launched of projects with budgets of five million and above, which highlights issues related to efficiency by which programmes can be evaluated.

   f. A database has been developed to monitor and track project evaluations and follow-up to recommendations. This database is also the means of developing a composite schedule of upcoming evaluations, and documenting follow up to evaluations.

   g. Full-time regional evaluation officers are now in place as well as part-time evaluation focal points in HQ Departments.

   h. An ILO evaluation network consisting of EVAL staff, Regional Evaluation Officers and Departmental Evaluation Focal Points is operational, carrying out tasks in the planning and management of decentralised evaluations. The ILO evaluation network participates in the biennial EVAL workshops that review and further develop the work of the evaluation function.

   i. A Community of Practice of staff interested and knowledgeable about Impact Evaluation carries out peer reviews of IE activities in ILO and serves as forum for exchanging experience on IE. Based on a decision in an Annual Evaluation Report that requires EVAL to work towards standards and consistency in IE work, an initial three tier strategy was developed focusing on building common understanding, providing technical support through a helpline and quality appraisal of implemented impact evaluations.

---

7   See http://www.ilo.org/eval/newsletter-and-think-pieces/WCMS_464253/lang--en/index.htm and http://www.ilo.org/eval/newsletter-and-think-pieces/WCMS_329163/lang--en/index.htm for most recent quality appraisals.

j. Quarterly newsletters and a minimum of two EVAL Think Pieces per year have been produced since 2011. An Evaluation Communication Strategy has been developed and initial elements implemented, such as a Knowledge Sharing Platform and Quick Fact sheets on key evaluations.

PURPOSE AND SCOPE

Purpose

22. The purpose of the IEE is to provide an independent assessment of the continued effectiveness and relevance of the 2005 policy and the Office's performance against the 2011–15 evaluation strategy. The IEE should also examine the relevance of the strategy (strategic fit) as well as the current arrangements and structures for the evaluation function within ILO, learning from the implementation experiences (involving both centralised and decentralised functions). The IEE should take note of international best practices including the UNEG[8] Norms and Standards and examine the extent to which the ILO has made progress in implementing the evaluation function with respect to the UNEG Norms and Standards.[9]

23. The IEE will have two main objectives:

(i) to extract lessons from the implementation of the three strategic outcomes of the 2011–15 results-based evaluation strategy, and the targets and milestones added to accommodate the transitional SPF 2016–17; and

(ii) to contribute towards the alignment of the next evaluation strategy to the SPF 2018–21

24. Recommendations are to be made, inter alia, in relation to the independence, credibility and utility of the evaluation function in ILO, and how these might be improved, and the role and contribution of evaluation within the ILO organisational reform, Strategic Policy Framework for 2018-2021, the Decent Work Country Programmes, the follow-up to the new Declaration on Social Justice for a Fair Globalisation and the 2030 Development Agenda with the Sustainable Development Goals (SDG). Recommendations on the operational framework and appropriate evaluation architecture is also required.

25. The IEE is also expected to provide a basis for improved accountability, learning of lessons, leadership and decision-making in the context of the ILO's continuing commitment to Results-Based Management (RBM). Although it is assumed that the evaluation recommendations will not translate into real increase in regular budget resources for the evaluation function, it is expected that potential resource implications are included in any recommendations put forward and that recommendations are prioritised as appropriate.

26. The main client for the IEE is the Governing Body, which is responsible for governance-level decisions on the findings and recommendations of this evaluation. Other key stakeholders include the ILO Director-General and members of the Senior Management Team, Regional Directors and ILO donors. The IEE is also expected to provide a basis for improved decision-making by the ILO.

---

[8] The United Nations Evaluation Group (UNEG) is a professional network that brings together the units responsible for evaluation in the UN system including the specialized agencies, funds, programmes and affiliated organisations. UNEG aims to strengthen the objectivity, effectiveness and visibility of the evaluation function across the UN system and to advocate the importance of evaluation for learning, decision making and accountability.

[9] The UNEG Norms for Evaluation in the UN system seek to ensure that evaluation functions within the UN follow agreed-upon basic principles. They provide a reference for strengthening, professionalizing and improving the quality of evaluation in all entities of the United Nations system. The UNEG Standards build upon the Norms, and are intended to guide the establishment of the institutional framework, management of the evaluation function, conduct and use of evaluations.

27. The IEE should, based on its findings and conclusions, propose a detailed outline of the new evaluation strategy, with corresponding Results Based Framework, to be put in place as of 2018 in line with the SPF. If the findings of the IEE suggest the need for a revision of the evaluation policy, this should be outlined.

