i-eval THINK Piece, No. 6

Impact Evaluation in the ILO

Stock-taking of current practice

Dr. Achim Engelhardt

October 2014

International Labour Office
Evaluation Office

IMPACT EVALUATION IN THE ILO

STOCK-TAKING OF CURRENT PRACTICE

By Dr Achim Engelhardt

The responsibility for opinions expressed in this document rests solely with the author. The publication does not constitute an endorsement by the International Labour Organization.

This document has not been subject to professional editing.

Table of contents

Abstract

Impact evaluation is increasingly in demand in international development. International organisations aim to better generate knowledge about "what works" through the analysis of interventions and policies. Those organisations and their donors ultimately need to demonstrate concrete changes in the lives of beneficiaries through direct interventions driven by an enabling policy environment.

This think piece is a short review of current impact evaluation techniques being pioneered in the ILO. The review begins with a guiding definition for impact evaluation. Methodologies currently being applied in ILO are explained and discussed. The review also presents some conclusions on the quality and usefulness of this work and then puts forward some recommendations for strengthening ILO future investment in impact evaluation.

Technical programmes surveyed on behalf of EVAL include: Better Work, the Employment-intensive Investment Unit (EMP/INVEST), Small and Medium Enterprises Unit (SME)[1], the Enterprises Department, ILO/AIDS, the International Programme for the Elimination of Child Labour (IPEC), the Social Finance Programme (SFP)[2], the Youth Employment Programme (YEP)[3], and the Research Department (RESEARCH). Given the available time and resources for this study, the focus was placed on methodology and to a lesser extent on the use of impact evaluations.

A rich variety of impact evaluation methodologies emerged from the nine technical programmes surveyed. Methodologies identified included the randomized control trial (RCT) using experimental or quasi-experimental designs, in the past often described as the "gold standard" for impact evaluation. Other methodologies included tracer studies, the use of difference-in-difference techniques in econometric studies, and a theory of change approach.

The review found that methodologies were reasonably successfully designed and applied. However, out of the eight impact evaluations using experimental or quasi-experimental design or techniques close to that standard, three had a focus on the interventions' outcome level rather than on long-term impact related to economic well-being. There were strong commonalities among the methodological constraints across the technical sections surveyed. Some of the main constraints included inadequate budget resources, time constraints, and lack of baseline or monitoring data.

Six evaluations, dating from 2012 to 2014, identified in this review as impact evaluations, i.e. with a focus on long-term results, assessed the direct intervention impact at the beneficiary level. Interventions included access to training and finance for micro entrepreneurs and informal entrepreneurs, access to micro insurance and child labour, education of child labourers and improvements for mainly female workers in the garment sector. Tracking the impact of policy advice provided by ILO to constituents is challenging in the development context. Establishing impact in this area has proven difficult and is therefore an underdeveloped but an important area where future ILO impact evaluation could be strengthened.

Impact evaluation results are used in the ILO for a number of purposes: to provide credible evidence to global and national decision makers of the appropriateness and

1 Now part of Enterprises Department.

2 Now part of the Employment Policy Department.

3 Ibid.

scalability of policy models, the relevance of the ILO's advice at the policy level, and to provide the means by which credible causal change can be measured for beneficiaries. Impact evaluations are particularly costly due to their more complex methodologies, longer timelines, and more robust data collection requirements. In addition to higher levels of expertise required, there is also the additional complication of new terminology and establishing a comprehensive agreement on the use of terms and methodologies. This review noted that the term "impact evaluation" was defined differently among ILO stakeholders, donors and staff, despite clear guidance on the term provided by the Evaluation Office (EVAL).

Several key lessons learned were identified regarding the need for a clear understanding of the definition of impact evaluation and what it entails: being realistic about the significant resource requirements and substantial time investments, as well as the complexities of designing and applying the correct methodologies. EVAL experience also indicates that, as part of the impact evaluation design process, an adequate review should be undertaken by the relevant stakeholders of how the results of the evaluation will be used and by whom.

The author proposes the following recommendations to address the issues identified above:

1. Prior to any impact evaluation, technical programmes should clarify the purpose of the impact evaluation with a clear focus on who would use the evaluation results and for which specific decision-making processes.

2. Technical programmes should undertake an evaluability assessment prior to each impact evaluation, with support from EVAL (e.g. through the use of the existing EVAL quality checklist for impact evaluation);

3. The ILO should start investing more in tracking the impact of advice on normative work and evaluating changes at the policy level following the support of constituents within specific national contexts; and

4. EVAL should further promote the consistent use of the term "impact evaluation" by playing an even more active role with regards to advocacy, knowledge sharing, technical support and quality assurance for impact evaluations.

Introduction

Impact evaluation responds to the growing demand among constituents and international partners for a more credible measurement of impact. This evidence-based impact can support sound decisions about policies and programmes and enable the ILO to actively improve its accountability for long-term results, the performance of its programmes and interventions, all of which is linked to effective use of resources.

Impact evaluation is not part of the compulsory evaluation policy in the ILO and technical programmes wishing to undertake them are primarily responsible for mobilizing the additional financial resources required. The Department Directors take responsibility for completing the impact evaluation according to ILO evaluation standards and guidelines. All impact evaluations labeled independent require the terms of reference (ToRs), the budget, the selection of consultants, the determination of methodologies, and the finalization of the report to be done in coordination with EVAL. EVAL, in collaboration with technical programmes and PARDEV, has developed guidance and quality standards, and offers advisory services for impact evaluation.

