



International
Labour
Organization

Evaluation
Office

► i-eval THINK Piece, No. 25

Impact evaluation: Ex-post review and way forward

Centre for Evaluation Development (C4ED)

ILO Evaluation Office

October 2023



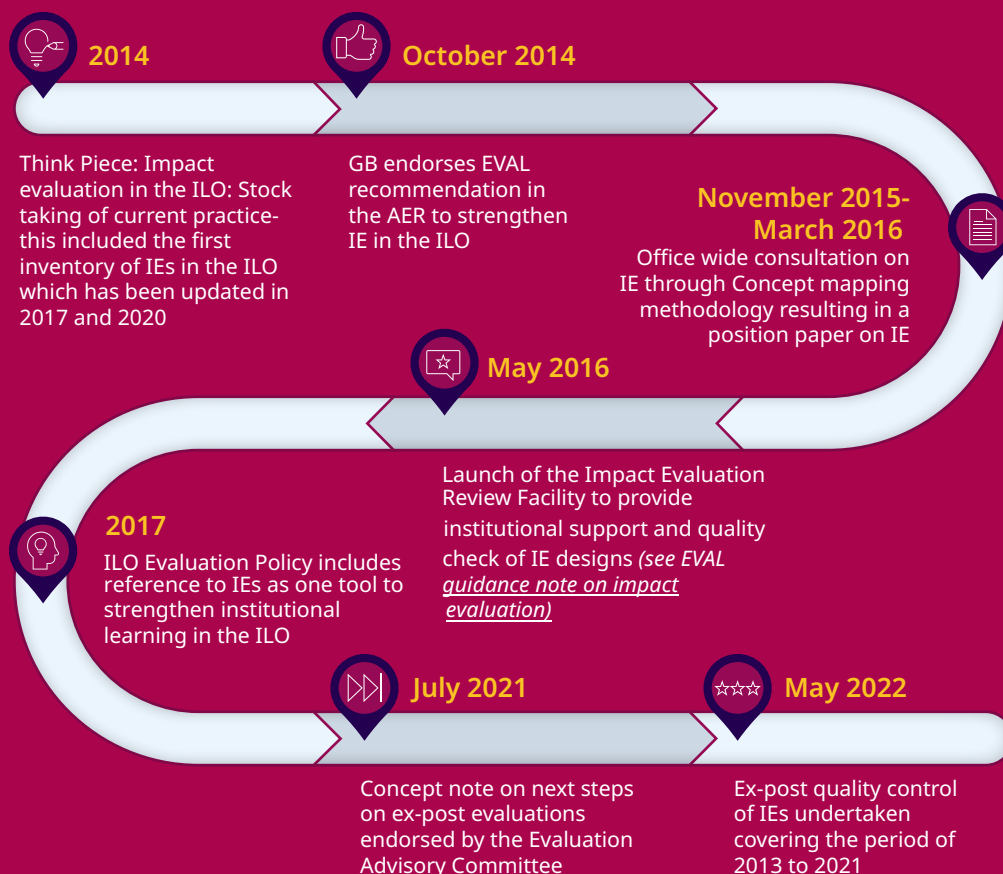
The data and findings presented in the Think Piece are derived from the report, “Ex-post Quality review of Impact Studies and Evaluations carried out by the ILO, 2013-2021¹” prepared by C4ED. The Think Piece also contains reflections by the ILO Evaluation Office (EVAL) on the way ahead for impact evaluations during the current ILO Evaluation Strategy period 2023-25. The Think Piece was drafted by Ms. Naomi Asukai with inputs from Mr. Guy Thijs and Ms. Amanda Mack of the ILO Evaluation Office and the team at C4ED but has not been subject to professional editing.

