



International
Labour
Organization

Evaluation
Office

► i-eval THINK Piece, No. 23

Assessing the effectiveness
and results of ILO's work:
key proxies of performance
A methodological review

By Magali Bonne-Moreau
March 2023



Introduction

For over a decade, the ILO's Evaluation Office (EVAL) has aimed to provide evaluative evidence to promote accountability, learning and decision-making, through annual assessments of independent evaluation reports of development cooperation projects, used as proxies to measure ILO's performance results.

►► EVAL excels at producing regular, timely and data-rich reporting on evaluation performance, evaluations undertaken, findings and recommendations, which is highly valued both internally and externally

► **Independent Evaluation of the ILO's Evaluation Function for the Period 2017- 2021 (2022)**

While there is recognition that evaluations provide credible and useful feedback to the ILO on its effectiveness in advancing decent work, this Think Piece aims to reflect on the extent to which the current proxies of performance used (independent project evaluations) in effectiveness reviews effectively provide a picture of ILO's overall performance; and the extent to which meta-analysis reports present results in a format that informs decision making in the ILO review; and therefore continues to be valid.

Following this short introduction, Section 2 provides an overview of the results from the ILO effectiveness review meta-analyses, as well as the framework and criteria used to assess performance. Section 3 presents a non-exhaustive summary of what is being done in other UN agencies to present and report on key proxies of performance based on evaluative evidence, looking at the different sources of evidence, approaches used, and key intended users. Section 4 discusses what can be learned from these findings, and highlights areas of performance that could be better addressed in the current meta-analysis framework. Based on this discussion, section 5 reflects on whether the current format for reporting and disseminating findings on performance is suitable, and if any changes need to be made to improve learning and decision-making.



How does the ILO Evaluation Office use evaluative evidence as key proxies of performance?

The ILO's Evaluation Office designed the first meta-analysis of independent project evaluations in 2011, to provide feedback on the effectiveness and operational performance of technical cooperation, based on evaluations completed in 2009 and 2010.¹ This was in line with the ILO Strategy for Evaluation 2011-15.

Methodological adjustments were made over time² to improve the application of the review framework and to address emerging concerns and cover key areas of interest, yet comparable approaches and methodologies have been used over time, as shown in Table 1 below.

► **Table 1. Overview of ILO Decent Work Effectiveness meta-analyses**

Coverage of meta-analysis	No of performance criteria	Type of scale	Use of rubrics	No of reports in sample
2009-2010	38	4 point	No	59
2011-2012	28	6 point	No	40
2013-2016	26	4 point	Yes	40
RBSA (2013-2017)	26	4 point	Yes	21
2017-2018	26	4 point	Yes	40
2019-2020	28	4 point	Yes	40
2020-21	29	4 point	Yes	48
2021-22	29	4 point	Yes	49

¹ ILO Evaluation Unit (2011) Decent work results and effectiveness of ILO technical cooperation: a meta-analysis of project evaluations, 2009-2010.

² A methodological review including a detailed discussion of changes made to the methodology between 2009 and 2016 was conducted by Universalia in 2017, and a second review was conducted in 2020 by the author of this Think Piece.

The main initial changes were a reduction in the number of criteria covered, descriptions for ratings, as well as changes in scale. The first meta-analysis used a 4-point scale to rate effectiveness. The second meta-analysis used a 6-point scale, while subsequent meta-analyses returned to a 4-point scale. From the 2013-16 meta-analysis onwards, a scoring rubric was developed by EVAL defining a description associated with ratings for each criterion. Particular attention was paid to the comparability of results, to facilitate analysis over time. Furthermore, from the RBSA meta-analysis onwards, the evaluative evidence for the assessments goes beyond lessons learned, conclusions and recommendations, to include all evaluation findings in the report.

Overview of existing methodological approach

The current methodology focuses on criteria grouped into three main performance areas: (1) Strategic relevance and alignment; (2) Effectiveness, sustainability and impact; and (3) Implementation performance and efficiency of management and resource use.

In 2021 EVAL conducted an internal methodological review, based on recommendations from an independent methodological review performed in 2020. Two main changes were made:

1. Re-wording of some of the rubrics for increased clarity of criteria definitions, while being mindful of the importance of comparability over time; and
2. The introduction of three new criteria in the appraisal framework, that had been piloted in earlier years, as discussed further below.



Strategic relevance and alignment

- | | |
|---|---------------------------------|
| 1.1 Link between project purpose and/or objectives with P&B outcome | 1.3 Constituent support |
| 1.2 Causal link between project objectives and DWCP outcome(s) | 1.4 Validity of design/approach |
| | 1.5 Pro-poor focus |
| | 1.6 Gender-responsive |

1



2. Effectiveness, sustainability and impact

- | | |
|--|---|
| 2.1 Quality and completeness of outputs | 2.7 Strategic importance of results achieved |
| 2.2 Achievement of immediate objectives | 2.8 Strategic relationships |
| 2.3 Knowledge development | 2.9 Tripartite processes and social dialogue |
| 2.4 Capacity building | 2.10 Sustainability of policies, knowledge & capacities |
| 2.5 Normative work / standards promotion | 2.11 Acknowledgement and Use of ILO expertise and involvement |
| 2.6 Policy influence | 2.12 Resource leveraging |