Purpose

The purpose of this think piece is to take stock of impact evaluation practice in the ILO with a focus on the methodologies being applied and, to a lesser extent, determine the utility and value of ILO impact evaluations. According to EVAL guidance, impact evaluation is an assessment of the causal effect of an intervention, programme or policy on target groups.

What is impact evaluation in the ILO?

"An impact evaluation aims to determine changes in economic well-being of individuals, households, communities or organizations that can be attributed to a particular intervention, programme or policy."

ILO also applies the OECD/DAC definition of impact as '*positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended*'.

Source: ILO. Evaluation Office. Guidance Note 13: Impact Evaluation, pp. 2 and 4.

EVAL's Guidance Note 13[4] identifies the following three main purposes of the impact evaluation approach in the ILO:

- Technical support to constituents for national impact evaluation
- Organizational accountability and learning from major programmes
- Testing effectiveness of specific innovations

[4] ILO. Evaluation Office: "Guidance Note 13: Impact Evaluation" in, 2nd ed., (Geneva, 2013), pp. 2 and 4.

Over the years a number of larger ILO programmes have experimented with impact evaluations at the intervention level with additional funding provided by donors for this purpose. Most of these programmes have direct beneficiaries but for some programmes engaged in upstream policy work it was more difficult for the impact evaluations to attribute change to ILO efforts.

Impact evaluation can provide valuable feedback to the ILO and constituents to determine the extent to which policy support, programmatic approaches and technical tools are appropriately designed to optimize delivery, results and sustainability. This in turn supports sound decisions about the policies and programmes to be supported directly by the ILO, and enables the ILO to actively improve its accountability for results linked to effective use of resources.

Current evaluation policy in the ILO is largely focused on performance evaluations through mandatory independent evaluations for projects above a certain budget threshold. These are aimed at assessing the effectiveness, relevance and efficiency of the achievement of project outcomes, usually within a well-defined situation and time frame. Very often, these evaluations support both ILO accountability for effective use of funds and organizational learning.

Background: Main findings of the stock-taking exercise

Appreciation of impact evaluation in the ILO

While some technical programmes cite donors as being the driving force behind requiring and funding impact evaluation, others stress their voluntary move toward impact evaluation. In fact, eight out of nine of the groups examined stated that the level of effort invested in impact evaluation is worthwhile; given that the ILO is the main user of the results, see Figure 1.[5]

While interviewed staff mostly expressed appreciation for impact evaluation, a number of practical questions emerged.

- Can changes in economic well-being of individuals, households, communities or organizations be shown within the lifetime of a two to three year intervention?

- Why is ex-post evaluation still underfunded in impact evaluation approaches if changes in economic well-being of individuals, households, communities or organizations are by definition long-term, and often well beyond the end of ILO interventions?

[5] Images are provided courtesy of ILO www.ilo.org, Thony Beliraire, Jenny Vaughn and Joseph Fortin (2014).

Figure 1: Purpose of impact evaluations across the ILO

- How can the ILO ensure that project intervention implementation teams understand and fully appreciate impact evaluation, when, for example, baseline surveys often result in delaying the start of projects?

Main impact evaluation methods used

The evaluation literature distinguishes between different impact evaluation designs[6], and these have been mapped against a sample of impact evaluations undertaken in the ILO in Table 1. This sample includes finalized impact evaluation reports or final drafts shared by technical programmes with the consultant during the review.

Using the designs listed in Table 1, the following impact evaluation methods are used in the ILO: randomized comparison trials (RCTs), tracer studies, difference-to-difference techniques, the regression discontinuity approach and qualitative methods such as theory-based models.

Seven out of the nine technical programmes[7] have used methods related to RCTs which are experimental approaches to impact evaluations.[8] Four technical programmes used tracer studies and three other technical programmes applied difference-in-difference techniques. The need to complement mainly quantitative data emerges in all technical programmes and is addressed through methods

[6] Examples have been taken from M. Bamberger, J. Rugh, and L. Mabry: *Real World Evaluation: Working under budget, time, data and political constraints* (Sage: Thousand Oaks, CA, 2011).
[7] Out of a predefined sample provided by EVAL, nine Departments, Bureaus or Offices shared experience in using impact evaluation methodologies: Better Work, the Employment-intensive Investment Unit (EMP/INVEST), Small and Medium Enterprises Unit (SME), the Enterprises Department, ILO/AIDS, the International Programme for the Elimination of Child Labour (IPEC), the Social Finance Programme (SFP)[7], the Youth Employment Programme (YEP)[7], and the Research Department. Experiences from the SKILLS technical programme were accessed through a previous EVAL assessment of impact evaluation practices.
[8] M.M. Gaarder: *Experimental and quasi-experimental designs* (Prague, 3IE, 2010).