1 Carried out by the Center for Evaluation Development (C4ED) Dr Atika Pasha, Irene DeJuan, Kevin Hempel.

Introduction

Brief history of impact evaluation in the ILO

The ILO Evaluation Policy of 2017 promoted the use of impact evaluations to support organizational learning and for knowledge building on effective policy interventions. Certain departments (Employment, Enterprises, Fundamentals (IPEC), Better Work), within the ILO have been undertaking impact evaluation or impact assessments² prior to the 2017 policy. EVAL first began work on impact evaluations in the ILO in 2009 and developed the first version of the guidance note and checklist on impact evaluation in 2013. Since then, EVAL has implemented several activities to strengthen ILO's work on impact evaluation. Some key highlights include:



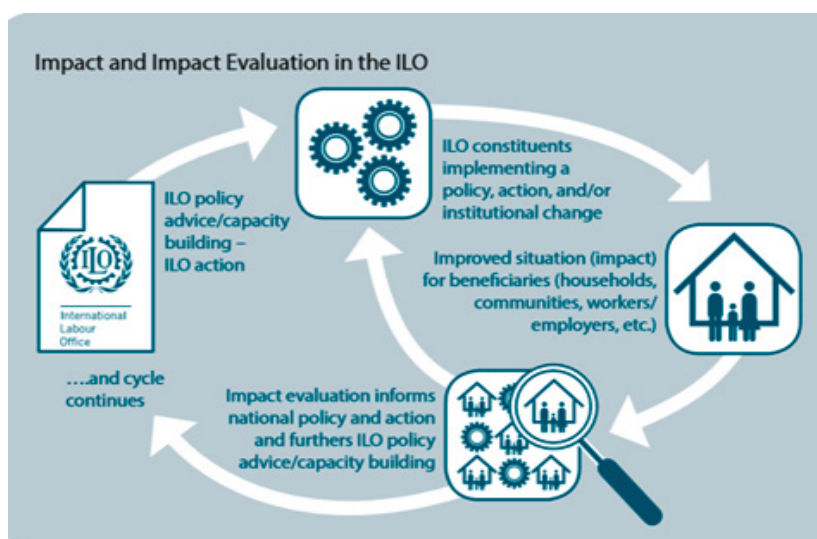
² An inventory of impact evaluations in the ILO is undertaken periodically by EVAL to have an overview of impact evaluations conducted in the various departments.

Role of EVAL

Past ILO evaluation policies and the current policy 2022-25 identifies the Evaluation Office (EVAL) as the entity to promote impact evaluation and assessments in the ILO through:

- Providing guidance on IE methods
- Technical review of IE methodologies
- Quality appraisals of completed IEs

Figure 1: [Guidance Note 2.5: Impact Evaluation](#) page 2



ILO Evaluation Policy 2022-25

While the previous Evaluation Strategy decentralized impact evaluation responsibilities to the regions and departments, the current strategy continues that approach with some important changes. The current strategy provides a more proactive role to EVAL and the departments and regions to jointly identify impact evaluation needs as part of the Criteria-based Integrated Evaluation Planning process (CIEPS). It also reiterates the role of the Impact Evaluation Review Facility (IERF) for review of impact evaluations carried out in the ILO at various stages of an impact evaluation. Furthermore, a quality assessment framework for impact evaluations is to be operationalized and applied ex-post to impact evaluations undertaken by ILO.

Definition of impact

In the ILO, the definition of impact follows the UN Evaluation Group (UNEG) definition on impact evaluations: they aim to assess “the extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended higher-level effects”. A fundamental element of impact evaluation is establishing cause and effect chains to show if an intervention has worked and, if so, how. It aims to determine the level of attribution that can be given to specific factors – including policies, programmes or other interventions, for observed and measured changes. It aims to answer the fundamental question of how Cause C influences Effect E. There is usually a sequence of “cause and effect” mechanisms, so effect 1 becomes cause 2 resulting in effect 2 and so on. This can also be expressed as the sequence: input, output, outcome, goal.

Results of an ex-post review

As part of its mandate on evaluations described above EVAL worked with the Center for Evaluation Development (C4ED) to undertake a quality review of impact studies and evaluations carried out by regions and departments during the period of 2013 to 2021 in order to identify the current areas of strength and weaknesses and to develop a baseline for the current state of impact evaluations in the ILO. Through the review, it was also proposed to develop a framework and quality assurance method for ex post reviews of impact evaluations in the ILO.

Methodology

An impact evaluation quality assessment tool was developed by C4ED through document reviews of ILO and C4ED tools and reports, as well as international best practices. The IE quality assessment tool included a weighing scheme and rationale for calculating the overall scores of the reports. Eighteen impact evaluation reports were selected through a modified stratified random sample to test the first version of the QA tool. In addition, a few key informant interviews were carried out with EVAL and technical departments who were involved in the commissioning, management and/or dissemination of impact evaluations from the sampled studies.