2



3. Implementation performance and efficiency of management and resource use

- | | |
|--|---|
| 3.1 Results-Based Management approach and goal orientation | 3.5 Monitoring and reporting, |
| 3.2 Implementation management | 3.6 Visibility and accessibility to knowledge and information |
| 3.3 ILO support to project | 3.7 Cost efficiency |
| 3.4 Internal ILO coordination | 3.8 Adequacy of resources |

3



4. New criteria: Current development cooperation concerns

- | |
|--|
| 4.1 Consistency and contribution of the intervention to SDGs |
| 4.2 Disability inclusion |
| 4.3 Integration of just transition to environmental sustainability |

4

The **Strategic relevance and alignment** performance area relates to the OECD/DAC definition of Relevance which aims to answer the question **“is the intervention doing the right thing?”** and is defined as *“The extent to which the intervention objectives and design respond to beneficiaries’, global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.”*

It is composed of six criteria. The first two criteria focus on the link between project objectives and strategic priorities of the ILO (Criterion 1.1) as well as those at the country level, as identified in Decent Work Country Programmes (DWCP) and Country Programme Outcomes (CPOs) (Criterion 1.2). The third criterion examines the extent to which constituents were involved in project formulation and implementation, and if the project reflected their decent work needs (Criterion 1.3). The fourth criterion (Criterion 1.4) addresses design aspects in terms of the alignment of strategies and services to project objectives, while the last two criteria (1.5 and 1.6) assess how poverty and the ILO cross-cutting policy driver of gender are addressed in interventions.

The **Effectiveness, sustainability and impact** performance area relates to the OECD/DAC definitions of:

3. **Effectiveness**, which aims to answer the question: **“Is the intervention achieving its objectives”** and is defined as *“The extent to which the intervention achieved, or is expected to achieve, its objectives and its results, including any differential results across groups.”* This is addressed by assessing the quality and completeness of outputs, and the achievement of immediate objectives (Criteria 2.1 and 2.2).
4. **Sustainability**, which aims to answer the question: **“Will the benefits last?”** and is defined as *“The extent to which the net benefits of the intervention continue or are likely to continue.”* This is partially covered by looking at Sustainability of policies, knowledge & capacities (Criterion 2.10); and Resource leveraging (Criterion 2.12). This is also related to Strategic relationships (Criterion 2.8), and possibly Acknowledgement and use of ILO expertise (Criterion 2.11). Scores related to Knowledge development (Criterion 2.3) and Capacity building (Criterion 2.4) also play a role in overall sustainability.
5. **Impact**, which aims to answer the question: **“What difference does the intervention make?”** and is defined as *“The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.”* This could be related to Policy influence and the Strategic importance of results achieved (Criteria 2.6 and 2.7).
6. This performance area also includes assessments of specific means used to achieve objectives/results, such as **Knowledge development** (Criterion 2.3) and **Capacity building** (Criterion 2.4). There is often an overlap in the evidence used to assess these criterion and Criterion 2.10. In addition, several criteria focus on ILO cross-cutting priorities, which are both a means to achieve results, as well as an end, such as **Normative work / standards promotion** (Criterion 2.5); and the extent to which **Tripartite processes and social dialogue** are embedded in the project approach (Criterion 2.9).

The third performance area, **Implementation performance and efficiency of management and resource use**, is associated with the OECD/DAC definition of Efficiency, which aims to answer the question “**How well are resources being used?**” and is defined as “*The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.*” This dimension is associated with Cost efficiency (Criterion 3.7), and should involve converting “inputs into outputs, outcomes and impacts in the most cost-effective way possible, as compared to feasible alternatives in the context”, thus reflecting the need for a thorough contextual analysis. The timeliness of delivery is another important element, which is partially captured by looking at Adequacy of Resources (Criterion 3.8). Operational efficiency (how well the intervention was managed) is also appraised by examining Implementation management (Criterion 3.2), ILO support to project (Criterion 3.3), Internal ILO coordination (Criterion 3.4) and Monitoring and Reporting (Criterion 3.5).

Goal orientation (Criterion 3.1) covers elements which are closely related to design aspects and are associated with a Results-based management approach.

Visibility and accessibility to knowledge and information (Criterion 3.6) focuses on the documentation and dissemination of knowledge outputs produced by interventions, both internally and externally.

Introduction of new criteria covering specific development cooperation concerns

To better represent the work done by the ILO, and in alignment with current development cooperation concerns such as the 2030 Agenda, United Nations Cooperation Frameworks, and the UN system-wide Disability Inclusion Strategy (UNDIS), two new criteria were defined by the evaluation team, with validation from EVAL, and piloted in the meta-analysis of reports from 2019 and 2020. Both are associated with the “Strategic relevance and alignment” performance area.

The first criterion, **Consistency and contribution of the intervention to SDGs**, aims to capture ILO's contribution to the SDGs, by focusing on the alignment of project objectives and SDG targets, and evidence of contributions of interventions to these targets, as well as coordination with other agencies in the UN system. This criterion also aims to contribute to measure the external coherence of ILO's interventions.

The second criterion, **Disability inclusion**, examines the extent of disability inclusion in intervention design and implementation, and the ratings definitions are based on the twin-track approach described in the UN Disability Inclusion Strategy (UNDIS).