Table 1: Main impact evaluation designs and use in the ILO

Impact evaluation designs	Description	Evidence
Comprehensive longitudinal design with pre-, midterm, post-, and ex-post observations on the intervention and comparison group	This is the most robust design but also the most expensive and time-consuming. The comparison group can be selected to match the intervention group as closely as possible.	X[9]** X[10] X[11] X[12]
'Before and after' and 'with and without' design/ pre-test-post-test intervention and comparison groups	This is often the best available design when provision for impact assessment is made at the start of a programme, a reasonable budget is available; no particular constraints on access to data or use of a comparison group.	X*[13] X*[14]*** X*[15]** X[16]
Truncated longitudinal pre-test and post-test intervention and comparison group design	Intervention and comparison groups observed at two or more points during programme implementation, but benchmarking is done while the intervention is already underway.	X[17]
Pre-test and post-test project group combined with post-test analysis of intervention and comparison group	No baseline data on comparison group	
Post-test intervention and comparison group	No baseline data collected	X[18]
'Before and after'/pre-test-post-test intervention group design	No comparison group	X[19]
Post-test intervention group only design	No baseline and no comparison group	
Theory-based design with a counterfactual	Focus on how & why impact occurred, uses scenario "without" intervention	X[20]

*Evaluation provides robust data predominantly at outcome level and to a lesser extent long-term impact
** Included also qualitative elements such as focus group discussions
***Study not undertaken by the ILO but on ILO intervention

such as focus group discussions, or qualitative surveys but also the use of theories of change complemented by monitoring data. In fact, current evaluation practice seems to indicate that a theory of change is the basis for any impact

[9] N. Fiala: *Stimulating Microenterprises Growth: Results from a Loans, Grants and Training Experiment in Uganda* (Berlin, German Institute for Economic Research, 2013).
[10] ILO and University of Mannheim: *Microfinance and Formalisation of Enterprises in the Informal Sector. An impact evaluation of ESAF (India)'s innovation (Geneva, ILO, internal report 2012).*
[11] A. Landmann, M. Frölich: *Can micro insurance help prevent child labour: An impact evaluation from Pakistan* (Geneva, ILO and University of Mannheim, Research Paper 32, 2013).
[12] Tufts University: Measuring the impact of Better Work (2011).
[13] F. Ahmed, M. Abedin: *Roads 2000 Nyanza Programme in Kenya (*Republic of Kenya, Nairobi, Socio-economic monitoring studies. Final Report, 2010).
[14] Y. Mano, et al: *How Can Micro and Small Enterprises in Sub-Saharan Africa Become More Productive? The Impacts of Experimental Basic Managerial Training* (Washington DC, World Bank, Policy Research Working Paper 5755, 2011).
[15] ILO: *The medium-term impact of Know About Business (KAB) on intermediate institute students in Syria* (Geneva, undated internal report).
[16] ILO: *The effectiveness of labour provisions in reducing the gender wage gap* (Geneva, draft report, 2014).
[17] Reality in impact evaluations seems to indicate that intervention implementation often starts before baseline data collection is completed.
[18] ILO: *Impact assessment: Start and Improve Your Business Sri Lanka* (Geneva, internal document, 2005).
[19] ILO. IPEC: *Kenya Tracer study: Measuring longer term impact on children and families of interventions against child labour* (Geneva, 2012).
[20] ILO. IPEC: *A guide to assessing the impact of time-bound programmes* (Geneva, TBP Map Paper V-02, 2003).

evaluation, irrespective of its methodology. This was still not the case in EVAL's last internal study on impact evaluation in 2009.

Appropriateness of methodologies used[21]

Randomized Control Trials (RCTs): This methodology is rooted in clinical trials conducted in the medical sector and was often described as a "gold standard"[22] for impact evaluations[23]. Its appropriateness outside the medical context and in the more generalized world of work faces substantial challenges.

The use of comparison groups in an experimental impact evaluation design is appreciated as a means to fill the attribution gaps of the causal effects of specific interventions. Comparison groups often constitute the main difference between a performance evaluation and an impact evaluation. They help to "differentiate out" effects of other factors that changed around the time of intervention and isolate the treatment effect.[24] Establishing comparison groups poses challenges for the ILO context, as presented in Table 2, which highlights some of those challenges identified by impact evaluation practitioners and suggests possible solutions discussed with the consultant during interviews.

Table 2: Impact evaluation designs using comparison groups: common challenges and possible solutions

Common challenges	Possible solutions
• Implementation partners such as financial institutions want to sell services even to comparison groups, for example in the area of social finance.	Use of rolling comparison groups, i.e. groups that will benefit from intervention in a later stage of the intervention.
• Ethical considerations hamper the use of comparison groups when addressing HIV/AIDS in the world of work.	Use of rolling comparison groups.
• Comparison groups are difficult to establish, for example, when working with few available large factories, as experienced in garment sectors.	Following evaluability assessment revise the impact evaluation approach away from an experimental design.

The robustness of impact evaluation requires significant financial and human resources as well as time. In the case of Better Work in Vietnam for example, a significant budget was used for the impact evaluation from different sources.[25] Nearly four years were needed from initial design to reporting final results.

[21] Interview comments suggested identifying the rational for using a specific impact evaluation method. While a valid comment, it goes beyond the scope of this study.
[22] V. Coates: *Randomised Controlled Trials: Still going for gold in 2012?* (Ulster, University of Ulster, 2012).
[23] For further reading, see also the 3ie website and the OECD/DAC Evaluation Network *NONIE guidance on impact evaluation*.
[24] Harvard University, Department of Government: *Econometric approaches to causal inference: Difference-in-Difference and Instrumental Variables* (Cambridge, MA, 2005).
[25] ILO. Better work: See baseline and impact measurement tools and methodologies on their website.

