WHAT WORKS
In Youth Employment

International Labour Organization

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes

Note 4

Enhancing youth employment learning through evaluation

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes

NOTE 4
Enhancing youth employment learning through evaluation

First published 2018

Guide on Measuring Decent Jobs for Youth - Monitoring, evaluation and learning in labour market programmes. Note 4: Enhancing youth programme learning through evaluation/ International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131662-6 (print)
978-92-2-131663-3 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلّم في برامج سوق العمل النشطة - الملاحظة 4: تحسين التعلّم الوظيفي للشباب من خلال التقييم, ISBN 978-92-2-630793-3 (print),978-92-2-630794-0 (web pdf), Geneva, 2018.

Design and layout by the International Training Centre of the ILO, Turin - Italy

Printed in Switzerland

Contents

Figures

Tables

Enhancing youth programme learning through evaluation

Prerequisites:

This chapter requires no prior knowledge. It introduces readers to the concept of evaluation, and the different types of evaluation options available to youth employment interventions.

Learning objectives:

At the end of this note, readers will be able to:

- understand why to evaluate, the main evaluation objectives of learning and accountability and the different internal and external audiences for evaluations
- formulate evaluation questions based on the core criteria of relevance, impact, effectiveness, efficiency and sustainability
- choose the appropriate type of evaluation, in line with intervention needs and the evaluation context – including performance evaluation, impact evaluation and cost-effectiveness and cost-benefit analyses.

Keywords:
Evaluability, relevance, effectiveness, efficiency, sustainability, impact, stakeholder consultation, descriptive research, normative research, causal research, performance evaluation, impact evaluation, cost-benefit, cost-effectiveness.

Good monitoring systems are critical to knowing whether our intervention is moving in the intended direction. However, they do not necessarily answer the question of how or why changes are coming about, nor do they provide credible evidence that any observed changes in outcomes are the result of our intervention. To complement the information we obtain from our monitoring, we need evaluations. There are different types of evaluation and the extent to which we want to rely on one or several of them will depend primarily on our information needs.

Before analysing how our learning objectives, together with the operational context, inform the choice of evaluation type(s) for our programme, project or intervention, we reflect on different motivations for and potential benefits of conducting evaluations. Throughout this note we focus on possible avenues for evaluating youth employment programmes.

Why conduct an evaluation?

Evaluations that build on well-designed results measurement systems are a critical means of improving decision-making, generating knowledge and providing verifiable evidence of effectiveness of the interventions we implement (ILO, 2017). There are two major goals in conducting evaluations: learning and establishing transparency. Both objectives can be realized with respect to both internal and external audiences, resulting in the four main benefits that evaluations generate (see figure 4.1).

Evaluations can help project management: By assessing the design, implementation or results of a youth employment intervention, evaluations enhance internal, organizational learning. They are, first and foremost, about learning for the benefit of our own project and organization. Furthermore, evaluations provide programme managers with the information needed to make strategic decisions about necessary changes in project design, planning or implementation. By examining how and why certain results were achieved, evaluations complement performance monitoring systems and results-based management

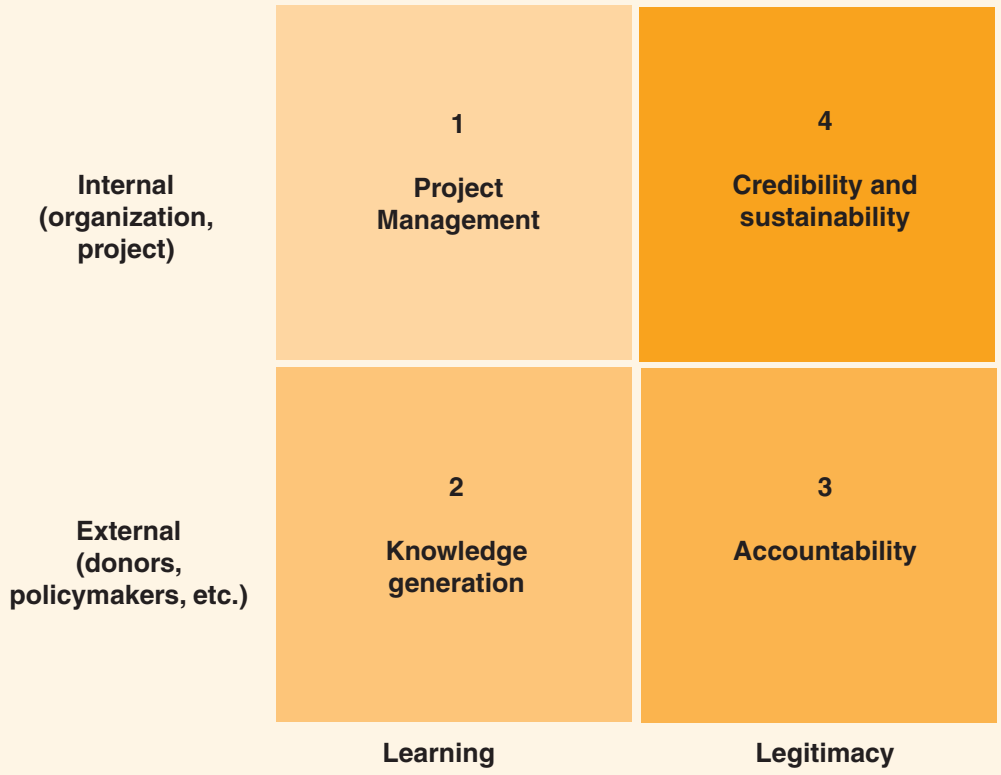
DEFINITION

Evaluation is an evidence-based assessment of strategy, policy or programme and project outcomes, by determining their *relevance*, *impact*, *effectiveness*, *efficiency* and *sustainability*. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors.

(RBM) practices. Although evaluations in general (and impact evaluations in particular) produce information periodically rather than continuously, they are nevertheless valuable parts of the project cycle. A well-designed evaluation helps practitioners to identify barriers to young people's access to employment, unexpected gaps or relevant contextual influence factors that were not diagnosed at the design stage (see Note 1), and address those issues, while sustaining programmes that are, or could be, achieving good results. Without this reflection and corresponding internal knowledge management, programmes run the risk of steering in the wrong direction, missing out on relevant barriers to the achievement of their objectives and repeating mistakes within the same institution. Thus, evaluations allow us to show the true value of our work and inform the design and planning of other interventions.

Evaluations generate knowledge: The youth employment field is characterized by a lack of sound evidence on what works, what does not work and what are the salient reasons for success and failure of interventions. Acquiring this knowledge typically demands evaluations that use specific methodologies to provide credible estimates of the success of particular interventions. In a systematic review of available evidence from impact evaluations, Kluve et al. (2016) find that ALMPs for youth do have a positive impact on labour market outcomes and that the type and design of interventions matter. To understand what

FIGURE 4.1: BENEFITS OF CONDUCTING EVALUATIONS



works, in which context, for whom and why is crucial for the replication of good practice. Large evidence gaps, in particular for youth employment interventions in low- and middle-income countries, remind us that much more knowledge is needed and promoting external learning through evaluations (and the dissemination of their results, lessons to be learned and implications) remains important.