A third criterion, piloted in the 2020-21 Effectiveness Review, aims to assess the extent to which the **integration of a just transition to environmental sustainability** is included in ILO's work. This cross-cutting policy driver was introduced in the ILO Strategic Plan for 2018-21. In this way, all the ILO's cross cutting policy drivers are covered in the meta-analysis framework.

The four-point rating system

Based on pre-defined rubrics and information provided in evaluation reports, assessors use a four-point rating system to score each criterion, as presented below.

Unsuccessful (1)	Partly successful (2)	Successful (3)	Highly Successful (4)
Highly deficient performance	Basic level of performance but stated need for improvement	Adequate level of performance and results	Very good level of performance and results

To ensure methodological robustness and inter-rater reliability, each evaluation report is rated independently by two reviewers, after a calibration exercise to ensure shared understanding of the concepts defined in the rubrics. The ratings are compared, and any discrepancies are discussed to produce a single rating.

Aggregation of results

Since 2020, the ILO has used a rolling approach to evaluation assessments, to allow for an account of performance results completed within the reporting year. Individual evaluation assessments are conducted, and all ratings for each criterion are recorded in a global performance matrix, along with meta-data and associated qualitative evidence. This allows for aggregation of results in response to emerging needs.

There have been several instances where performance results have been aggregated to take a comparative approach and examine trends over time, across geographic regions, departments, and P&B outcomes, based on median ratings for each criterion. The most recent comparative assessment was performed in the rolling meta-analysis for the period 2021 – 22.

Most recently, the 2022 report examined performance trends through median results of projects evaluated over the past five years, disaggregating them by technical intervention area and region. While the sample size was robust, results needed to be interpreted with caution, due to the uneven distribution of projects across the different sub-categories. Nevertheless, this approach highlighted specific trends related to systemic issues that would be worth observing over time, reflecting the added value of such an exercise.

Furthermore, clustering of criteria has been used in past Annual Evaluation Reports to give a rapid overview of performance ratings over a given timeframe, as presented below.

► **Figure 1. Clustering of criteria for Annual Evaluation Report**

Strategic relevance and alignment criteria 1.1 to 1.6	Effectiveness and impact criteria 2.1 to 2.9	Efficiency criteria 3.2 to 3.4, 3.8
M&E Framework criteria 3.1 and 3.5	Financial/cost efficiency criteria 3.7 and 3.8	Sustainability criteria 2.10

How do other UN agencies present and report on key proxies of performance based on evaluative evidence?

A non-exhaustive review of performance assessments by evaluation offices in different UN agencies was conducted to explore the evaluation sources and criteria used by other agencies to report on overall performance results, based on evaluative evidence, and the objectives of these assessments. The following section highlights key findings from this review.

Only some of the UN evaluation offices reviewed had a regular practice of using and reporting on key proxies of performance based on evaluative evidence. When reporting took place, findings were presented within specific sections of Annual Evaluation Reports or in stand-alone documents, sometimes accompanied by summary sheets or presentations highlighting key findings.

Intended audiences and objectives for performance-related analyses

Performance-related analyses based on evaluation contents were often conducted to report on results to Governing/Executive bodies or General Assemblies, as well as senior management, partners, and stakeholders.

The objectives for these analyses were related to learning, accountability and decision-making, with different emphases on these three elements, depending on the UN Agency. Some of these objectives are quoted below.

► **Objectives of different UN Agencies for carrying out performance analyses based on evaluative evidence**

"The annual synthetic review contributes, at an organization-wide level, to the three purposes of evaluation defined in the UNESCO Evaluation Policy 2022-2029 — **learning, accountability, and decision-making** – by presenting and analyzing evidence of UNESCO's performance and progress on goals and directions set in its strategic guiding documents." (UNESCO)

"The objective of this year's Annual Evaluation Synthesis Report (AESR) [...] is to facilitate **learning** by synthesizing the key messages emerging from recent evaluation reports and related reviews. [...] The 2022 AESR also includes a new perspective. As part of the management response to the 2021 AESR, the IEU has been tasked with analyzing the lessons learned from the implementation of ITC's Strategic Plan 2022-25. The 2022 AESR will contribute to this task by assessing the findings from evaluation reports and related reviews against the main dimensions of the Strategic Plan, thereby establishing a form of **baseline against which to track progress** in the implementation of the Strategic Plan in the coming years." (ITC)

"The report provides stakeholders such as Governments, UNEP senior management and partners with **an evaluative assessment of UNEP's programme and project performance** in the 2020-21 biennium." (UNEP)

"The introduction of a programme performance rating system (hereafter 'rating system') for use in UNDP Independent Country Programme Evaluations (ICPEs) to determine the contribution of UNDP country programmes is another effort to consolidate the IEO resources and data architecture to **support organizational learning**." (UNDP)

"It offers a unique insight into the aggregate results of UNODC evaluations from 2019-2021, ensuring that lessons can be learnt at a strategic and programmatic level **to foster learning and accountability**". (UNDOC)

"The main purpose of this Country Programme Evaluation (CPE) synthesis is to **contribute to organizational learning** above and beyond the country- and WCO-specific learning generated in the individual CPEs." (WHO)