Financial resources invested in impact evaluation seem to range from two per cent to 15 per cent of the intervention budget, according to ILO interviewees.

The Youth Employment Programme[26] benefitted from successfully applying for 3ie (International Initiative for Impact Evaluation) funding and other technical programmes of the ILO have also started applying for 3ie funding. It is notable that funding from external sources - other than those provided by the intervention donor - is frequently required to fund the robust analysis, often through RCTs, demanded by impact evaluation.

Tracer studies: ILO's experience with tracer studies, as a more qualitative means to assess impact, is positive and openness about strengths and weaknesses of this impact evaluation approach prevails. As stated by IPEC, the methodology "takes development work to the next level – determining whether development interventions are making real and sustainable impact on the lives of intended beneficiaries."[27] This longitudinal approach is a clear advantage to measure change over time, often beyond the end of the ILO intervention. Some tracer studies use *before and after* comparisons, including baselines and in some cases tracer studies were also applied ex-post. A main disadvantage can be the absence of comparison groups[28] which disables the attribution of change. SME's[29] experience with tracer studies, where comparison groups were used to evaluate impact in entrepreneurship education, point toward challenges in establishing true comparability of the comparison and intervention group, as well as challenges in the traceability of beneficiaries. The latter is a common challenge for most tracer studies.

IPEC's experience with tracer studies in Kenya shows that the methodology depends also on the quality of secondary data. In this respect, the Statistical Information and Monitoring Programme on Child labour (SIMPOC) provides the overall impact information on changes in prevalence of child labour in a country or for a specific sector or geographical area. The programme is not geared, however, to providing data for individual policies or interventions, unless these are covered by the SIMPOC survey being conducted, and for which attribution can therefore be looked into. Some challenges in using SIMPOC data were reported, as they do not contain information on the five impact areas examined by tracer studies.[30] Tracer studies also face challenges in maintaining valid sampling frames given the difficulty of locating beneficiaries well past specific intervention time periods.

Difference-in-difference: In EMP/INVEST, the Social Finance Programme and RESEARCH, econometric modeling for impact evaluations is based on the "difference-in-difference" method. This is an estimation method with the determination of the counterfactual being the core of the evaluation design. RESEARCH for example evaluates the causal relationship between the implementation of a trade agreement with labour provisions and the

[26] Now part of Employment Policy Department
[27] ILO. IPEC: *Kenya Tracer study,* op. cit., p. 9.
[28] ILO. IPEC: *Tracer methodology to assess long-term impacts of interventions on child domestic labour implemented by CHODAWU/ILO – Tanzania: Analytical data report* (Geneva, internal report, 2004).
[29] Now part of Enterprise department.
[30] ILO. IPEC: *Child labour impact assessment toolkit; Tracer study manual*. 3 V. (Geneva, 2011).

improvement of working conditions in Cambodia. The evaluation specifically focuses on the gender wage gap in the garment sector.

The use of the "difference-in-difference" method is appropriate mainly for econometric modeling and environments where large datasets are available at a reasonable quality. Its appropriateness is established if the intervention group is random, which may not be the case when natural experiments, i.e. interventions, are evaluated.[31]

Theory of change: EVAL conducted an internal assessment of baselines, impact monitoring and impact assessment tools and methodologies for six technical programmes in 2009. Results of that assessment revealed a lack of theories of change in many of the impact evaluation approaches. In fact, the development of a logic model or theory of change is important to illustrate explicitly the problems to be addressed by the intervention. The theory of change approach should include a series of assumptions, and include strategies on how to solve the identified problems. Links from the input to the outcome level are also included, followed by links to the short-, medium- and long-term results. The value of using a theory of change in an impact evaluation is to identify how and why impact can be expected to occur.

ILO/IPEC[32] provides detailed guidance on the use of the theory of change in impact evaluations.

Results of ILO impact evaluations

At least seven examples of ILO impact through impact evaluation emerged for 2012, 2013 and 2014. All examples relate to impact at the intervention level but the one presented for 2014 included policy level changes.

2012

1. The Microfinance for Decent Work (MF4DW) action research evaluated the impact from the voluntary option to extend health and accidental death insurance to household members outside the nuclear family and issues of child labour in Pakistan. Specific intervention impacts include:
    - The innovation did in fact lead to increased insurance coverage at the individual level;
    - Child labour decreased by almost seven per cent, with the largest effect on boys, as they were more active as child labourers;
    - Around five per cent to six per cent lower risk of hazardous occupations caused by the innovation for boys and girls;
    - Many of these effects were driven by a shift from heavy wage work or day labour in shops, factories or the street towards less intensive work in the house and the family;
    - The innovation did not affect school attendance, as economic reasons were less often stated as a reason to stop schooling.

[31] Harvard University, Department of Government, op. cit.
[32] ILO. IPEC: *A guide to assessing the impact of time-bound programmes* (Geneva, TBP Map Paper V-02, 2003).