Findings

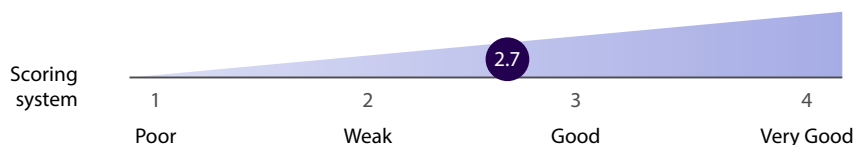
The existing ILO IE inventory has a total of 74 studies, with a significant spread and diversity. Reports correspond to 9 thematic areas (Research, Employment, Enterprises, Better Work, Fundamentals, Work Quality, SME, Impact Insurance, and Soc Pro) and included a vast diversity in terms of methodology, as follows: 35 Randomized Control Trials (RCTs), 24 Quasi-experimental designs (QEDs), seven alternative designs without a counterfactual (i.e., process tracing, tracer studies, etc.), three synthesis reviews and meta-analyses and four studies of unknown/unclear designs. The studies included a large geographical spread (Africa (30), Asia (24), Latin American and Caribbean (9) and Multi-country/Other regions (11)).

However, the review of the ILO IE inventory also helped identify several drawbacks to the database. It showed that the current inventory is in need of updating, to ensure the veracity of information included therein, as the stages of the evaluation or the method used were incorrectly entered or completely missing in several cases. Updating this database appeared to be a self-reported voluntary exercise, where much of the information that would be helpful for establishing a baseline on the impact studies was not available or correct. In its current form the inventory does not effectively capture the stages of the evaluation to which the reports correspond (design, inception, baseline, midline and endline), and also does not include a clear categorisation for quantitative, qualitative or mixed-methods studies.

The 18 reports that were reviewed from this database were also able to provide interesting insights into the quality, utility, and relevance of ILO's IE:

- **Overall, ILO's impact evaluations are in moderate compliance with global quality standards** although there is a significant heterogeneity in the quality of these evaluations. On a scale from 1-4 (1 as "poor" and 4 as "very good"), the scores ranged from 1.7 (between poor and weak) to 3.6 (good and very good). The average score across the sampled studies was 2.7 (rating of 3 considered to be "good"). This suggests that providing stronger guidance and quality assurance would help bring the overall quality of IE in the ILO to a "good" rating.

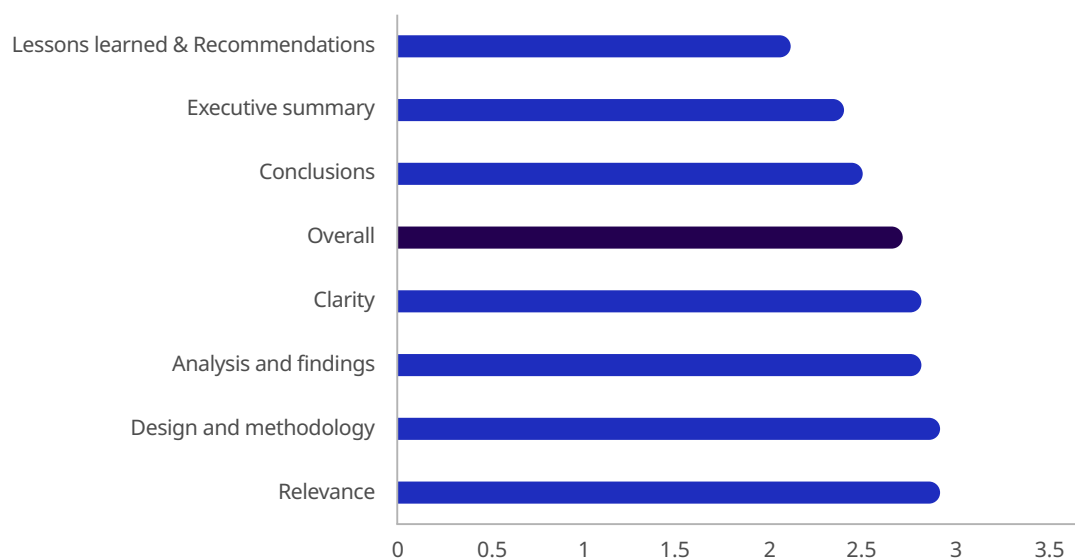
Figure 2. Overall quality of ILO's impact evaluation reports (sample of 18 studies)