"The objective of the synthesis is to identify lessons from evaluations that can contribute to an informed discussion on FAO's strategic positioning and contribution to results in the region and **provide reflections and suggestions for future work planning** at the regional level. The preparation of the regional synthesis **is part of a broader objective of enhancing the usefulness and utilization of evaluations** by members of the governing bodies, in particular, at the regional levels." (FAO)

"In providing an overview of the performance of IFAD-supported operations based on independent evaluations, the ARIE continues to be pivotal in **ensuring accountability** for results." (IFAD)

"The purpose of this report is to aggregate and synthesize information generated from 2017 evaluations and transform it into **accessible knowledge for future systematic strengthening of programming, organizational effectiveness**, and the evaluation function. This is intended to help inform the implementation of the United Nations Women Entity for Gender Equality and Women's Empowerment (UN Women) Strategic Plan 2018-2021." (UN Women, 2018)

"This meta-synthesis focuses on types of UN Women support that have not been covered in recent syntheses and evaluations, and that are highlighted in the new Strategic Plan. The meta-synthesis **contributes to UN Women's organizational knowledge base** and the primary audience for the synthesis is UN Women stakeholders, including the Executive Board and UN Women personnel." (UN Women, 2022)

Types of evaluative evidence for reporting on performance

There were different types of evaluations used to synthesize findings on performance, with agencies drawing evidence from self-evaluations, project completion reports, joint, external and donor-led evaluations, independent project and programme evaluations, as well as higher-level strategic/corporate reviews and evaluations. In some cases, these were supplemented by interviews or statistical data.

One recurring key limitation was that it was sometimes difficult to draw strong conclusions due to the disparate evidence provided in reports, as well as significant variations in scope and approaches of evaluations and the interventions they covered.

Criteria used to report on performance

There is a variety of criteria used to report on performance and effectiveness. Several outputs, such as the ones presented for UNEP, IFAD, and UNESCO (see box below), were structured around the OECD/DAC evaluation criteria of relevance, coherence, effectiveness, efficiency, impact and sustainability, with indicators adapted to the specific mandate and objectives of the organization. In some cases, agencies also linked these to their own strategic objectives and/or cross-cutting issues.

In contrast, reports from other agencies, such as the [2021 FAO Programme Evaluation Report](#), used evidence from evaluations to succinctly report on factors related to FAO's contributions to the SDGs³. UN Women changed their approach over time, from one that focused on OECD/DAC criteria and drivers of change, to one that synthesizes insights and evidence on the key types of support provided by UN Women during the Strategic Plan period 2018 to 2021.

Ratings associated with these criteria were not consistently used, with some agencies only providing qualitative evidence from selected evaluation reports, some agencies referring to the use of a four-point scale to rate performance, and others to a six-point scale, while other reports did not provide this information. Annex 1 provides an overview of selected UN organizations have used and rated evaluative data to provide aggregated measures and grouped different criteria to examine trends.

► **Several agencies, such as UNEP, UN Women and UNESCO, had annual or biennial reports synthesizing information from project evaluations as a proxy for assessing project performance, in a similar way to the ILO, yet very few of these reports score performance.**

► The [UNEP Evaluation Synthesis Report](#)⁴ summarizes and rates performance criteria based on the **OECD/DAC evaluation criteria** using a six-point scale, for all independent evaluations undertaken by the Evaluation Office over a biennium. There are no qualitative details regarding specific projects in the presentation of results⁵. Overall performance is also calculated using a weighted ratings scale, and the 2020-21 report assessed trends in overall performance over the last five biennia. In addition, the report also presents highlights from evaluations and reviews with high strategic importance, but there is no narrative linking these to the performance ratings of independent project evaluations.

3 The report was centered around contributions to Hunger, food insecurity and malnutrition; Sustainable agriculture, environment and climate action; Supporting people in crisis settings; Leaving no one behind; and Enabling Factors

4 https://wedocs.unep.org/bitstream/handle/20.500.11822/40650/2020-2021%20Biennial%20Evaluation%20Synthesis%20Report_Aug2022.pdf

5 The 2020-21 report rated 42 independent project evaluations.

- IFAD's "Annual Report on Results and Impact of IFAD Operations" (ARRI), subsequently renamed "Annual Report on the Independent Evaluation of IFAD" (ARIE), compares findings across evaluations and presents a time-series of ratings to identify performance trends, based on country strategy and programme evaluations and project-level evaluations. It also highlights evaluation findings on key themes central to IFAD's mandate, including from corporate-level evaluations and evaluation syntheses. Analysis focuses on the **OECD/DAC criteria** of relevance, effectiveness, efficiency, coherence, sustainability and impact, as well as IFAD-specific criteria. Different aggregate measures are used, as presented in Annex 1.
- The UN Women 2017 meta-analysis⁶ combines qualitative analysis of 39 evaluation reports with a quantitative assessment of development effectiveness based on the OECD/DAC criteria and methods. Lessons, insights, and good practices are highlighted throughout the report, including key strategies for change and external drivers of UN Women performance. Performance is assessed qualitatively and quantitatively, using a 4-point scale.
- The 2022 meta-synthesis of UN Women evaluations presents qualitative evidence and lessons on **the key types of UN Women support provided during the 2018-2021 Strategic Plan period**, based on 67 evaluations. Evaluative evidence is organized around normative support; integrated policy advice and thought leadership; capacity development and technical assistance; and advocacy and social mobilization. Key insights, good practices and impediments are highlighted throughout, with a geographic and thematic focus. There are no ratings in the report.
- The annual [Synthetic Review of UNESCO Evaluations](#) draws from evaluative evidence from five corporate and 22 decentralized evaluations to highlight cross-cutting areas of achievement, gaps, opportunities and other programming/operations issues; to present trends in performance against the standard **OECD/DAC evaluation** criteria; and to show how the work covered by the yearly evaluation portfolio contributes to the UNESCO Programme and Budget and the Expected Results of its Major Programmes. While there are no references to the use of ratings or quantitative data, aside from the Quality Assurance exercise, qualitative evidence from specific projects is used throughout the different sections.