2. In India, the ILO worked on awareness raising campaigns and business development services for the formalization and strengthening of growth-oriented enterprises. In this microfinance project, also evaluated under Microfinance for Decent Work (MF4DW) action research, the ILO's partner offered loans, insurance, training and consulting services to beneficiaries. The 2012 impact evaluation showed a 70 per cent increase in business registrations with relevant authorities. As a result of these innovations the share of clients who stated that their household enterprise was the main income source, increased by 17 per cent. However, the purchase of new durable goods for the household developed at a similar rate in the intervention and comparison group.[33]

3. In Kenya the ILO measured the longer-term impact on children and families of interventions against child labour in 2012 using a tracer study.[34] The approach used plausible attribution in the absence of a comparison group. Interventions provided education as the alternative to child labour, which constitutes the most critical pillar in the campaign against child labour. Results from the analysis revealed that the number of beneficiaries that were engaged in child labour reduced during the duration of the project but gradually began to rise when the project ended. The proportion of beneficiaries who attended formal education did not change very much before, during or after the project. There was however, a sharp increase in enrolment and attendance to vocational and non-formal training institutions during the project period but it declined once again after the intervention period. The tracer study recorded improvement in the health status of beneficiaries and their communities. The latter seemed to be mainly related to increased government funding. Attribution to the ILO interventions was not possible in the absence of a comparison group.

2013

4. The Start and Improve Your Business Programme (SIYB) in Uganda detected changes on programme beneficiaries in education, health and household consumption in 2013.[35] Beneficiaries were drawn from a sample of 1,550 microenterprise owners and the impact was attributed to a combination of SIYB training modules combined with access to loans and grants. The impact evaluation, using an RCT, showed an increase in income and household labour for men while the programmes had little or negative effects on women's income. The latter was due to family pressure to spend money outside the business, which is normally much higher for women than for men. Measurements taken six months after the interventions were delivered, showed an increase in missed school for household children, consistent with men demanding family support. This was accompanied by a significant decrease in spending on child health for women who were part of the *loan only* program due to pressures to repay the loan. The effects on missed school had disappeared by the nine-month

[33] ILO and University of Mannheim, op cit.
[34] ILO. IPEC: *Kenya Tracer study,* op. cit.
[35] N. Fiala, op. cit.

follow-up. Children were no longer being pulled out of school to work in the business. Despite increased profits for men in the loan and training program, household consumption did not increase.

5. The Better Work Programme in Jordan,[36] working with 58 garment factories and covering over 39,000 workers (67 per cent female), contributed to a 19 per cent increase in compliance in forced labour issues. In 2012, 94 per cent of workers in Jordan reported having comparison of their own passports, a 19-point increase from 2010 and especially important for the two thirds of Jordanian garment workers who are international migrants. From 2010 to 2012, workers report significant health improvements. Reports of frequent exhaustion or fatigue were down nearly 50 per cent.

6. In Vietnam, the Better Work Programme operated in 198 garment factories with over 213 workers (76 per cent female).[37] Between 2010 and 2013 the programme contributed to a 10 per cent increase in worker income and a 3 per cent improvement in workers' health. The capacity utilization rate of Better Work Vietnam factories increased by 15 per cent while the proportion of factories that were preferred suppliers, or most important customer for their buyers, increased by over 20 per cent.

2014

7. RESEARCH is currently undertaking an impact evaluation related to the 1998 *ILO Declaration on Fundamental Principles and Rights at Work* in the textile sector in Cambodia.[38] The impact evaluation uses annual household survey data for Cambodia as well as data from Better Work's interventions in the country to measure the impact of the 1999 *Bilateral Textile Agreement* between Cambodia and the United States of America. The agreement is based on US core labour standards, following the 1998 *ILO Declaration on Fundamental Principles and Rights at Work*. For the impact evaluation a difference-in-difference approach is used with the non-textile sector as comparison group. Preliminary results show there is a significant reduction of the gender wage gap in the textile sector, which can be attributed to the *Bilateral Textile Agreement* between Cambodia and the United States of America, aligned to the 1998 ILO Declaration.

Common constraints

This study found that eight out of nine technical programmes of the ILO state that the level of effort invested in impact evaluations was worthwhile, given the robustness of results. However, a number of common constraints emerged during the planning and implementation phases of impact evaluation, as found in the present study and shown in Table 3 below.

The two main challenges identified were resource limitations and identification of an appropriate methodology which are related to the planning of impact

[36] ILO and International Finance Corporation: *Better Work Jordan: Impact Brief* (Geneva, 2013).
[37] ILO and International Finance Corporation: *Better Work Vietnam: Impact Brief* (Geneva, 2013).
[38] ILO: *The effectiveness of labour provisions in reducing the gender wage gap* (Geneva, draft internal report, 2014).

evaluation. At the implementation stage, the constraints become increasingly manifold, combined with timing constraints, lack of or access to baseline and monitoring data, as well as the lack of suitable comparison groups. This is further compounded by a general lack of knowledge and expertise in impact evaluation within the ILO to appropriately oversee and guide researchers. There are, consequently, considerable management issues which need to be addressed if ILO intends to expand its work in impact evaluation.