- The robustness of methodology can be considered a relative strength of ILO's impact evaluation reports. The majority of the studies were designed using a robust methodology, and also received a good score in the implementation and internal validity of the design. However, a larger focus on (the reporting of) ethical standards in evaluations (score of 1.5), handling of multiple causal factors (score 2.2), managing data quality during and after the data collection (score 2.2), as well as attention to the limitations and constraints of the methodology (score 2.3) is needed. Moreover, the lack of a clear evaluation framework and a Theory of Change limited the guidance on the specific short and longer-term outcomes to actually measure change.
- **Similarly, the analysis and findings sections in ILO's impact evaluation studies were typically robust.** Common strengths across ILO reports included a clear and comprehensive presentation of findings, which were mostly substantiated by evidence derived from the evaluation. Some limitations that were observed were the lack of identification (and measurement) of unintended outcomes (score 2.3), the limited use of mixed methods or multiple data sources to triangulate the findings (score 2.4), and the vague (or missing) reference to the generalisability of the findings to other contexts and populations (score 2.4).
- **Many of the reports are not able to assess the sustainability of the outcomes,** and in fact many do not even find an effect, given the short time period between the intervention and the final evaluation (or the lack of adequate deliberation on long term outcomes). Future evaluations would benefit from evaluation designs that are more aligned with the timeline of the project, without hampering the validity of the findings. This overall finding is related to the need to include a dedicated section on evaluation framework or theory of change, which was not present in all reports.
- **Relevance of the evaluations is relatively high.** The sampled reports performed best in presenting the context and scope of the evaluation, identifying the existing knowledge gaps and the setup of the intervention. They also generally identified (some of) the users of the report and their needs, and frequently tackled cross-cutting issues related to, for example, gender or vulnerable groups.
- **The formulation of lessons learned and recommendations represents the main weakness in ILO's impact evaluation reports.** Most reports lack a dedicated section on lessons and recommendations. If provided at all, lessons and recommendations were mixed with findings and/or conclusions, making it difficult to isolate and implement these clearly. Moreover, even when recommendations were provided, they were not actionable. Instead, they tended to be generic and superficial, thereby impeding the translation of evaluation findings into concrete changes to be made by different stakeholders.

- **Another fundamental weakness of ILO IEs is the common lack of self-standing executive summaries** that adequately reflect the purpose and methodology of the evaluation, key findings, and recommendations to inform decision making of stakeholders. Some studies had no executive summary at all, while many studies used academic-style paper abstracts (with lengths of approximately one paragraph).
- **ILO IE and their reports are not reaching their full potential as instruments of evidence-based policy and programming.** The aforementioned poor formulation of lessons learned and recommendations, as well as the missing or weak executive summaries are the most notable weaknesses across the ILO IE, hindering proper take-up on findings. Further, the fact that IEs do not seem to feature at programming (and budgeting) levels limits the case for their utilisation and further advancement. Institutional capacity constraints are considered key barriers for ILO's staff interests, research ideas and cultivated expertise to further develop, and they also negatively influence the possibility of IEs overall to be featured into a systematic and aligned policy and practice.

Figure 3. Average report ratings by sections (score 1-4 with 3 and above being satisfactory)





Lessons learned and recommendations

Based on the broad findings, as well as insights provided by key informants³, the following lessons and recommendations were drawn:

► **Identify a workable institutional strategy that will allow IEs to be advanced and streamlined at ILO, including a feasible budgeting mechanism.**

EVAL who is responsible for orienting IEs at ILO needs to be provided with both decision-making power to systematise IE policy as well as with adequate financial and human resources. One option that can be explored is the allocation of an Evaluation Fund, a resource pool which interested ILO departments/offices could access via competitive internal mechanisms to implement IEs of their projects and activities, and thus contribute to building evidence on ILO results on a more systematic basis. The existence of this Fund would need to be advocated with donors, and one possibility would be to tie its emergence with allowing ILO to address a key constituents' expectation, improving the understanding of the impact of ILO on decent work conditions globally.