Kluve et al. (2016) document the fact that most of the available evaluations are in the area of training and skills development, while evidence on all other types of interventions, such as subsidized employment for youth, employment services, youth entrepreneurship, youth-inclusive financial services and targeted programmes for excluded groups, is relatively scarce. Moreover, there is growing evaluation evidence from youth employment programmes implemented in sub-Saharan Africa, but limited information from the Middle East and North Africa, South Asia and East Asia and the Pacific regions.

Evaluations fulfil accountability requirements: Evaluations expose programmes to external scrutiny and hold programmes to account over the achievement of results and use of resources. Given the scarcity of resources, it is imperative to ensure that money, time and effort are spent in the best possible way to help young people find and create decent work, be it through training young people, linking them to job opportunities, providing them with business opportunities or offering counselling on career paths. It is important, and increasingly necessary, to be accountable to the ultimate beneficiaries who, for instance, might spend their time (and sometimes money) in participating in a youth employment progamme, or various constituents of our organization, such as governments and social partners, civil society and donors.

Evaluations can help to establish credibility: Evaluations increase the transparency of the project and, thereby, the reputation of the implementing organization. The simple fact that an organization or project agrees to carry out independent evaluations that follow predetermined protocols already indicates high standards in programming. If the evaluation shows good results, then the payoff for both organization and programme can be immense. Proving that "our" method and approach is working can make a big difference in the eyes of donors and policy-makers,

Box 4.1: Youth Employment Evidence Gap Map

The International Initiative for Impact Evaluation (3ie) has developed a Youth Employment Evidence Gap Map. Structured as an interactive matrix, the map provides an accessible overview of evidence from systematic reviews and impact evaluations for programme developers and evaluators. Evidence is presented by major ALMP category (skills training, entrepreneurship promotion, employment services and subsidized employment) as well as by outcome category (employment, earnings, business performance). See http://gapmaps.3ieimpact.org/evidence-maps/youth-employment-evidence-gap-map

who, prior to the evaluation, were unable to differentiate the impacts of our intervention from the alleged impacts of numerous other programmes

All evaluation results convey learning: While, positive evaluation results can be used in advocacy efforts to obtain greater support from donors, governments and the general public, negative findings are arguably even more useful in helping to iterate and refine intervention modules. As in any other field, building successful products and services requires testing, prototyping and adapting to local circumstances. Failures are a necessary step toward state-of-the-art programming. And, especially when evaluation is not simply asking what works or doesn't work but also how and why certain approaches work, negative results will improve both programme design and ongoing operations. If, early in the process, we are able to understand the problems that may reduce the effectiveness of our intervention, then we are in a good position to build successful projects in the long run.

Box 4.2: The link between programme design and evaluation

There are important linkages between programme design and evaluation. One of the major roles of evaluation is to support learning and, in turn, future planning. The usefulness and feasibility of the evaluation is highly dependent on the quality of the original programme design, as set out in Note 1. Bear the following points in mind:

- **Evaluation does not make up for poor design:** Later evaluation does not replace early thinking. A carefully thought-out programme design, based on existing research and experience, puts the programme on the best possible path for success.
- **The evaluation strategy will depend on the knowledge gaps identified during the design stage:** Knowing the evidence base and identifying potential knowledge gaps are important factors in choosing the right evaluation strategy. For example, impact evaluations will be particularly valuable for innovative and untested programmes that provide an opportunity to fill global knowledge gaps.
- **The right programme design can facilitate evaluation:** Some programmes are easier to evaluate than others. For example, if an impact assessment is not planned during the design stage of the programme, the tools available to conduct the evaluation may be severely constrained. However, choosing clear, fair and transparent targeting criteria, such as random assignment for oversubscribed programmes or eligibility scores, can significantly ease the evaluation process. Thus, if there are multiple acceptable ways of delivering a particular programme, it may be wise to plan ahead and choose a design that also suits the evaluation.

Assessing the evaluability of our intervention

Before starting on the evaluation activities, it is advisable to assess whether it is actually possible to evaluate the intervention under consideration. A scoping exercise can determine how well the intervention, as planned, can be evaluated. **Evaluability** assessments can be carried out for both performance and impact evaluations (Davies, 2013). In both cases, the assessment establishes whether an intervention features the necessary requirements for a performance or impact evaluation to be conducted in a way that will generate useful results. Moreover, the coherence and logic of an intervention, project or programme is reviewed and the timing of the planned evaluation, the political, social and economic context, as well as the availability of sufficient resources for the evaluation, are all taken into account.

Assessing evaluability for performance evaluation also includes clarifying both data availability and the adequacy of available data for reflecting progress towards results.

Furthermore, such a scoping exercise can assess the presence of a well-developed results measurement framework and the availability and quality of baseline data for assessing changes (Note 3). For impact evaluations, these assessments also consider how likely it is that an impact study will lead to real improvements in programme performance and success and, thus, whether the costs in terms of effort and money will realistically be outweighed by the anticipated benefits. Following this assessment, the scope, methods and timing of the evaluation can be adjusted accordingly (ILO, 2017).

TIP

The ILO has developed a methodology to systematically assess evaluability, based on best practices among OECD/DAC members. The evaluability instrument scores individual projects and programmes based on objective, indicators, existence of baseline data, milestones, risks and assumptions, as well as monitoring and evaluation. Detailed guidance on this tool is available in ILO (2017). Further information can be requested from the ILO's Evaluation Office.

DEFINITION

Evaluability is the extent to which an activity or programme can be evaluated in a reliable and credible fashion.

Evaluation criteria and evaluation questions

As already mentioned, evaluations are periodic assessments of the **relevance, effectiveness, efficiency, impact and sustainability** of our intervention. These are the common **evaluation criteria** as originally defined by the Organisation for Economic Co-operation and Development (OECD) (see table 4.1 for more detail).

Table 4.1: Evaluation criteria

Criteria	Description
Relevance	The extent to which the objectives of an intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies
Effectiveness	The extent to which the intervention attains its objectives The major factors influencing the achievement or non-achievement of the objectives
Efficiency	A measure of how economically inputs (money, expertise, time) are converted into results If possible, an indication of costs per output/beneficiary should be determined and a comparison made with similar interventions
Impact	A measure of the positive and negative changes produced by the intervention Estimates of causal differences that the intervention has made to the beneficiaries
Sustainability	The likelihood that the results of the intervention are durable and can be maintained by intervention partners The extent to which benefits of a project will continue after donor input ceases

Source: Based on OECD/DAC, 2002 and ILO, 2017.