Scope

28. The IEE will cover the period from the initiation of the new evaluation strategy 2011–15, including the transitional extension covering 2016–17 as relevant to the evaluation, taking into account the 2005 Evaluation Policy and the recommendations from the Independent Evaluation of the Evaluation function in 2010. Appropriate reference to work since 2005 will be made when needed.

29. The IEE will examine the following aspects:

Independence

■ The quality of the evaluation function in the ILO, with regard to its independence, credibility and utility as assessed against international norms and standards;

■ Structural aspects of the evaluation function in the ILO and whether the current organisational relationships and reporting arrangements contribute to both learning and accountability;

■ Relevance and effectiveness of portfolio of type and nature of evaluations (centralised versus decentralised; independent v. internal and self-evaluations; etc.);

■ The relevance and effectiveness of the reporting arrangements internally to the Director General (DG) and Evaluation Advisory Committee and externally to the Governing Body;

Credibility

■ Relationships regarding how the evaluation function and responsibilities are meeting the needs of tripartite constituents supporting the ILO governance process; and ways in which these relationships could be strengthened;

■ The extent to which ILO evaluations respect UNEG Norms and Standards and relationships with evaluation units and mechanisms within the United Nation system, donor agencies, and other relevant evaluation networks;

■ Adequacy and relevance of capacity and competencies for evaluation, and the use of evaluation techniques and methodologies;

Utility of the evaluation function

■ The extent to which the products and services of the evaluation function are useful;

■ The extent to which the evaluation function contributes to informing the strategic directions, policies, programmes and projects of the ILO, including the focus on results-based management (RBM), and how to make it more effective in this respect; and

■ The extent to which evaluation results are incorporated and used in follow-up activities (at strategic, policy, programme and project level) and within the knowledge management strategies of the ILO, and disseminated to wider audiences.

Enabling environment

■ Organisational relationships of the evaluation function including EVAL's mandate, scope and work, including its relationships to the various evaluation operations within the ILO, and respective roles and utility of centralised and decentralised evaluation activities as well as of independent evaluations and self-evaluations;

CONSIDERATIONS ON THE METHODOLOGY

30. In accordance with the ILO evaluation policy, participation of ILO tripartite constituents and both internal and external key stakeholders during all phases of the evaluation process should be ensured. This involvement will be based on suitable methodologies, focussing on interviews, consultation meetings, surveys, and document reviews.

31. EVAL will prepare all relevant background documentation; prepare a list of proposed key stakeholders; carry out initial analysis and presentations of what is considered key performance data; consolidate the on-going reflection and review exercises such as the discussions during the biennial EVAL workshops and rolling plans of action; and update key background documents from IEEF 2010 such as list of references in documents, list of mandates, list of guidance material, list of relevant monitoring and evaluation instruments.

32. An analytical framework will be used that, based on the UNEG Norms, develops a set of criteria with corresponding evaluation questions that takes into account:

   a. independence, credibility and utility as the normative framework;

   b. specific policy objectives and outcomes of the evaluation strategy:

   c. key function/overall purpose of accountability, planning and learning.

33. The details of the analytical framework and the methodology will be elaborated by the independent evaluation team on the basis of these Terms of Reference (TORs), guided by the overseeing entity, using the materials and related support provided by EVAL as the Secretariat and documented in an inception report. It is expected that the evaluation team will apply mixed methods which draw on both hard and soft evidence and involve multiple means of analysis. These could be, but is not limited to:

   ■ Reviewing appropriate evaluation policy-related documentation using UN Norms and Standards as quality benchmarks; including performance reports, budget information, internal governance documents, etc.

   ■ Conducting a review of current evaluation arrangements and practices

   ■ Reviewing an inventory of the portfolio of evaluation work done since 2011, including the extent to which these have been managed according to ILO evaluation policy and guidelines.

   ■ Reviewing a sample of evaluations reflecting regional diversity, centralised vs. decentralised reports, technical themes, and type (project, country programme, global strategy, thematic) against recognized evaluation good practices;

   ■ Reviewing the electronic systems supporting the evaluation function to facilitate transparency, accountability and knowledge sharing;

   ■ Reviewing evidence of follow up to evaluation recommendations and use of lessons learned by ILO management and other stakeholders as relevant;

   ■ Interviewing key stakeholders reflecting a diversity of backgrounds inside the Office according to sector, technical unit, regions and country situations, and representing both subjects and users of evaluations;

   ■ Carrying out visits to 4 Regional Offices (Lima, Bangkok, Beirut, Abidjan) to engage with ILO stakeholders and constituents there and undertake maximum 3 visits to HQ Geneva to engage with ILO key stakeholders, including GB Officers and donor represented in missions in Geneva; The visits to Geneva will cover the initial scoping and interviews with key stakeholders; presentation of the inception report; and discussion on revised evaluation strategy and management response.