Table 3: Impact evaluation challenges

Planning	ILO organizational units, n=9
Adequate resources allocated in project budget	7
Evaluation too early to assess real impacts on the ground	5
Proposing adequate impact evaluation methodology	5
Implementation	
Time constraints	8
Access to/lack of baseline data	8
Lack of comparison groups	7
Inadequate financial resources	6
Lack of monitoring data	6
Lack of conceptual clarity/guidance on methodology	6
Lack of sufficient impact evaluation knowledge within the programme	6

Risks and effects on the ILO and its constituents

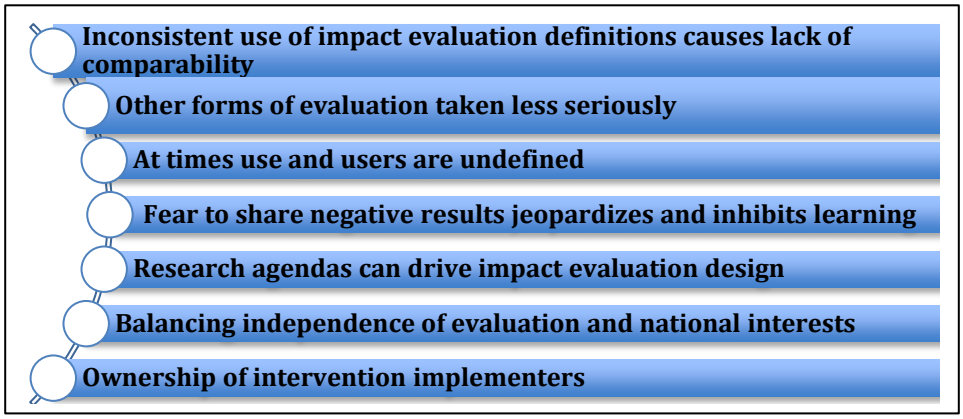
Impact evaluation results are used to feed in to ILO policy dialogue, changes in legislation or regulations, enhance the design and replication of successful approaches, to underpin the relevance of the organization, more generally and to attract donor funding.

> "Impact evaluations are really good. They serve as an active research and development approach. It helps us to test and modify, test and modify again. This is particularly true in my work for example to enhance our gender focus and better reflect the needs of women. "
>
> Source: ILO staff working on impact evaluation

Figure 2 summarises the main risks identified during this stock-taking exercise. Most ILO staff interviewed expressed uneasiness with the variety of definitions used for impact evaluation, due in part to inconsistent donor practices, but also due to a disregard for guidelines established by EVAL.[39] The current result is that ILO has produced a range of impact evaluations that are not necessarily comparable in purpose (credible assessment of attribution and causal effects for long-term results related to economic well-being).

Figure 2: Risks related to impact evaluation identified in the ILO

[39] ILO. Evaluation Office, op. cit.



While the evaluation literature and existing guidance, including ILO evaluation guidelines reflect different impact evaluation methodologies with different degrees of rigor, challenges seem to occur when performance in delivering outputs or outcomes are evaluated and being "sold" as impact evaluation.

> "Somehow I feel that there is a conceptual contradiction between aiming to prove long-term impact as early as at the immediate end of the intervention cycle."
>
> Source: ILO staff working on impact evaluation

One side effect of impact evaluations, intentional or unintentional, is that ILO constituents and the ILO itself can take other forms of evaluations less seriously, i.e. that impact evaluation results gain more weight than results of other evaluation types without properly understanding that each evaluation type fits a distinct purpose. Linking impact evaluation to the target users and to understand how the evaluation results will be used to inform decisions is another, often unresolved issue. In addition, a tendency to selectively use impact evaluation results and to highlight mainly positive aspects can hamper learning from intervention implementation challenges. In fact, there is fear that negative impact evaluation results could jeopardize the work of technical programmes that pioneer impact evaluation.

> "Are impact evaluations encouraged in the ILO? We need a clear signal of senior management buy-in. Otherwise we could become victim of our own success."
>
> Source: ILO staff working on impact evaluation

An **unexpected risk** of impact evaluation is the close alignment of the ILO with the research agenda of specialized academic institutions. The more value primary data collection shows for the academic partner, the higher the likelihood that fees get reduced and the ILO achieves a "better deal" and saves significant amounts of money. While this seems at first glance a "win-win" situation, in fact research agendas of academia can drive impact evaluation designs which need to be closely managed. High-level technical expertise in impact evaluation is

required to dialogue about the detailed evaluation design, if ILO wishes to maintain control.

Another risk experienced is the fine balance to be struck between the independence of impact evaluations and national interests and sensitivities with regard to assessing impact of their policies and programmes.

Currently, the intellectual property of impact evaluation data is handled unevenly across the ILO. As no common practice prevails - despite ILO rules to systematically address this issue in procurement and External Collaboration contracts - instances emerge where the ILO faces challenges to access primary data of its own interventions when it is held by impact evaluation partners.

Finally, project staff can perceive impact evaluation as a rigorous accountability tool providing sensitive information to donors, causing at times project implementers' resistance to fully own impact evaluation processes and results.

Key lessons learned

Time and resource requirements

- Impact evaluation is a very technical, time and resource intensive form of evaluation. Demands on management time in the ILO are high to oversee impact evaluation researchers and this can increase indirect costs of an impact evaluation by as much as 50 per cent.
- Access to a network of high level researchers can help to mobilize impact evaluation funding.