Formulate evaluation questions: In practice, we may have many evaluation questions across all criteria that we would like to assess. While evaluation criteria are formulated in a general manner, evaluation questions need to be adapted and tailored to the specific intervention. This also helps to define the menu of appropriate monitoring and evaluation (M&E) tools that will allow those questions to be answered (GAO, 2012). There are several important considerations when formulating evaluation questions:

DEFINITION

Evaluation criteria: To ensure a high quality of evaluation studies, it is good practice to assess a number of standardized evaluation criteria (e.g. relevance, effectiveness, efficiency, impact and sustainability) that are formulated in a general manner.

DEFINITION

Evaluation question: Evaluation questions adopt and tailor evaluation criteria to the specifics of the intervention.

1. **Involve stakeholders:** Before getting started, it is important to identify the audience for the evaluation and what that audience wants to know. The set of stakeholders depends on the intended goals of the evaluation (see the section "Why conduct an evaluation?" above) and typically includes project managers and staff members in the originating organization as well as key national and local partners (for example, line ministries and social partners involved) as well as donors. Involvement of our target audience to jointly identify evaluation questions is important for the assessment of all evaluation criteria and particularly helpful for evaluating the relevance and sustainability of an intervention.

2. **Choose complementary questions for each evaluation criterion:** All types of evaluation questions examine different aspects of the project and provide different kinds of information. Rather than being substitutes, they often complement or build on each other. For most assessed evaluation criteria, it makes sense to ask a mixture of *descriptive*, *normative* and *cause-and-effect* questions (Imas and Rist, 2009):

   - **descriptive questions** seek to define processes, conditions, organizational relationships and stakeholder views (*What is going on in our project?*)
   - **normative questions** compare what is taking place to what should be taking place. They compare the current situation with the specific objectives and targets that have been defined (*Has our project been implemented as intended? Is it performing as expected?*)
   - **cause-and-effect questions** examine outcomes and try to measure the difference that an intervention makes. They ask whether objectives have been achieved *as a result of* our project (*To what extent can we attribute observable change to our intervention?*).

3. **Organize evaluation questions around the results chain, where possible:** If a reliable results measurement system is in place (see Note 3), there should be consensus around our project logic in terms of implementation and results, which in turn helps to identify critical learning objectives of the intervention. Descriptive and normative questions can relate to all levels of the results chain; however, cause-and-effect questions mainly refer to outcomes and impact-level outcomes. This strategy is particularly useful for identifying questions to assess a project's effectiveness, efficiency and impact (see table 4.2).

Table 4.2: Examples of evaluation questions formulated along a results chain

	Inputs	Activities	Outputs	Outcomes	Higher level goals
Descriptive	• How does the cost of the programme compare to similar interventions? • What qualifications do the service providers have? • What other ongoing interventions are there?	• Do youth know about the programme and how they qualify to join? • What delivery mechanisms are being used? • To what extent does the programme implementation differ by site?	• How many youth participate (by age, sex, etc.)? • Who drops out? • Which services are used the most?	• Are participants satisfied with the programme? • Are there any observable changes in participant's basic, technical or core skills? • How many programme participants find employment within 3 months?	• What are the current local youth unemployment rates? • How high is the average household income?
Normative	• Do we spend as much as we have budgeted? • Are the staff and financial resources adequate? • Is the programme duplicating other efforts?	• Is the process for selecting participants fair and equitable? • Is the programme implementation delayed? • Are operational manuals being followed?	• Do we achieve the desired gender balance in participants? • Will we reach the goal of training 5,000 youth per year?	• Does participant income increase by 20%, as planned? • Do 80% of beneficiaries find a job within 3 months of graduation, as required?	• Is local youth unemployment falling, compared to the programme start? • Are household incomes evolving? • Are more households becoming self-sufficient?
Cause- and -effect	• n/a	• n/a	• n/a	• As a result of the job training, do participants have higher paying jobs than they otherwise would have? • Do 80% of beneficiaries find a job within 3 months of graduation due to their participation in the programme? • Does including internships increase the effectiveness of technical training offered? • Does the programme affect boys and girls differently? • What, if any, unintended positive or negative direct effects are there?	• Does the project contribute to reducing poverty in the area? • What other (positive or negative) impacts does this intervention have on the living conditions of the wider community?

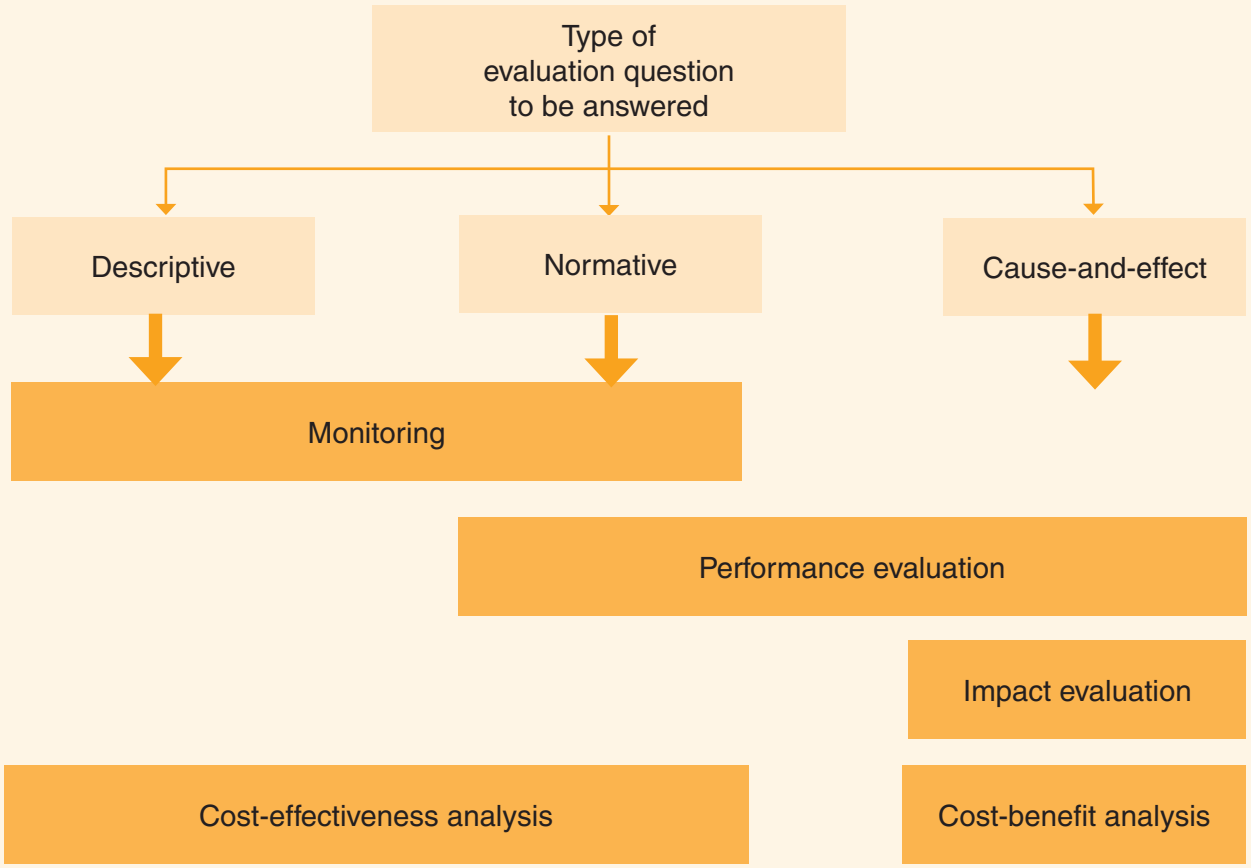
Source: Based on OECD/DAC, 2002 and ILO, 2017.