⁶ See Meta-synthesis of UN Women evaluations 2017/2018 - <https://www.unwomen.org/en/digital-library/publications/2019/08/meta-synthesis-of-un-women-evaluations-2017-2018>

⁷ <https://www.unwomen.org/sites/default/files/2022-05/Evaluation-Meta-synthesis-of-UN-Women-evaluations-en.pdf>

What can ILO learn from these outputs and approaches?

The criteria used in the ILO effectiveness reviews allow readers to know where ILO stands in relation to most of the OECD/DAC evaluation criteria, as well as to some ILO cross-cutting policy issues and development cooperation concerns. However, certain performance criteria are not included in the current framework. The following sections draw from what is being done in other agencies to make suggestions to provide a more adequate picture of ILO's overall performance. These are areas where further investigation is recommended for more effective use of the ILO's effectiveness assessment framework. This would allow EVAL to go beyond assessing how well work is done – in terms of doing things right and meeting project objectives – to also gain insights on what work is being done, and whether it is helping to meet the ILO's strategic objectives.

Including additional performance criteria:

Coherence

One OECD/DAC criterion which is not fully addressed in the current ILO effectiveness reviews⁸ is **Coherence**, both in terms of internal coherence and external coherence, including complementarity with existing interventions.

► **COHERENCE: OECD definition⁹:**

The compatibility of the intervention with other interventions in a country, sector or institution.

Note: The extent to which other interventions (particularly policies) support or undermine the intervention, and vice versa. Includes internal coherence and external coherence:

Internal coherence addresses the synergies and interlinkages between the intervention and other interventions carried out by the same institution/government, as well as the consistency of the intervention with the relevant international norms and standards to which that institution/government adheres.

External coherence considers the consistency of the intervention with other actors' interventions in the same context. This includes complementarity, harmonization and co-ordination with others, and the extent to which the intervention is adding value while avoiding duplication of effort.

⁸ The methodological framework was developed before the OECD/DAC evaluation criteria were modified.

⁹ Better Criteria for Better Evaluation. Revised Evaluation Criteria Definitions and Principles for Use. OECD/DAC Network on Development Evaluation (2019)

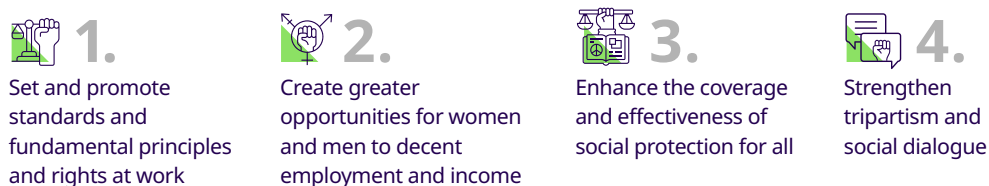
Such a criterion would rate how well the intervention design and implementation has taken into account ongoing and planned initiatives, to ensure that synergies are maximised, and duplication of efforts is avoided.

Currently, some of the information related to internal and external coherence is partially captured by some of the criteria and associated rubrics, including Causal link with DWCP outcomes and CPOs (Criterion 1.2); Strategic partnerships (Criterion 2.8); Internal ILO coordination (Criterion 3.4); and Linkages to SDG targets (Criterion 4.1). Certain evaluation reports provide evidence related to this criterion, but there is currently no space to fully reflect these findings in the current framework. Including this dimension would be especially pertinent in the context of cluster, regional, thematic and joint evaluations.

Links to the Decent Work Agenda

Another aspect of ILO performance that could be more strongly captured is related to intervention results and progress towards the ILO's strategic goals. In particular, evidence provided in the current report does not provide sufficient information on Policy Outcomes and the four strategic objectives at the heart of the Decent Work Agenda.

► **Figure 2. Strategic objectives of the Decent Work Agenda**



The current framework allows for an assessment of Strategic objectives 1 and 4, through performance criteria 2.5 and 2.9 respectively, while the other two strategic objectives are not captured and reported on in a systematic manner, although Strategic objective 2 is partially included in performance criteria 1.5 and 1.6. While evaluation reports often provide details regarding these goals, they are not aggregated in a manner that allows for a rapid assessment of performance in this regard.

One way to address this would be to include a specific theme related to the DWA strategic objectives in the meta-data used when assessing an evaluation report. Each evaluation report could be associated with one or several DWA strategic goals, as applicable, based on the main objectives of the intervention evaluated. This would allow to have an overview of the number of evaluations related to each strategic goal in a given timeframe. If one wished to go further, an analysis of performance could be performed for each DWA strategic goal, by selecting specific reports based on the meta-data.