Impact refers to long-term results

- Expectations of impact evaluations need to be managed. Only if elements of ex-post assessments some years after the end of the intervention are foreseen, it might be possible to evaluate developmental impact at the beneficiaries' livelihood level.
- Policy level impact requires long-term tracking. It might take multiple years for the impact of policy work within a particular country or context to be measured adequately at the beneficiary level.

Value of a clear purpose and realism

- The purpose of an impact evaluation should be clarified from the outset. The identification of scalable models, accountability to constituents and the assessment of the ILO's policy influence seem worth the investment.
- It seems important to identify ex ante who and how the impact evaluation results will be used. There should be a clear need or plan for using the results to affect decisions to establish the utility of the impact evaluation.
- Realism in impact evaluation is important and the approach needs to be adapted to realities. Using experimental or quasi-experimental designs with randomized control groups are not always feasible, for example, where only a few players comparable to the intervention group are identifiable.

What would ILO impact evaluation practitioners replicate?

Impact evaluation design starts at intervention design. A reasonable and useful point in time to design robust impact evaluation is at the very beginning of an intervention planning. This should ensure robustness and planning for the collection of baseline and end line data, including for comparison groups, and a clearly agreed intervention logic or theory of change.

Establish a functioning monitoring system. This should include databases which are a cornerstone for impact evaluation in EMPLOYMENT and IPEC, and directly affect the quality, availability and robustness of good quality data.

Engage staff and partners actively. Data collection through specialized research partners contributes to capacity building, enhanced ownership and ultimately sustainability, as proven in the social finance programme in Pakistan and the Philippines. Research consultants and implementation staff should be involved in the impact evaluation design, the latter to fully appreciate and support the approach taken.

EVAL's role in impact evaluations

ILO staff working on impact evaluations and interviewed for this think piece predominantly want EVAL to play a more prominent role with regard to impact evaluations. The degree of recommended EVAL involvement could be as follows:

Advocacy	○ Advocate and educate more actively to inform ILO staff about impact evaluation, for example through the use of the evaluation focal point network; giving particular importance to field staff ○ More proactive information should be made available through EVAL about issues such as impact evaluation purpose, users of evaluation results, timelines, budgets and data requirements
Knowledge sharing	○ Continue with the peer-review seminars to 1) present impact evaluation methodologies; 2) share experience; and 3) strengthen ILO's impact evaluation community of practice ○ Include field staff through video conferencing when knowledge sharing events take place in HQ
Technical support	○ Develop a more active sounding board for impact evaluations in the planning and analysis stage ○ Install a roster of approved impact evaluation researchers
Quality assurance	○ Require impact evaluation strategies from organizational units

EVAL already addresses a couple of important gaps identified by ILO staff. Impact evaluation is clearly defined in *Guidance Note 13: Impact Evaluation* (2013) which is part of the *i-eval Resource Kit - ILO policy guidelines for results-based evaluation: Principles, rationale, planning and managing for evaluations (2012).* However, a lack of awareness about this guidance prevails among most impact evaluation practitioners interviewed for the stock-taking exercise.

A list to assess the quality of impact evaluation also exists. EVAL's *Checklist 9: Impact evaluation planning* (2014) serves also as a very useful evaluability checklist. Practitioners will swiftly find out whether it is possible to include impact evaluation into the design of a new intervention.

The section on impact evaluation in ILO policy guidelines was purposefully expanded and detailed in the guidance note and checklist which are consistent across these three documents, and in line with both the OECD/DAC Evaluation Network *NONIE guidance on impact evaluation* and the guidance issued by the United Nations Evaluation Group (UNEG). [40]

Stock-taking exercise conclusions

1. Purpose of impact evaluations in the ILO
While EVAL guidance identifies three distinct purposes to undertaken an impact evaluation, interview results suggest that the rational for impact evaluation is even more diverse. Staff point to additional purposes such as positioning the ILO and showing its comparative advantage and relevance to donors. The identification of a clear purpose of any impact evaluation can help to assess ex-ante its usefulness and to inform methodological choices.

2. Methodologies used
In the nine technical programmes surveyed across the ILO in July 2014, a rich variety of impact evaluation methodologies were noted. Methodologies included the randomized control trail, followed by tracer studies, the use of difference-in-difference techniques in econometric studies and regression discontinuity. Increasingly qualitative elements such as a theory of change approach, for example in time-bound programmes, are used for impact evaluation to complement quantitative techniques for better understanding of how and why change takes place.

3. Appropriateness of methodologies
Methodologies were reasonably successfully applied. However, the review shows that out of eight impact evaluations, using quantitative technics or techniques close to that standard, three had a focus at the interventions outcome level rather than impact at the level of economic well-being of individuals, households, communities or organizations. Feedback from EVAL and ILO staff indicates that ambitious impact evaluation methodologies need to be adapted to reality where baseline data might be limited or access to comparison groups is not always feasible.

4. Driver of impact and final beneficiaries
Six out of seven evaluations from 2012 to 2014 identified as impact evaluations with a focus on long-term results, assessed impact at the beneficiary level (for example, access to training and finance for micro entrepreneurs, informal entrepreneurs and education of child labourers). Impact evaluation at the policy

[40] United Nations Evaluation Group: *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*, (New York, 2013).

level and relevant tracking seems still underdeveloped with few exceptions.[41] In fact, it might take multiple years to see the impact of the ILO's policy work at the beneficiary level for specific Programme and Budget outcomes in a given biennium, as shown in the examples provided in this review. Stronger institutional commitment might be required to better address this issue.