Linking evaluation questions to evaluation types

There is no one-size-fits-all evaluation template. Ultimately, the choice of the evaluation should depend on the type of information we want to collect (*What do we want to assess: programme progress/performance/attributable impacts?*), as well as on available resources, such as skills, money and time.

Figure 4.2 provides an overview of available evaluation options, depending on the type of questions we want to answer. There are other types of evaluations, which focus on assessing high-level policies, national strategies in a certain sector or thematic area, or macro implications of development cooperation. They are not considered in this note, which is limited to evaluation of projects and interventions that directly impact beneficiaries, such as through ALMPs for young people.

FIGURE 4.2: DIFFERENT EVALUATION TYPES



MONITORING (NO EVALUATION)

If a programme manager requires only descriptive information about the intervention, for example, because the project is in a very early stage and the objective is to obtain some general information about how the programme is being implemented, then a full-fledged evaluation may not be necessary. In that case, the knowledge obtained from monitoring may well be sufficient. Obviously, this requires the existence of a well-functioning results measurement system, with a clearly defined results chain, indicators, data collection tools and so on (see Notes 2 and 3). A solid monitoring system is the backbone of all processes of results measurement and internal learning and provides a firm foundation for conducting thorough and informative evaluations. If such a system is in place, descriptive information about the programme should be relatively east to obtain.

PERFORMANCE EVALUATION

Performance evaluations are a systematic, objective assessment of an ongoing or completed project or programme, its design, implementation and results. They evaluate how well programme objectives have been formulated (see Note 1) and what has been achieved to date. This requires an understanding of whether there are gaps between planned and realized activities and outputs and, if so, why. Building on descriptive information, such as what activities are being conducted and who is participating in the programme (or who is not), performance evaluations identify ways to improve the quality of the services offered. Performance evaluations often take a comprehensive approach and assess most, or all, of the five evaluation criteria (see table 4.1).

This implies that they also assess whether the results framework is appropriate; that is, whether there are inconsistencies between the resources, activities and objectives, and whether priorities or timelines should be adapted to ensure that the agreed objectives are achieved effectively. For this purpose, the whole theory of change is being revised. An important question is whether the envisaged outcomes were achieved. Moreover, performance evaluation also examine in detail the linkages between activities, outputs and outcomes, as well as whether underlying assumptions turned out to be realistic. For example, when conducting a performance evaluation of a skills training programme for unemployed youth, the evaluator will assess the progress of previously defined indicators for outputs (e.g. number of young people trained), intermediate outcomes (e.g. number of applications sent, number of job interviews obtained) and impact-level outcomes (e.g. number of jobs obtained). They will aim to understand the assumptions underlying each

DEFINITION

Performance evaluation: A performance evaluation assesses the quality of the service delivered and the outcomes (results) achieved. It typically covers short-term and medium-term outcomes, but not casual impact.

connection (e.g. participation in the training contributed to an increase in job-search behaviour) and test whether they are applicable and appropriate.

Such evaluations can be carried out across all stages of implementation, but they are particularly common for mid-term reviews (when their focus is on learning for programme management) or at programme completion (when their focus is on accountability and lessons learned for future interventions) (see box 4.3). However, it is important to consider programme evaluation from the planning stage onwards. The formulation of the results chain, the design of SMART indicators (see Note 3) and the gathering of baseline data should be undertaken with due consideration for the type of evaluation to be conducted later on, in order to ensure a coherent process and access to good quality data.

Typically carried out by an independent evaluator, performance evaluations can be implemented relatively quickly and at moderate cost. They tend to rely heavily on desk research (e.g. the analysis of available administrative data), key informant interviews, field visits and observations, as well as focus group discussions. Sometimes, however, performance evaluations may incorporate more extensive data collection, such as the implementation of surveys, a before-and-after comparison of participant outcomes or additional qualitative tools (see figure 4.3 for the key steps in an evaluation).

Box 4.3: Categorizing evaluations by timing

- **Annual reviews** focus on outputs and outcomes of projects, programmes, strategies or policies. They are a form of internal evaluation during which the stakeholders reflect on how well the intervention is progressing towards achieving its objectives, taking into account available M&E data. Reviews with this type of focus may also be organized to look at specific issues.
- **Mid-term evaluations** should take place during the implementation of projects, programmes, strategies or policies. The exact timing will vary and should be flexible, if justified. They are most useful when a number of planned activities have been delivered and a considerable percentage of funds have been spent. Mid-term evaluations aim to assess the continued relevance of an intervention and progress made towards achieving its planned objectives, offering an opportunity to make modifications to ensure they are achieved.
- **Final evaluations** focus on the outcomes of projects, programmes, strategies or policies and the likelihood that they will achieve their intended impact. These evaluations provide an opportunity for in-depth reflection on the strategy and assumptions guiding the intervention. They assess the extent to which an intervention achieved its objectives and may recommend adjustments to its strategy. They are also a means to assess how well intervention-level actions support higher-level strategies and objectives, as articulated in Decent Work Country Programmes (DWCPs) and the ILO's Programme and Budget (P&B).
- **Ex-post evaluations** take place after completion of the project with the aim of assessing longer-term effects of specific interventions. They can be part of strategy/policy, thematic or country programme evaluations that also consider linkages between different interventions and longer-term development outcomes. The primary purpose of these evaluations is to examine the sustained impact of a particular intervention.

Source: ILO, 2017.

Typically, performance evaluations cannot assess if a certain outcome happened explicitly due to the programme activities. Claiming the causal attribution of an observed change for an intervention requires robust impact evaluation methodologies to be applied (see the following section). However, the transition between the two evaluation types is fluid. In particular, qualitative tools can be used in performance evaluations in order to establish a causal link between the project activities and the measured outcome. This is usually done by identifying and revising alternative hypotheses on the causal process underlying the outcome in question. In contexts where a counterfactual-based impact evaluation is not feasible, reasonable or desired, it is recommended that the range of qualitative methods for establishing causal claims in performance evaluations be explored and assumptions about contributions of the intervention to broader development impacts and goals be assessed (ILO, 2017). A useful overview of qualitative methods for addressing cause-and-effect questions that can also form part of a performance evaluation can be found in White and Phillips (2012).