► **Direct efforts at the promotion of a 'culture for evidence' within ILO, where impact evaluations are seen as key instruments for its policy-based work, through:**

► the generation of an established **community of practice** around IEs in ILO, including dedicated internal communication channels (and ILO website), pulling the expertise already present in the organisation (i.e., staff experienced in IEs).

► promote **dissemination** on the usefulness of impact evaluations for ILO work, including examples of how good IEs have contributed to relevant policy work.

► **build capacity at all levels** for the inclusion and understanding of IEs across all levels of ILO, such that budgeting, design and implementation of IEs can be streamlined across various stakeholders.

► **Expand access and usability of EVAL's inventory by including it as part of a dissemination-oriented website on impact evaluation within ILO.**

► **Inventory adjustments:** Ensure that all ILO departments are required to periodically fill in and update the internal EVAL inventory with the evaluations performed.

³ Key informant interviews were undertaken with EVAL and eight ILO focal points from various technical departments, who were involved in the commissioning, management and/or dissemination of impact evaluations from the sampled studies.

- **Provide corporate support and guidance to technical and regional departments for the commissioning, design, management, and quality control of the impact evaluations, to ensure quality standards.**
- **Ensure the inclusion of best practices for impact evaluation from the TOR stage and throughout the implementation of the evaluation, and enforce their compliance,** in particular regarding:
 - The incorporation of a complete evaluation framework, featuring a programme/project Theory of Change (ToC) with expected chain of results and assumptions, distinguishing time horizons (shorter- and longer- term impacts) and specifying which will be measured by the evaluation; evaluation questions to be answer, derived from this framework.
 - The requirement to disclose the full methodology in all reports, including type of design, tools, field protocols, data quality checks, limitations and mitigation measures, discussion of applicability and processes followed to determine internal and external validity.
 - The requirement to link the impact evaluation design with relevant ILO indicators and results frameworks, as relevant to each department.
- **Focus efforts on improving the user-friendliness of evaluation reports to boost take-up and discussion of results.** This could be achieved by:
 - Guiding collaborating researchers leading the IEs (and ILO staff as applicable) on the materials they are expected to produce, including standardised report structures, full-length executive summaries which address all key elements of the study for their target audiences, clear conclusions linked to the interventions' results chain and actionable recommendations.
 - Ensure reports are accompanied by user-friendly Impact Evaluation Briefs and that these are also part of the widely available database.

Quality assessment tool

The ILO IE Assessment Tool (or IEA tool) was developed following the completion of a thorough document review, comprising a wide set of data streams, which included: ILO core evaluation documents -in particular, ILO's Evaluation Policy (2017), Results-based Framework (2018) and Guidance notes on Impact Evaluation (2020); internal and external reviews commissioned by ILO on impact evaluations (e.g., ILO, 2016; ILO, 2014); international guidance for impact evaluations from other peer organisations aligned with international standards (NONIE, 2009; UNICEF Innocenti, 2014; UNEG, 2020); quality assessment tools from various international development agencies (e.g., United Nations International Children's Emergency Fund – UNICEF, the Joint United Nations Programme on HIV/AIDS – UNAIDS and the U.S. Agency for International Development – USAID); as well as internal C4ED's quality assurance tools and lessons learned derived from our experience conducting IEs.

Through this review, key strengths and weaknesses of other evaluation tools for the case of impact evaluation were identified. This was helpful in singling out relevant criteria and internationally recognised standards against which the quality of any rigorous impact evaluation should be assessed, including technical aspects (e.g., design, implementation, and analysis), formal aspects (e.g., structure and clarity), as well as higher-level, policy-related aspects (e.g., policy relevance for ILO, the broader UN system, and the Sustainable Development Goals – SDGs). The IE Quality Assessment Tool review also considered criteria to account for limitations and knowledge gaps noted by the IE reports. In addition, a weighting scheme and rationale for calculating the overall scores of the reports was discussed.