Another way to address this would be to rate each project using a composite weighted score, similarly to what is done by UNEP, and link the project to its Country Programme Outcomes (CPOs), to report on performance by Policy Outcome and/or Priority Area.

Changes in circumstances

The evidence gathered under the performance area of Relevance and Strategic alignment does not capture whether intervention objectives and design continue to be relevant if circumstances change. This is addressed by several UN agencies, such as UNEP, that has a specific criterion on “Nature of External Context”, which is not used to rate overall performance, yet allows to take unexpected disruptions into account. This can provide a more nuanced qualitative analysis in the main report and would be particularly pertinent in the context of COVID-19 and interventions in complex socio-political situations, for instance.

Including “Those at risk of being left behind”

As the ILO aligns with the objectives of Agenda 2030, the topic of “those left behind” could be better integrated and reported on based on the existing “Pro-poor focus” criterion (criterion 1.5) as well as the criterion on Disability Inclusion (criterion 4.2). The use of an aggregated performance metric including criteria 1.5 (Pro-poor focus), 1.6 (Gender responsiveness) and 4.2 (disability inclusion) could be explored further.

Innovation

Another aspect of ILO performance that is not captured in the current criteria is the promotion of innovative practices - both in project operations and internally. Several UN agencies, including UNDP, IFAD and UN Women, include this in their performance assessments in several different ways under external coherence; effectiveness – prioritisation of development innovation; and efficiency – looking at the use of innovative management and work practices to address challenges or foster efficient delivery of results. While including a specific criterion on innovation may skew performance results for otherwise strong projects that successfully use existing ILO approaches, rather than seek innovative practices, this aspect could be better integrated in the overall narrative, through specific boxes or examples, to showcase results from pilot projects and reflect on innovative approaches.

Advocacy

One area of support that is not reflected in the existing framework is advocacy, although related elements may be captured through ILO's capacity to build and mobilise different relationships, establishing dialogue with key actors, and through communication efforts with different stakeholders.

Going beyond independent project evaluations

While the number of evaluation reports used in performance-related meta-analyses has varied over time, with increased coverage, the type of evaluative evidence has not changed: only evidence from independent project evaluations has been used to assess and report on performance. Examining approaches from other UN agencies has revealed that including evidence from strategic/policy-level evaluations and associated synthesis reviews may give increased insights into ILO performance. While it would not be possible to include these in the current rating framework for analysis, and such reports could not be used to rate performance related to individual criteria, qualitative evidence from these sources could provide rich insights into overall performance and complement results from the meta-analysis.

Are there different angles that could be used to analyse and disseminate results on performance?

Is the current format adequate to inform decision-making?

Two main outputs are produced based on the meta-analysis of ILO effectiveness assessment for a given period:

1. Key findings from the meta-study feed into a section of the ILO's Annual Evaluation Report (AER); which is presented to the Governing Body. To provide information on the ILO's overall effectiveness, the report has graphics showing "Successful" and "Highly successful" ratings for "Strategic relevance and alignment"; "Effectiveness, sustainability and impact", and "Implementation and efficiency of management", as well as "Specific development cooperation concerns". Median scores by performance category are also presented. In the 2022 Annual Evaluation Report, a comparative assessment of performance ratings from 2017-2022 was included, with results disaggregated by technical department and region; this was a target for 2022 in the ILO's Evaluation Strategy.
2. A more comprehensive report, presenting findings from the different performance areas, based on the same ratings as for the AER, supplemented by qualitative evidence from the assessed evaluation reports for each evaluation criterion. The report also includes a section on factors contributing to the effectiveness of the most successful projects, a comparison of median performance over time; and the latest report included key highlights from the comparison of effectiveness results by technical intervention area and region, over the past five years, as well as recommendations for a way forward. The report is made available on the ILO's website, together with a Quick Fact sheet, and shared with all ILO staff through EVAL's internal Broadcast and external audiences through its widely disseminated Newsletter.

The structure of the “Decent Work Results and Effectiveness of ILO Operations Meta-analysis” report is driven by the three main performance areas of Strategic relevance and alignment; Effectiveness, sustainability and Impact; and Implementation performance and efficiency of management and resource use; and their associated criteria, as well as the three more recent criteria associated with specific development cooperation concerns: SDGs, disability inclusion and environmental sustainability.

The existing approach to assess ILO's effectiveness includes performance-related evidence about what the ILO does, and how it does it. The three main performance areas and the criteria related to specific development concerns cover a combination of OECD-DAC-related criteria, as well as cross-cutting issues and approaches that are at the core of how ILO works, including capacity-building, knowledge development, social dialogue and tripartism, and normative work and ILS, for instance, that influence criteria such as policy influence, strategic importance of results achieved and sustainability. This results in a combined assessment of substantive results and operational performance, especially under the second performance area of Effectiveness, sustainability and Impact.

The current approach principally takes an accountability perspective, providing concrete reflections of how ILO performed with regards to the different evaluation criteria, with rich, illustrative evidence from individual evaluations throughout the report. However, specific contributions that could foster learning and facilitate decision-making are not reported as explicitly, although evidence is provided in the main report, either in the narratives, or in boxes.