In sum, performance evaluations aim to determine whether a programme is being implemented effectively with regard to its objectives, what is going right or wrong and why, and to generate lessons learned to inform future decision-making processes.

FIGURE 4.3: KEY STEPS IN PLANNING AND MANAGING AN EVALUATION

Source: ILO, 2017.

IMPACT EVALUATION

Impact evaluations answer cause-and-effect questions. Such questions require us to determine not only whether the desired outcomes occurred but also if those outcomes occurred *because the programme was implemented*. In other words, impact evaluations aim to determine whether observed changes in the economic or social well-being of beneficiaries can be attributed to a particular intervention, project or programme (ILO, 2013).

Impact evaluations require methods to explicitly tackle the challenge of isolating the causal effect of an intervention on outcomes of interest. This is usually achieved by constructing a counterfactual scenario, i.e. aiming to answer the question "What would have happened in the absence of the programme?". This is what makes impact evaluations different from performance evaluations and monitoring systems that typically focus on programme beneficiaries alone. They therefore tend to require more time and statistical skill, and they typically cost more than other evaluation types.

Based on the information they provide, impact evaluations are particularly useful for informing strategic approaches, from scaling up effective interventions to curtailing unpromising programmes (Rubio, 2011). Decisions linked to replication and expansion often require evidence on whether smaller pilot projects worked for a comparable population. As governments, donors and non-governmental organizations are accountable for spending scarce resources efficiently, they increasingly demand the reliable evidence that well-conducted impact evaluation can deliver (ILO, 2013). Importantly, the current global evidence base on what works in youth employment is still limited, so more impact evaluations are needed to support external learning processes. Impact evaluations also help us to understand which programme design options (dosage, delivery channel, etc.) are most important within a specific programme category, such as skills training, employment services or entrepreneurship promotion.

There is a huge spectrum of impact evaluation approaches available for exploring the causal effects of youth labour market programmes. Depending on the intervention and context, some methods may be more practical, useful and, at the same time, less costly and less time-consuming than others. A more detailed description of impact evaluation methods, including the advantages and disadvantages of their application in different contexts, can be found in Note 5.

While performance evaluations can be part of every programme, impact evaluations and cost-benefit analyses should be applied more selectively. According to Gertler et al. (2011), the additional effort and resources required for conducting impact evaluations are best mobilized when the programme is (1) strategically relevant and influential, (2) innovative or untested and (3) replicable:

- **Strategically relevant and influential:** How important would the results be for informing future programmes, policies or policy dialogue? If the stakes of an intervention are high, then an impact evaluation should

DEFINITION

Impact evaluation: An impact evaluation establishes a causal link between a programme or intervention and a set of outcomes. An impact evaluation tries to answer the question of whether a programme is responsible for the changes observed in the outcomes of interest.

be considered. This may apply to new initiatives as well as to existing programmes when we need to make decisions about their continuation, expansion or termination. In fact, even an expensive impact evaluation can be highly cost-effective since its findings may help to produce important improvements in programme performance (World Bank, 2009).

- **Innovative or untested:** What is the current state of evidence or knowledge on the proposed programme's impacts? If little is known about the effectiveness of the type of intervention, globally or in a particular context, an impact evaluation can add powerful knowledge to our organization and the entire field. This is the case for most youth employment programmes, for which the evidence base is still slim.

- **Replicable:** To what extent and under what circumstances could a successful pilot or small-scale programme be scaled up or replicated with different population groups? If the programme can be scaled up, or can be applied in different settings, then an impact evaluation is an important step in providing the justification for programme replication. However, aspects and sensitivities related to the specific context, target group or design of the initial pilot need to be carefully analysed before taking a decision on upscaling or replication in order to prevent failure at a larger scale.

COST-EFFECTIVENESS AND COST-BENEFIT ANALYSES

Cost-effectiveness and cost-benefit evaluations assess both monetary and non-monetary programme costs and compare them with alternative uses of the same resources and the benefits produced by the intervention (Baker, 2000).

Cost-effectiveness analysis (CEA) measures the cost per output or outcome (e.g. US$300 per youth trained, US$500 per job created) and compares this cost to similar interventions of our own and other organizations. It thus answers the question of how much output or outcome we get per dollar spent (descriptive) and whether there is a gap between those findings and our expectations (normative).

Cost-benefit analysis (CBA), in turn, weighs the total expected costs against the total expected benefits (outcomes) of an intervention, where both costs and benefits are typically expressed

DEFINITION

Cost-effectiveness analysis (CEA): CEA measures the cost per output or outcome (e.g. $300 per youth trained, $500 per job created) and compares this cost to similar interventions of our own and other organizations.

DEFINITION

Cost-benefit analysis (CBA): CBA weighs the total expected costs against the total expected benefits (outcomes) of an intervention, where both costs and benefits are typically expressed in monetary terms.

in monetary terms. For instance, if our programme were to help 500 youth to find and retain jobs or set up sustainable small businesses, we would (1) estimate the aggregate benefits in terms of higher incomes, better health, lower crime rate etc. and (2) compare these benefits to the overall costs of the intervention.

Two summary measures are typically used in a cost-benefit analysis. The first is a *benefit-cost ratio*. To find this ratio, divide the programme's net benefits by its net costs. The result is a summary measure that states "for every dollar spent on programme X, Y dollars are saved". This type of summary measure is popular with policy-makers because it is readily understandable. If the benefit-cost ratio is greater than US$1, it implies that the programme or intervention produces more benefit than it costs. Another summary measure for benefit-cost analysis is the *net benefits* approach, derived by subtracting net costs from net benefits. According to this method, programmes show a positive return on investment if net benefits are greater than zero.

Historically, major development finance institutions, such as the World Bank, made decisions based primarily on use of cost-benefit analysis. Such analysis served to demonstrate a commitment to measuring results and ensuring accountability. However, the percentage of World Bank projects justified by cost-benefit analysis has been trending downward for several decades, due to both a decline in adherence to policy and the difficulty of applying cost-benefit analyses.

In recent years, the "managing for results" agenda has been dominated by discussions about measuring results, using logical frameworks to frame the monitoring and evaluation efforts, and impact evaluation to measure impact in a more accurate and rigorous way. These efforts complement each other. Yet, in practice, they are often treated separately, leading to unnecessary fragmentation.

According to the World Bank's Independent Evaluation Group (IEG, 2010), cost-benefit analyses often do not mention or use impact evaluation results, despite the fact that measurement of benefits against the counterfactual is integral to cost-benefit assessment. Similarly, it is rare to find impact evaluation studies that embed the results they obtain in a cost-benefit setting. For example, suppose that an intervention is designed to raise youth incomes. The value of the increase in income would be part of the benefit flow in the cost-benefit analysis, and the researcher would typically make an informed estimate of the income that would have accrued without the intervention and then compute the value of the change in income. An impact evaluation would provide that figure more accurately. Similarly, an impact evaluation that went no further than providing estimates of the increase in income would be an incomplete evaluation for decision-makers who want to know whether to repeat the project. What was the value of the increase in incomes, how does that compare with the costs? In this example the information from the impact evaluation and the cost-benefit framework complement each other and provide better analysis.