Based on this review, the evaluation tool was developed, following quality standards and covering the key knowledge needs as expressed by the ILO. The objective of the ILO IEA Tool is to offer a uniform and assessable methodology for IE reports and their rigorous quality assessment. The current tool is comprised of an introductory page, with an overview of the object of the evaluation and the details of the ranking system; and seven sections outlining key areas for review. A description of the sections is provided here:



Table 1: ILO IE Assessment Tool Structure and weighting

Sections and their weight	Evaluation guide
I. Executive Summary (10%)	Executive Summaries are key tools for dissemination and programming guidance. Therefore, the necessary key points to fulfil the evaluation requirements should be available and elaborated here. This section should capture the main features that define a good executive summary, i.e., having a good balance between conciseness and clarity , is complete , containing all necessary elements, and the elements included are relevant to inform decision making. A well written executive summary would succeed at informing the reader of the relevance and purpose of the evaluation, its scope, the methods used, and key aspects such as the main findings, conclusions and recommendations derived for future programming and policy.
II. IE design and methodology (25%)	The design and methodology are critical to allow for a robust evaluation of the intervention. In this section, we review: ► The selected design and its justification, including: i) the use of a ToC, ii) the design's appropriateness to address the evaluation objectives, iii) the causal attribution strategy, iv) the clarity, comprehensiveness and appropriateness of the sampling strategy, and v) the way multiple causal factors are considered. ► The methodology developed and justification of any deviations, limitations, and/or constraints, including a description of data collection tools and field protocols and a description of the analysis strategy. We also assess the data quality (i.e., the accuracy and reliability of primary and secondary data utilised, and quality checks carried out), and the data sources. ► Cross-cutting principles, such as ethical considerations and the application of gender and human rights-responsive methodology and tools.
III. Analysis and Findings (20%)	The analysis of the data collected, and the write-up of the findings are also key components of reporting. In this section, we review i) whether the report uses analysis methods that are comprehensive and suitable to the context and objectives, ii) whether the report analyses the fidelity of the intervention's implementation and its take-up, iii) the appropriate use of data sources and triangulation, iv) whether credible findings are constructed from these data against all evaluation questions, on all result levels (outputs/outcomes/impact), considering both short- and long- term effects and an analysis of causality; vi) issues such as external and internal validity and are tackled.
IV. Conclusions (10%)	In this section we review the extent to which conclusions are credible and logically derived from findings . Any conclusion must provide a description of the strengths and weaknesses of the intervention with a reasonable explanation of what short/ intermediate/long-term impacts were identified and for whom, what contribution/s (if any) did the intervention make, and how can these be linked to future programming.
V. Stock-taking, recommendations, and a way forward (15%)	In this section we assess whether the lessons learned, good practices and recommendations logically stem from findings and concisely capture the contexts from which they were derived. Presented recommendations and lessons learned are pertinent, reasonable, actionable , time-bound, and targeted, to be useful for programming, management and policy advice.
VI. Clarity and Comprehensiveness (10%)	We will finally assess the clarity of the report's presentation and its alignment with ILO requirements, the adequacy of the report's length and its completeness (inclusion of all relevant information). Additionally, we review the report's language adaptation to its target audience.
VII. Relevance (10%)	This section considers the degree to which the report's content meets the needs and interests of its intended users . The report should also properly explain the context of the evaluation , providing a relevant description of its background. This section also assesses whether the report covers the evaluation scope thematically, chronologically, geographically and for the selected target groups, and gives reasons for any constraints in coverage. This section also considers if there is an identification of all intended stakeholders and their level of participation and contribution to the evaluation design and its implementation.

The final score of the report is the weighted mean of the sectional scores (see Table 1 for the section weights).

In addition to the quantitative ratings, the tool includes a built-in space for comments next to each item, and once for the overall report. Here, reviewers need to justify and explain the score they give, including by providing examples or page references. The comment space is also meant to reflect the strengths and limitations noted in the report, and to provide suggestions for improvement of each item/section assessed.