Although this structure is comprehensive and systematic, **it is suggested to present the information in a different way, by initially focusing and reporting on key strategic areas of interest based on evaluative evidence, highlighting areas of success, challenges, lessons learned and examples of innovating work, followed by the analysis of performance results based on the OECD/DAC criteria, and drawing from high-level evaluations could be a way to facilitate learning and decision making.** This has been the approach taken in recent years by some agencies, such as UN Women, UNESCO, and ITC. In some cases, these meta-analyses only focus on a few thematic areas of interest, while in others, they are more comprehensive. Nevertheless, it is worth noting that the majority of these meta-analyses did not report on overall performance or aggregated results, unless they were presenting trends over time.

Figure 3 below provides suggestions to cluster criteria used in the assessment framework in new ways, to improve reporting on performance based on evaluative evidence and associated ratings. This may provide a more nuanced account of the different types of achievements of ILO interventions and to reflect on different areas of ILO's performance. Some of these suggestions take into account the changing maturity of certain criteria .

► **Figure 3. Suggested clustering of criteria**

Strategic relevance 1.1, 1.2, 1.3, 1.4, 4.1	Coherence (internal and external) 1.2, 2.5, 2.8, 4.1	Effectiveness 2.1, 2.2
Sustainability and impact 2.6, 2.7, 2.10	Means to achieve results 2.3, 2.4, 2.5, 2.8, 2.9, 3.6	Including those at risk of being left behind 1.5, 1.6, 4.2
Cross-cutting policy areas 1.6, 2.5, 2.9, 4.2, 4.3 (possibly 1.5)	Operational efficiency 3.1, 3.2, 3.3, 3.4, 3.5	Resource efficiency 3.7, 3.8 (possibly 2.12)

In order to strengthen the messages around performance based on the synthesis and rating evaluative evidence, it is strongly suggested that EVAL moves away from a reporting format that focuses mainly on the OECD/DAC criteria, and takes a more thematic approach, aligned with ILO strategic priorities. The areas of focus could vary over time, while keeping a section on the different OECD/DAC and ILO-specific criteria. For instance, upcoming meta-analysis reports could focus on the four Priority Action Programmes, keeping in mind that interventions evaluated in 2023 were not necessarily designed based on these priorities:

- Transition from the informal to the formal economy
- Just transitions towards environmentally sustainable economies and societies
- Decent work outcomes in supply chains
- Decent work in crisis and post-crisis situations

Presentation of performance results based on CPOs, Policy Outcomes, and ILO Priority Action Areas, would allow to track progress and generate lessons related to the implementation of these objectives. Alternatively, reporting could include key strategic areas, and be supplemented by findings related to the clusters presented in the figure above.

In terms of communication products, a short, interactive document adhering to EVAL's communication strategy, to highlight key facts, with direct links to the longer report, could improve knowledge-sharing and dissemination to a wider audience.

► **Annex 1**

Agency	Criteria used and rating system (as applicable)	Type of aggregation	Grouping by time, departments, or other dimensions
IFAD	Performance is summarized by OECD/DAC criteria and related sub-criteria, including Relevance; Effectiveness (sub-criteria: innovation); Efficiency; Impact; Sustainability (including Scaling up, National resource management, and adaptation to climate change); Coherence (including “non-lending activities” which cover the sub-domains of knowledge management, partnership development and policy engagement). IFAD-specific criteria are: Gender equality and women's empowerment; and Partner performance (assessed separately for IFAD and Governments) IFAD uses a six-point rating scale to assess performance under each evaluation criterion, according to which ratings 4 to 6 signal positive performance, while ratings 1 to 3 mark weak performance.	► Project performance: an arithmetic average of the ratings related to the core performance criteria of relevance, effectiveness, efficiency and sustainability of benefits ► Overall project performance: arithmetic average of all the key evaluation criteria ratings for a programme or project, based on the OECD core performance criteria and IFAD-specific criteria, rounded to the second digit ► Overall project achievement: based on the analysis and ratings of all criteria, except for the performance of IFAD and of government. ► Non-lending activities: partnership building, country-level policy engagement, and knowledge management.	Performance related to ITAD-specific criteria over time: When looking at trends over time, IFAD uses an average rating for project performance by three-year moving period based on an arithmetic average of the ratings for the four core criteria, by year of evaluation. Overall project performance by year of project completion Project performance on selected criteria by IFAD regional divisions Combined criteria: rural poverty impact and scaling up, sustainability, and efficiency
ITC	Performance is summarized by OECD/DAC criteria of relevance, coherence, effectiveness, efficiency, impact and sustainability, with qualitative narratives for each criterion and sub-criteria. A six-point rating system is used to assess performance related to each criterion. Along with a qualitative justification, a quantitative measure based on the achievement of planned targets set out in the logical framework, and a composite rating for overall project performance based on consideration of the individual ratings are to be provided. When logical frameworks are not available, only a qualitative assessment on each of the evaluation criteria, and an overall rating of the project, is expected.	► Findings from all evaluation reports are aggregated and presented using a color-coded table to provide a baseline assessment for the implementation of ITC's strategic plan 2022-25. The assessment includes four colors: ► Green: well on track for achieving the Strategic Plan 2022-25; Orange: on track for achieving the Strategic Plan 2022-25, but with moderate risks; Red: not on track; Grey: inconclusive / no assessment possible due to lack of data. ► The baseline for assessing the ITC Strategic Plan 2022-25 focuses on Vision, mission and principles; the Matrix approach and framework for MSMEs; Responding to country needs; Engaging with the WTO and the UN (partnering with purpose); Leading the global conversation; Purpose-driven identity, including cross-cutting objectives; Organizational strengths/Value proposition; Resource mobilization. ► A methodological framework has been developed. Each of its fields contains criteria, in the form of research questions, against which the respective Strategic Plan objective (columns) can be assessed in relation to each of the OECD/DAC evaluation criteria (rows).	