Understanding evaluation needs based on the operational context

After formulating evaluation questions and identifying potential types of evaluation, we need to explore which evaluation can be carried out under which conditions. This implies analysing the operational context of our intervention to gain a clearer understanding of how time and resource constraints, as well as political considerations, might affect any potential evaluation. This also involves taking factors outside our intervention into account.

TIMING

Early identification of evaluation demands: Planning an evaluation should ideally be part of the programme planning phase. However, often information needs may arise suddenly; for example, as a result of unexpected problems on the ground, or a request from a donor. Similarly, operational constraints, such as having to implement a programme quickly to disburse funds, may dictate the timetable for evaluation. These constraints are unavoidable in real life but reduce the options for evaluation that may be available.

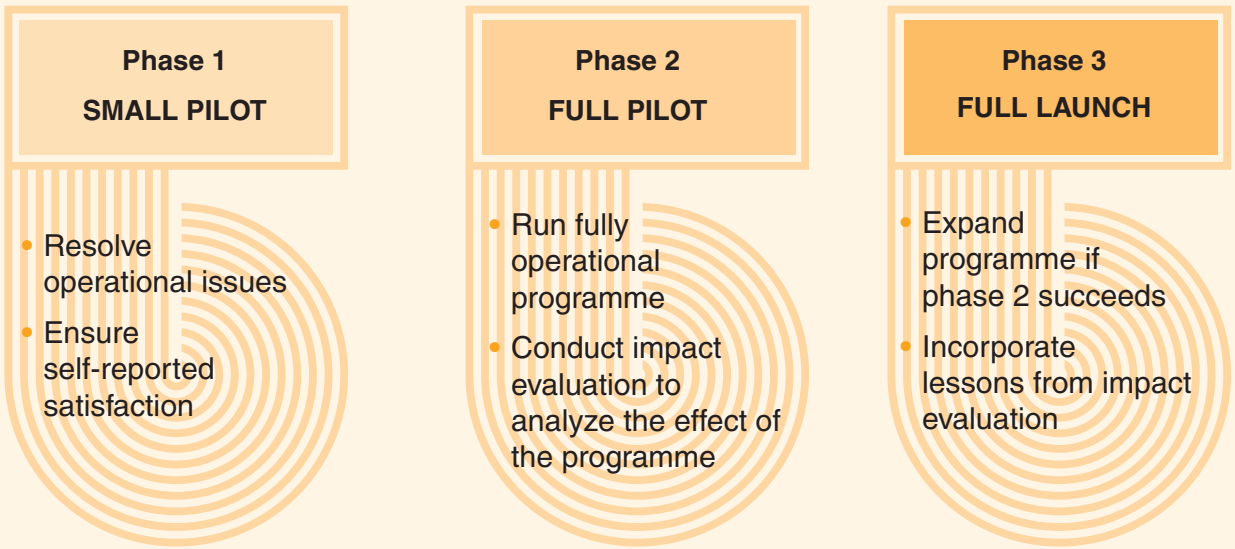
Varying information needs over the programme life cycle: A programme that has just been planned may require a cost-effectiveness analysis to help determine whether or not it should be implemented. Alternatively, for a recently launched intervention, we may need to know how closely programme procedures are being followed and whether any adjustments are necessary to guarantee successful programme operation in the future (Rubio, 2011). In many cases these information needs can be estimated even before the programme begins, as can the approximate timing of the evaluation.

Evaluation time frames differ significantly, in part dependent on the methods used for collecting and analysing data. In general, it is fair to assume that a performance evaluation can be carried out within one to six months. They can be carried out relatively quickly when relying heavily on desk research and a limited number of interviews, but the time frame will extend as soon as complex processes are being analysed and more data are collected. Impact evaluations tend to be the most time consuming of all (taking six months to two or more years), since their methodology needs to be carefully planned and new data collection is often required. Cost-benefit analysis itself can take less than a month if all the necessary data are available. However, if information first needs to be collected, the process can take much longer.

Box 4.4 illustrates the point in a programme's life cycle at which different evaluation strategies are best conducted.

Box 4.4: Life cycle of a programme and suitable evaluation strategies

FIGURE 4.4: OVERVIEW OF TYPICAL PROGRAMME PHASES



Phase 1: *The first pilot of an innovative and relatively untested youth employment intervention is about to start. What evaluation should be used?*

At the earliest stages of a programme, we usually need to make sure that everything is being done as planned. Conducting an impact evaluation at this time is not recommended because the results would not reflect the true quality of the programme. It is more appropriate to focus on monitoring until the programme is fully operational and the implementation issues common in setting up new programmes have been resolved. Qualitative data collection methods (e.g. key informant interviews, focus groups) can be particularly useful in these early stages as they may answer *why* certain elements are or are not working as intended.

Phase 2: *The intervention has been running for one year, and early operational issues have been resolved. Monitoring shows that beneficiaries are satisfied with the programme. Should we expand the programme or replicate it elsewhere?*

Now may be the time for an impact evaluation. The programme is up and running and we are confident about the quality of implementation. An impact evaluation will allow us to confirm that the programme is having an effect on the outcomes of interest. We can also use the impact evaluation to compare the effectiveness of programme design alternatives (e.g. different combinations of activities, different intensities of activities) if we are still uncertain about specific design elements. By applying a mixed method design we can understand why certain elements are working and identify any remaining bottlenecks to fine-tune our intervention. As a result of the information obtained through an impact evaluation, we can make the decision on whether or not substantial funds should be invested in the programme.

Phase 3: *The impact evaluation yielded very positive results overall. Do we still need to evaluate?*

Although positive results do not imply that the programme would work equally well in different contexts, we can now be fairly confident about the accuracy of our theory of change and the combination of activities. This is a good basis for expanding the programme to include more participants or replicating it in similar sites. Unless we want to significantly modify our intervention, another impact evaluation will probably not be necessary. However, we need to be certain that the quality of implementation remains high and that we achieve our objectives. Monitoring on all levels, including outcomes, must remain a fundamental component of our programme. Moreover, independent performance evaluations at regular intervals can help to verify the continued relevance and quality of the programme.

RESOURCES

Some otherwise desirable evaluation methods may not be feasible if we don't have the human and financial resources to carry them out. It is important to assess the skills and funding available in our programme or organization to ensure that they are in line with the requirements of the evaluation we envision.

Conducting quality evaluations calls for special skills that may not always exist in a programme or organization. In that case, and to ensure neutrality, it is often advisable to hire external evaluators.