The following considerations have been observed for the design of the tool:

- Balance between specificity and comprehensiveness
- Adaptability to methodological diversity
- Integration of ILO IE guidance and information needs
- Balance between sectional and overall review of IE report content:
- Consideration of relevance and usability
- Differential weighting according to importance of sections in determining IE quality





EVAL reflections and way forward

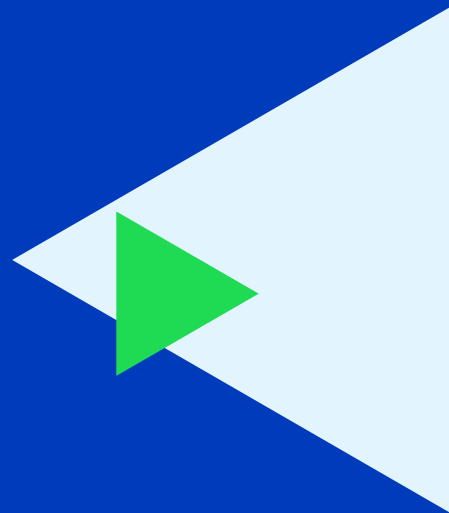
There is a conducive environment to strengthen ILO's work on impact evaluation and EVAL's role in overseeing impact evaluations given the new outputs of the ILO Evaluation Strategy (2023-25). The Strategy clarifies EVAL's role in identifying and articulating how evaluations can support to complement ILO's learning and knowledge agenda and to fill needs and gaps for further evidence-based decision-making information and data. The recent ex-post review study further reaffirms the need for an entity to oversee the quality of impact studies and evaluations and to work together with regions and technical departments to bring the data and information from these IEs to an institutional decision-making level. The review further confirmed the need for further quality assurance and assessments of the myriad of impact studies and evaluations being undertaken by the Office to strengthen the credibility and thus use of these evaluations.

We can see from the average report ratings by section that current ILO impact evaluation studies fall below 3 and that the categories of "design and methodology" and "relevance" are at 2.9 meaning that with more robust inputs at the design stage, ILO impact evaluations could see the required improvement to reach the 3 and above scoring. While EVAL has not been able to fully immerse itself in the review of impact evaluation designs over the prior strategy period, those that have been reviewed by EVAL confirm some of the findings of the ex-post review specifically reviewing the whole context of the intervention under evaluation and the handling of multiple causal factors. Furthermore, as found in the definition of impact evaluation, it is important to have a theory of change to hypothesize the "cause and effect", yet the review found that many of the reviewed reports had not clearly identified an evaluation framework nor was there any articulated "Theory of Change" that would have provided this theoretical chain of results. With more attention paid to these particular areas in the new cohort of planned impact evaluations, the IEs have the potential to provide more robust findings on a particular ILO intervention including why and how something worked in a particular context. This type of evidence would provide senior managers EVAL, as laid out in the current ILO Evaluation Strategy period, will pursue a six-pronged approach to strengthen the quality assessment of evaluations in order to ensure that impact studies and evaluations can serve as important knowledge sources for evidenced-based decision-making.

EVAL, as laid out in the current ILO Evaluation Strategy period, will pursue a six-pronged approach to strengthen the quality assessment of evaluations in order to ensure that impact studies and evaluations can serve as important knowledge sources for evidenced-based decision-making.

- Upgrade the **IE inventory** to become an annual exercise that is put together with the consultations during the CIEPS process with departments and regions. The CIEPS process will identify upcoming evaluations and other learning opportunities and all planned impact evaluations should be identified by DEFP and REOs with the regions and departments. All completed impact evaluations would be included during this process in the rolling inventory list.
- Revitalize the **institutional review board** of impact evaluations pre-implementation ILO EVAL has already set up the **Impact Evaluation Review Facility (IERF)** with limited uptake by the technical departments to use the function for review of their impact evaluation design. EVAL highly recommends that new impact evaluation designs should be reviewed by the IERF or through a **peer review mechanism comprised** of an internal impact evaluation board.
- **Institutionalize two yearly ex-Post quality assessments of impact evaluations** to be conducted by a reliable independent entity based on the tool developed in the context of the review above.
- **Reconstitute the Communities of Practice on impact evaluation** to be created and moderated by EVAL for department and regional colleagues to exchange information and communicate on impact evaluation related issues and questions and to request peer to peer support and share opportunities for trainings.
- **Assess the potential of using the ILO Evaluation Trust Fund (IETF)** modality to serve as a resource pool which interested ILO departments/offices could access via competitive internal mechanisms to implement IEs of their projects and activities.
- **Dissemination and awareness raising** of the services EVAL provides to the department and regional colleagues, with recommendations for the use of the IERF and how the ex-post quality assessment tool will be applied to IEs.





ilo.org

International Labour Organization
Route des Morillons 4
1211 Geneva 22
Switzerland