■ Carrying out a series of electronic surveys both with Office staff and Governing Body members and prepare regional and country case studies, including based on visits to the Regional Offices and other field offices (optional).

MANAGEMENT ARRANGEMENTS

The GB in November 2015 approved the following management structure to ensure independence, credibility and utility of the IEE:

■ a management panel (EAC) to oversee the independent evaluation of the evaluation function and to ensure that it is conducted in a manner that enhances the utility of its findings for the ILO while also maintaining the independence of the process;

■ an evaluation team of external consultants (the team); and

■ a support secretariat composed of EVAL staff (the secretariat).

Management Panel (Technical Committee)

35. The EAC decided that a Technical Committee (TecCom) should serve as the management panel. The TecCom has been identified by the EAC, consisting of the head of ILO Office of Internal Audit and Oversight, an ILO official based in the region with a strong evaluation background but not currently involved in evaluation activities in the ILO and a chief of a branch of a technical department with strong development cooperation elements.

36. The TecCom will oversee the IEE process on behalf of the EAC through updates to the chair of EAC in order to maintain its independence. The responsibilities of the TecCom is to provide the ILO Governing Body with assurance that the IEE is conducted independently and transparently, and that it complies with established procedures and standards, including with regard to the bidding process for the selection of the evaluation team and the approved TOR. The TecCom will meet virtually and through electronic means. Specific responsibilities of the management panel, timing and level of effort will include:

Task	Timing
Review the work plan and terms of reference for the evaluators	April-May 2016
Seek feedback from the Governing body on the TOR through the Chair of the EAC.	April-May 2016
Review the profiles of the members of the evaluation team and the technical proposals presented to the competitive bidding process leading to the selection and contracting of the team of external evaluators	May-June 2016
Oversee the evaluation process, ensuring that the evaluators have access to the necessary materials and relevant staff, and other required facilities;	Ongoing
Review the draft and final evaluation report to ensure that it complies with the agreed TOR;	October 2016 and December 2016
Be a focal point for ILO's management comments and response on the draft report; and	Ongoing
Provide, if required, oral or written feedback to the Governing Body on the independence, transparency and credibility of the process, including compliance with established rules and procedures.	

Independent Evaluation Team

37. The IEE will be carried out by a team of senior evaluation consultants, experienced in similar exercises and independent from the ILO. The consultants will be selected through a competitive bidding process in accordance with normal ILO procurement procedures. A team of up to three consultants will be identified. ILO estimates that the total effort required will be 100 working days.

38. The TecCom will review technical proposals and screen candidates based on relevant evaluation experience; independence (i.e. no current or past close working relationships with the ILO), familiarity with the ILO and its mandate, and understanding of UN evaluation norms and standards. Travel and per diem cost for field visits will be calculated separately. Bids will be invited from companies and teams providing consultancy services. The call for bids would be published through various sources so as to attract a broad range of responses.

39. The combined profile of the evaluation team should include the following in a suitable combination (technical criteria for the evaluation team):

    a. The evaluation team will have a senior level internationally recognised evaluation expert as the team leader. The team leader should have a minimum of 15 years documented experience in UN system international development evaluation. The other team members depending on role require a minimum of 5 to 10 years of relevant experience.

    b. Experience with the full range of evaluation types and levels from strategic to specific project level at both global, regional and country level; Experience with evaluations carried out considering OECD/DAC and UNEG norms and standards; Quality assurance and organisational review experience; Impact evaluation; theory-based evaluation and mixed methods experience; Results Based Management; UN system development planning, programing, monitoring and evaluation approaches experience; and knowledge of recent developments in international development will be an advantage.

    c. The following language skills will be required (English, French and Spanish)

    d. No prior evaluation engagement with ILO within the time frame covered by the evaluation (2011 to 2016).

Evaluation Secretariat

40. The Evaluation Office (EVAL) will provide support services to the TecCom and to the evaluation team as requested, but will not be involved in the actual evaluation process.