The differences in scope and varying forms of data collection and analysis create a wide range of evaluation costs. Relying on desk research and key informant interviews is naturally much cheaper than designing and running new surveys involving a large number of people. Performance evaluations are usually the cheapest type of evaluation, with costs ranging from $10,000 to $60,000, depending on the scope of the evaluation and salary of the evaluator, as well as the data collection tools employed. In contrast, costs for impact evaluations vary widely (from around $30,000 to over $500,000) depending on the methodology used: the more data collected, the more expensive the evaluation becomes. Finally, cost-effectiveness and cost-benefit analyses can run from $10,000 to $30,000, depending on whether benefits have previously been measured (otherwise, see the costs for impact evaluations) and whether data are readily available.

THE POLITICAL CONTEXT

Different stakeholders within and outside our organization may have potentially competing interests in terms of whether or not an evaluation should take place, the issues to be studied, the type of evaluation and its methodology, the data collection strategy, and who, if anyone, should be hired for the evaluation. All of these factors can exert pressure on the choice of an evaluation and influence the relevance and quality of the planned

research. Such pressure may range from hints that certain issues should not be studied to the expression of official disapproval on the part of public authorities, effectively vetoing the interviewing of certain groups of youth, families or communities.

It is therefore important to try to understand the various interests and the political environment that exists in the specific context. The following questions will help us begin our analysis:

1. What is the local political context and the distribution of power?
2. What are the relationships among beneficiaries, programme managers, policy-makers, donors and other stakeholders?
3. What are the interests of and incentives for each group of stakeholders in terms of influencing the conduct of the evaluation and the design of the programme? For example, if the programme is narrowly targeted at one particular group of young people, those not included will have an incentive to influence the programme and evaluation in such a way that they, too, can receive benefits.
4. If the evaluation shows impact, who are the potential winners and losers from any programmatic or policy reform that could derive from the evaluation?
5. What are the conclusions and implications if the evaluation shows no impact?
6. Will the local environment allow a rigorous and independent evaluation, and will it support the evaluators in publishing their evidence-based findings, regardless of political consequences?

Working to understand stakeholder concerns through continuous and open interaction can help us to identify ways in which to address the pressures and competing interests and to build support for the evaluation. In this context, it is desirable for the programme to both ensure that the political will exists at higher levels of the hierarchy (to avoid the evaluation being thwarted) and have good relationships with key stakeholders at the grassroots level (in order to ensure ready access to data, co-operation and honesty). Moreover, it is usually helpful to bring in external evaluators who, in addition to contributing a specific skill set, may find it easier to maintain their independence. Table 4.3 presents a summary of the evaluation types.

Table 4.3: Overview of the main evaluation types

	Performance evaluation	Impact evaluation	Cost-effectiveness and cost-benefit analyses
What are the main questions answered by this type of evaluation?	• Do programmes have clear objectives? • Is the programme design appropriate to achieve the objectives? • Are adequate resources and systems (management, information, etc.) in place? • To what extent have programme objectives been achieved? • Do priorities need to be changed? • Is the programme being implemented according to design?	• How has the well-being of participants changed as a result of the intervention? • Are there any unintended consequences, positive or negative, on programme participants?	• Are programme costs justified, compared with similar interventions? • Are aggregate programme costs justified in terms of the benefits achieved?
When can this evaluation be conducted?	It may be conducted at early stages of implementation, for mid-term review or at programme completion	It should be designed during the planning of a programme but the final results will typically not be available until after the programme (phase) has been completed	It is commonly conducted during an ex ante analysis to determine whether the programme is worth implementing or continuing, or after the programme is completed to determine the final costs and their relation to the achieved benefits
How long does it take?	1–6 months (more if additional data collection is involved)	• At least 6 months (retrospective evaluation) • Approximately 12–24 months (prospective evaluation)	1–3 months
What data collection and analyses are required?	Desk review of existing documents and selected field visits and interviews with programme staff and clients. Possibly complemented by monitoring data analysis, beneficiary and stakeholder interviews, mini-surveys, focus groups, etc.	Statistical and econometric analysis of survey and administrative data, ideally combined with qualitative data analysis	Desk review of existing programme documents and relevant literature as well as key informant interviews
Who carries out the evaluation?	Usually independent evaluator (but can also be internal)	Independent evaluation team, including, for example, lead evaluator, field coordinator, survey firm	Independent evaluator (can be the same as for performance or impact evaluation)
What skills are needed?	Programme analysis, possibly qualitative and simple quantitative methods	Statistical and econometric analysis, possibly qualitative methods	Valuation and econometric analysis of programme costs and benefits
What are the costs?	$10,000–$60,000	Can range from $15,000 to $1 million or more, depending on the size and complexity of the programme	$10,000–$30,000
What programmes are best suited for this evaluation?	Any programme	Programmes that are • innovative and untested • strategically relevant and influential • replicable	• For cost-effectiveness analysis: any programme • for cost-benefit analysis: same as impact evaluation

Source: Adapted from Rubio, 2011.

KEY POINTS

1. **Learning needs are the point of departure for any evaluation**. This requires formulating evaluation questions at each level of the results chain and prioritizing the most relevant ones. In general, evaluation questions can be descriptive, normative or cause-and-effect and relate to one or several evaluation criteria, which are the relevance, effectiveness, efficiency, impact and sustainability of the intervention under consideration.
2. **The choice of evaluation strategy and the right mix of evaluation types depends on the evaluation questions**. Purely descriptive information needs may not require an evaluation and monitoring may suffice. Normative questions are most commonly answered through performance evaluations. If cause-and-effect questions are the priority, impact evaluations are needed. Cost-effectiveness and cost-benefit analyses determine whether the costs involved in an intervention are justifiable.
3. **Impact evaluations aim to determine whether and to what extent an intervention had a causal effect** (either positive or negative) on beneficiaries. Lessons learned from impact evaluation can potentially be applied beyond the intervention itself, thus allowing for broadly applicable knowledge generation.
4. **Choosing an appropriate type of evaluation depends on the operational context**. It is therefore crucial to understand whether the costs in terms of money, staff and time for each evaluation are appropriate for a given intervention.

KEY RESOURCES

- ILO. 2017. *Policy guidelines for evaluation: Principles, rationale, planning and managing for evaluations*, 3rd edn (Geneva).

- Rubio, G.M. 2011. *The design and implementation of a menu of evaluations*, PREM Notes, The Nuts and Bolts of M&E Systems, No. 6 (Washington, DC, World Bank).