5. Solidity of findings and value for money of investing in impact evaluation

Eight out of nine technical programmes of the ILO stated that given the robustness of results, the level of effort invested in impact evaluations was worthwhile. In fact, the impact evaluation designs reviewed seem robust and would normally include a comparison group to address the attribution issue of long-term results. However, the use of complex, highly elaborate and expensive evaluation approaches for the assessment of medium-term results and outcomes is problematic, as observed in three out of six evaluations reaching or at reach of the "gold standard". Donor pressures to show impact within the lifetime of an intervention as well as inconsistent donor definitions and understanding of what impact evaluations actually entail are just some explanations for this costly discrepancy. In fact, outcome level evaluations using impact evaluation methodology might satisfy donor's demand for information on changes in beneficiaries' knowledge, awareness and practices following ILO capacity building, though this falls short of assessing the specific developmental impact.

6. Common constraints

Despite the methodological strengths of many impact evaluations in the ILO, the complexity of this evaluation type is significant. During the planning phase of impact evaluation, ILO staff must face challenges related to financial resources, timing and methodology. When implementing impact evaluations, challenges increase. Additional constraints emerge beyond the resources and methodological challenges faced at the planning stage: to undertake the evaluation on time, a lack of or access to data, including baseline and monitoring data, as well as lack of suitable comparison groups. A lack of capacity in the ILO to appropriately oversee and guide researchers is another limiting factor. The latter could be another reason for some impact evaluations stopping with the assessment of mid-term results rather than focusing on long-term results, and the broader developmental impact.

7. Use of evaluation results

In the ILO the use of impact evaluation results is manifold. It includes proving the scalability of models, establishing the relevance of the ILO at the policy level, integrating the World of Work thinking in sectors such as microfinance and providing examples of real changes for beneficiaries on the ground. Ultimately, impact data gives the ILO a comparative advantage over other organizations by generating evidence about what works and why.

[41] One exception identified in this study is the impact evaluation related to the 1998 *ILO Declaration on Fundamental Principles and Rights at Work* in the textile sector in Cambodia. The ILO finds a significant reduction of the gender wage gap in the textile sector which can be attributed to the *Bilateral Textile Agreement* between Cambodia and the United States of America, aligned to the 1998 *ILO Declaration*.

8. Risks and effects on constituents

- *Inconsistent use of terminology:* Despite a common definition of impact evaluation in the ILO and guidance notes and instruments of EVAL on the topic based on international standards (OECD/DAC and UNEG), ILO staff use the terminology "impact evaluation" inconsistently. The survey shows that this is in fact due to a lack of awareness about EVAL guidance on the topic introduced in 2012. Inconsistent definitions and expectations of impact evaluation amongst donors further increase the complexity of this issue.

- *Credibility of impact evaluations versus other evaluations:* Given the methodological rigor and credibility of expensive impact evaluations, ILO constituents and the ILO tend to take traditional performance evaluations less serious, which is a negative side effect of impact evaluations. Serious evaluability analysis is required to justify the investment in impact evaluations.

- *Dominating research agendas:* The fact that research agendas of academics can drive impact evaluation design needs to be closely managed. Otherwise the ILO might be in danger of losing control over methodological choices. High-level technical expertise in impact evaluation is required in the ILO to dialogue about the detailed evaluation design.

- *Fear to show challenges in achieving impact:* ILO staff also expressed their fears that negative impact evaluation results might not be well received by their line managers with the result that mainly positive aspects of impact evaluations are highlighted and therefore hamper learning from challenges.

- *Demand for EVAL support:* Finally, many staff interviewed looked for more EVAL support with impact evaluations, covering issues such as advocacy, knowledge sharing, technical support and quality assurance.

Recommendations based on the conclusions

For each technical programme, with backstopping from EVAL

1. Prior to any impact evaluation, technical programmes should clarify the purpose of the impact evaluation with a clear focus on who would use the evaluation results and for which specific decision-making processes. This could inform methodological choices to be taken.

Linked to conclusion 1

2. An evaluability assessment prior to each impact evaluation should guide the technical programmes of the ILO whether it is worth investing in an impact evaluation. For this purpose, EVAL's *Checklist 9: Impact evaluation planning* should be consistently used and annexed to the final evaluation report. The ILO should stay clear from the significant investment in an impact evaluation if reality on the ground tells that only medium-term results can be assessed during the proposed timing of the impact evaluation, as opposed to long-term results linked to MDGs and subsequent development goals.

Linked to conclusions 1, 2, 3, 5 & 6

For each technical programme

3. The ILO should start investing more in tracking the impact of advice on normative work and evaluating changes at the policy level following the support of constituents within specific national contexts.

Linked to conclusion 4

For EVAL

4. The terminology "impact" and "impact evaluation" should be applied consistently across the ILO, based on EVAL *Guidance Note 13: Impact Evaluation* to ensure that long-term results, the attribution and causal effects of development interventions at the beneficiary or policy level are the purpose of any impact evaluation. In this respect, EVAL could play an even more active role with regards to advocacy, knowledge sharing, technical support and quality assurance for impact evaluations.

Linked to conclusions 7 & 8