UNEP	Project performance is summarised by OECD/DAC criteria and related sub-criteria, including Strategic Relevance; Quality of project design; Effectiveness; Financial management; Efficiency; and Sustainability. Factors affecting performance of interest to UNEP are also included: Environmental and social safeguards; quality of project management and supervision; stakeholder participation and cooperation; communication and public awareness; country ownership and driven-ness; preparation and readiness; human rights and gender equity)	► Individual evaluation criteria are aggregated using a weighted ratings scale to derive an “overall performance rating”. ► Greatest emphasis is placed on the achievement of outcomes (30%), sustainability/ durability of outcomes (20%), thus reflecting an emphasis on assessing performance from a results perspective rather than activities perspective.
------	--	---

► Sources consulted

IFAD (2022) [2022 Annual Report on the Independent Evaluation of IFAD](#)

IFAD (2022) [Evaluation manual, 3rd edition](#)

IFAD (2015) [Evaluation manual, 2nd edition](#)

International Trade Center (2022) [Annual Evaluation Synthesis Review](#)

UNEP(2022) [Evaluation Synthesis Report 2020-2021](#)

UN Women (2020) [Insights on organizational effectiveness and efficiency: a meta-synthesis of UN Women Evaluations](#)

UN Women (2021) [Transform Magazine](#)

UN Women (2019) [Meta-synthesis of UN Women Evaluations – 2017/2018](#)

UN Women (2019) [Transform Magazine](#)

FAO (2021) [Programme Evaluation Report](#)

UNESCO (2022) [2022 Synthetic Review of UNESCO Evaluations](#)

UNDP (2022) [UNDP Country Programme Performance Rating System Manual](#)

UNHCR (2021) [UNHCR Evaluation Service Annual Report](#)

UNDOC (2022) [Meta-synthesis: 2019-2021 UNDOC Evaluations](#)

WFP (2022) [Annual Evaluation Report](#)

WFP (2023) [Evaluation Synthesis of WFP's Performance Measurement and Monitoring 2018-2021](#)

WHO (2021) [Synthesis of WHO country programme evaluations](#)

[Better Criteria for Better Evaluation. Revised Evaluation Criteria Definitions and Principles for Use. OECD/DAC Network on Development Evaluation \(2019\)](#)

Mathiason et al. (2010) Independent External Evaluation of the International Labour Office Evaluation Function

ILO Evaluation Unit (2011) Decent work results and effectiveness of ILO technical cooperation: a meta-analysis of project evaluations, 2009-2010.

ILO Evaluation Unit. Decent work results and effectiveness of ILO technical cooperation: a meta-analysis of project evaluations, 2011-2012.

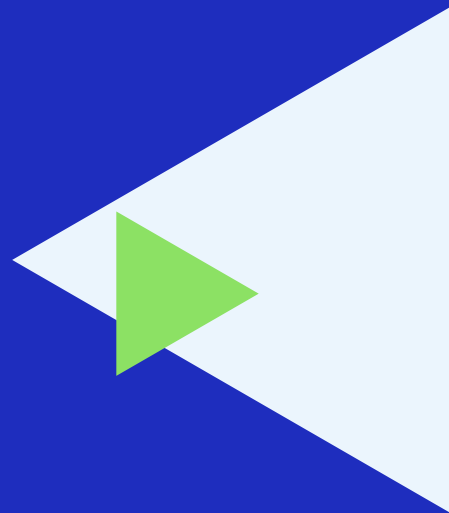
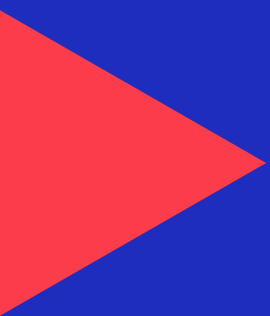
ILO Evaluation Office Decent work results and effectiveness of ILO technical cooperation: a meta-analysis of project evaluations, 2013-2016.

ILO Evaluation Office Decent work results and effectiveness of ILO operations: ex-post meta-analysis of development cooperation evaluations, 2017-18

ILO Evaluation Office. Decent work results and effectiveness of ILO operations: Meta-analysis of development cooperation evaluations, 2019-2020

ILO Evaluation Office. Decent work results and effectiveness of ILO operations: Meta-analysis of development cooperation evaluations, 2020-2021

ILO Evaluation Office. Independent review of effectiveness of Regular Budget Supplementary Account (RBSA)- supported interventions in delivering Decent Work Results, 2013-17.



ilo.org

International Labour Organization
Route des Morillons 4
1211 Geneva 22
Switzerland