- OECD DAC. 1991. *Principles for the Evaluation of Development Assistance.*

REFERENCES

- Baker, J.L. 2000. *Evaluating the impact of development projects on poverty: A handbook for practitioners* (Washington, DC, World Bank).
- Davies, R. 2013. *Planning evaluability assessments: A synthesis of the literature with recommendations*, DFID Working Paper 40 (Cambridge, DFID).
- Gertler, P.J.; Martinez, S.; Premand, P.; Rawlings, L.B.; Vermeersch, C.M.J. 2011. *Impact evaluation in practice* (Washington, DC, World Bank).
- Independent Evaluation Group (IEG). 2010. *Cost-benefit analysis in World Bank projects* (Washington DC, Independent Evaluation Group, World Bank).
- International Labour Organization (ILO). 2011. *Evaluation tool: Evaluability assessment for DWCP and projects* (Geneva, Evaluation Unit, ILO).
- ILO. 2013. *I-eval Resource Kit: Guidance Note 13: Impact evaluation* (Geneva, Evaluation Unit, ILO).
- ILO. 2017. *Policy guidelines for evaluation: Principles, rationale, planning and managing for evaluations*, 3rd edn (Geneva).
- Imas, L.G.M.; Rist, R.C. 2009. *The road to results: Designing and conducting effective development evaluations* (Washington, DC, World Bank).
- Kluve, J.; Puerto, S.; Robalino, D.; Romero, J.M.; Rother, F.; Stöterau, J.; Weidenkaff, F.; Witte, M. 2017. *Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services, and subsidized employment interventions, Campbell Systematic Reviews 2017:12.*
- OECD DAC. 1991. *Principles for the Evaluation of Development Assistance.*
- Organisation for Economic Co-operation and Development/Development Assistance Committee (OECD/DAC). 2002. *Glossary of key terms in evaluation and results based management* (Paris).
- Rubio, G.M. 2011. *The design and implementation of a menu of evaluations*, PREM Notes, The Nuts and Bolts of M&E Systems, No. 6 (Washington, DC, World Bank).
- United States Government Accountability Office (GAO). 2012. *Designing evaluations* (Washington, DC).
- White, H.; Phillips, D. 2012. *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework*, International Initiative for Impact Evaluation, Working Paper 15 (London, ALNAP).
- World Bank. 2009. *Institutionalizing impact evaluation within the framework of a monitoring and evaluation system* (Washington, DC, World Bank).

CASE STUDY

Developing terms of reference for a mid-term evaluation of a youth employment project in Egypt

This case study is based on the mid-term evaluation of the “Decent Jobs for Egypt’s Young People project” funded by Global Affairs Canada.

Learning objectives

This case study will guide the reader through several (stylized and simplified) steps of this mid-term evaluation. By the end of this case study, readers will be able to demonstrate the following learning outcomes:

- an understanding of how to formulate key parts of terms of reference for (independent) project evaluations
- the ability to develop specific evaluation questions for a youth employment intervention, given general evaluation criteria
- a clearer understanding of the scope of different evaluation and data collection methods and their links to evaluation questions.

Introduction and case study context

The greatest threat to Egypt's tenuous economic progress is its lack of decent work opportunities for young people (aged 15 to 29 years old). High unemployment levels for young women and men are only the tip of the iceberg, as many young people are forced to accept low productivity, poorly paid and insecure jobs, which are far below their capabilities.

The "Decent Jobs for Egypt's Young People" (DJEP) project aims to enable the Government of Egypt, actors at national, governorate and community level, civil society partners, the private sector and young people to create and access decent work opportunities. It focuses specifically on groups that find it particularly hard to gain access to the labour market, such as those in households headed by women, people with disabilities, economically disadvantaged people in rural areas and unemployed graduates.

Through an integrated multidimensional approach, the eight-year project (2011–2019), with an overall budget of US$12.5 million, funded by Global Affairs Canada, contributes to the development and implementation of youth employment initiatives in collaboration with national and local partners. The project engages extensively with policy-makers to strengthen youth employment policies and programmes. The project's activities are implemented at the national level and in the four governorates of El-Minya, Luxor, Port Said and Red Sea.

In line with the ILO's evaluation policy, a mid-term evaluation of the project was conducted in 2015. The evaluation served two main purposes: (1) to give an independent assessment of progress of the project across the major outcomes to date; (2) to provide strategic and operational recommendations, highlight successful interventions for scaling-up and capture lessons learned to improve performance and delivery of project results. Clients of the evaluation are ILO's management in the country office in Egypt as well as in several technical departments involved in the intervention at ILO's headquarters in Geneva. Clients also include several government agencies and ministries, as well as the project's main donor – Global Affairs Canada.

Part I: Formulating evaluation questions

A crucial step in planning a project evaluation is to formulate terms of reference (TOR) for the evaluator. The TOR clarify the purpose and scope of the evaluation and summarize the expectations of different stakeholders. TOR also formulate specific evaluation questions, tailored to the context of the project.

For the following exercise, consider the background of the DJEP project's mid-term evaluation described above and also take note of the following project components and outputs (selected and adapted for the purpose of this case study):

Component A: Strengthened capacities of the relevant ministries of the Government of Egypt to design and implement youth employment related policies and programmes.

Key outputs and activities: (a) Conduct training programmes and workshops for at least 200 government officials on how to design youth employment projects (including diagnostic analysis, project design and monitoring); (b) conduct quarterly knowledge-sharing workshops through a "Youth Employment Forum" to be attended by ministries, civil society organizations and development partners.

TIP

As part of its i-eval Resource Kit, the ILO Evaluation Office (www.ilo.org/eval) provides detailed guidance on how to formulate quality TOR for evaluators, including checklists to ensure that all relevant elements are addressed (see ILO, 2017).

Component B: Employment and entrepreneurial skills for youth and women in four selected governorates enhanced.

Key outputs and activities: (a) Offer ILO's Know About Business (KAB) course, an entrepreneurship promotion and sensitization training, at colleges to at least 50,000 students in two governorates; (b) offer ILO's "Get Ahead for Women in Enterprise" training to at least 400 young women across three governorates, working together with local non-governmental organizations (NGOs).

Discussion Topics

1. Formulate at least one evaluation question for each of the following evaluation criteria: relevance, effectiveness, efficiency, impact, sustainability (see also table 4.1 in Note 4).

Part II: Defining the evaluation's methodology

Based on the evaluation criteria and questions, the TOR for an evaluation need to specify the evaluation methods to ensure transparency and appropriate use of the evaluation budget. This includes describing the general methodological approach as well as the data collection instruments to be used. Importantly, the TOR needs to specify how the evaluation methods involve key stakeholders in the implementation of the evaluation.

Assume that the mid-term evaluation is supposed to employ a mixed-method approach and will, among other things, rely on the following data collection tools:

- key informant interviews
- focus group discussion.

Discussion Topics

Select one of the two data collection methods and:

1. define who should be interviewed (target group), at which level, (national level, governorate) and why
2. discuss for which of the evaluation criteria and questions formulated above these methods would be particularly helpful
3. discuss which additional data collection methods should be included in the TOR for this mid-term evaluation.

International Labour Organization
Youth Employment Programme
4 route des Morillons
CH-1211 Genève 22
Switzerland

youth@ilo.org
www.ilo.org/yep
www.ilo.org/taqeem

978-92-2-131662-6