Request for proposals

41. Interested companies and teams are expected to provide a short (3 to 5 pages maximum) technical proposal with ideas on how to make these terms of reference operational. The technical proposal should include a preliminary analytical framework, including evaluation criteria, indicators and benchmarks with proposed associated evaluation questions, as well as a brief description of the methodological approach to be followed. The proposal should also describe the division of labour among team membersand other operational considerations deemed appropriate.

42. Interested bidders should also submit the CVs of all the team members detailing their role in the team, their evaluation experience and including references,

43. Finally, all bidders should submit a financial proposal including only the total number of working days by consultant and the consultants' fees (per day) in a separate envelope, following ILO's established procurement guidelines.

OUTPUTS

44. The following written outputs in English will be produced by the independent evaluation team and submitted to the TecCom, who will ensure input and review as required:

    ■ An inception report detailing initial findings and proposed methodology, including key questions to answer (July 2016); this will include the analytical framework, relevant survey

instruments (if any) and interviews guides. A detailed work plan with distribution of responsibility within the team will also be included.

■ A detailed draft evaluation report (maximum 50 pages plus annexes) drawing conclusions and making recommendations based on factual information and well-reasoned judgement based on credible analysis of sources and documentation consulted (October 2016). This will include elements of the new strategy. The draft evaluation report will be circulated to key stakeholders for comments on factual accuracy and requests of clarifications.

■ An executive summary for presentation to the Governing Body of summary findings in March 2017 (draft required by latest mid-February 2017).

■ A final evaluation report to be posted on the ILO website and disseminated to key stakeholders, considering comments on the draft report; if some of the comments are not taken into consideration, the team should submit separately a reasoned explanation of the causes.

45. EVAL will produce as required a series of internal and supporting outputs such as an annotated bibliography, list of evaluations and other annexes as required by the evaluation team.

PROVISIONAL PLAN OF WORK AND TIMETABLE

46. The proposed evaluation time frame is from July 2016 to March 2017.

Dates	Events
May 2016	Preparation of TORs and consultations with GB members TOR
May 2016	Request for proposals
Jun 2016	Selection of Consultants
July 2016	Inception Evaluation Report
July-Sep 2016	Evaluation process
Oct 2016	Draft Evaluation Report reviewed against TOR
Nov-Dec 2016	Briefing of constituents based on which management response can be formulated
Dec 2016	Report Finalised
Jan 2017	ILO Management's response to the evaluation finalized
Mar 2017	Presentation of summary IEE report to GB
Mar 2017	Public dissemination of full IEE report

KEY REFERENCE DOCUMENTS

(See http://www.ilo.org/eval/lang--en/index.htm for full list of ILO evaluation documents)

- ILO Evaluation Policy, 2005 (http://www.ilo.org/public/english/standards/relm/gb/docs/gb294/pdf/pfa-8-4.pdf)
- ILO policy guidelines for results-based evaluation (2012, 2nd edition July 2013) (http://www.ilo.org/eval/Evaluationguidance/WCMS_168289/lang--en/index.htm)
- Results-based strategies 2011–15: Evaluation strategy – Strengthening the use of evaluation (2011) (http://www.ilo.org/wcmsp5/groups/public/---ed_norm/---relconf/documents/meetingdocument/wcms_152025.pdf)
- Evaluation in the ILO. Director General's Announcement IGDS No. 75 (V.2) (2011)
- The ILO Evaluation Unit. Office Directive IGDS No. 74 (V.1) (2009)
- Evaluation Advisory Committee. ILO Circular No. 245, Series 2 Đ (2008)
- Annual Evaluation Reports (http://www.ilo.org/eval/Evaluationreports/annual/lang--en/index.htm)
- ILO i-eval THINK Piece, No. 10 - February 2016; Quality assessments of ILO project evaluations: What are the next steps to better evaluations? (http://www.ilo.org/eval/newsletter-and-think-pieces/WCMS_464253/lang--en/index.htm)
- ILO i-eval THINK Piece, No.8 - December 2014: External Quality Appraisal: Implications for evaluation quality and utilization (http://www.ilo.org/eval/newsletter-and-think-pieces/WCMS_329163/lang--en/index.htm)
- ILO, October 2009: Independent external evaluation of International Labour Organization Evaluation Function (report prepared by an external evaluation team) (http://www.ilo.org/eval/Evaluationreports/WCMS_182078/lang--en/index.htm)
- United Nations Evaluation Group (UNEG), April 2005: Standards for Evaluation in the UN System ((http://www.unevaluation.org/document/detail/22)
- United Nations Evaluation Group (UNEG), April 2005: Norms for Evaluation in the UN System ((http://www.unevaluation.org/document/detail/21)