

# Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Note

# 6

## A step-by-step guide to impact evaluation





Guide on Measuring Decent Jobs for Youth  
Monitoring, evaluation and learning in labour market programmes

**NOTE 6.**  
**A step-by-step guide to impact evaluation**

Copyright © International Labour Organization 2018

First published 2018

Publications of the International Labour Office enjoy copyright under Protocol 2 of the Universal Copyright Convention. Nevertheless, short excerpts from them may be reproduced without authorization, on condition that the source is indicated. For rights of reproduction or translation, application should be made to ILO Publications (Rights and Licensing), International Labour Office, CH-1211 Geneva 22, Switzerland, or by email: [rights@ilo.org](mailto:rights@ilo.org). The International Labour Office welcomes such applications.

Libraries, institutions and other users registered with a reproduction rights organization may make copies in accordance with the licences issued to them for this purpose. Visit [www.ifrro.org](http://www.ifrro.org) to find the reproduction rights organization in your country.

---

*Guide on Measuring Decent Jobs for Youth - Monitoring, evaluation and learning in labour market programmes.*  
*Note 6: A step-by-step guide to impact evaluation*/International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131666-4 (print)  
978-92-2-131667-1 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلم في برامج سوق العمل النشطة - الملاحظة 6: دليل خطوة بخطوة لتقييم الأثر, ISBN 978-92-2-630797-1 (print), 978-92-2-630798-8 (web pdf), Geneva, 2018.

---

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the International Labour Office concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the International Labour Office, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

Information on ILO publications and digital products can be found at: [www.ilo.org/publns](http://www.ilo.org/publns).

---

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

Printed in Switzerland

# Contents

<b>A step-by-step guide to impact evaluation</b>	<b>1</b>	<b>Step 7: Conduct follow-up survey and analysis</b>	<b>26</b>
<b>Step 1: Prepare for the impact evaluation</b>	<b>2</b>	Timing	26
Troubleshooting: Preparing for the evaluation	3	Tracking	27
<b>Step 2: Define timeline and budget</b>	<b>4</b>	Follow-up survey design and data collection	28
Timeline	4	Final analysis and evaluation report	29
Budget	4	Troubleshooting: Conducting a follow-up survey and analysis	33
Troubleshooting: Defining timeline and budget	5	<b>Step 8: Disseminate findings</b>	<b>35</b>
<b>Step 3: Set up an evaluation team</b>	<b>7</b>	Internal dissemination	35
Troubleshooting: Setting up an evaluation team	9	External dissemination	35
<b>Step 4: Develop an evaluation plan</b>	<b>10</b>	Troubleshooting: Disseminating findings	35
Defining the sample for the evaluation	11	<b>Key points</b>	<b>36</b>
Planning the data collection	13	<b>Key resources</b>	<b>36</b>
Troubleshooting: Developing an evaluation plan	16	<b>References</b>	<b>37</b>
<b>Step 5: Develop and pilot a survey instrument</b>	<b>17</b>	<b>Case study: Survey design and implementation for Neqdar Nesharek in Egypt</b>	<b>38</b>
Designing and testing the survey	18		
Training the fieldwork team	19	<b>Tables</b>	
Human subjects protection	19	6.1: Sample impact evaluation budget	6
Troubleshooting: Developing and piloting a survey instrument	22	6.2: Impact evaluation team and responsibilities	8
<b>Step 6: Conduct a baseline survey and analysis</b>	<b>23</b>	6.3: Overview of ethical considerations when conducting research on children and youth	21
Timing	23	<b>Figure</b>	
Supervising the data collection	23	6.1: Steps to conducting an impact evaluation	2
Analysis and report writing	24		
Troubleshooting: Conducting a baseline survey and analysis	24		





## A step-by-step guide to impact evaluation



### Prerequisites:

Notes 3 and 4 set out the elements that need to be in place and the decisions to be taken before initiating an evaluation. Note 5 is essential pre-reading to introduce readers to the different options for impact evaluation. This note is a guide to implementing an impact evaluation for youth employment interventions, covering all aspects of the evaluation process from preparation to the dissemination of results.



### Learning objectives:

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

- ▶ prepare for an impact evaluation by clarifying programme objectives
- ▶ define the timeline and budget for the evaluation, based on realistic costings
- ▶ allocate the various roles and responsibilities to members of the evaluation team
- ▶ develop the evaluation plan, including a sufficiently sized sample and data collection schedule
- ▶ pilot test the survey instrument and train the field team, taking into account good research practice and ethical considerations
- ▶ conduct baseline surveys, analyse and report on preliminary results
- ▶ conduct follow-up surveys, produce the final evaluation report and disseminate findings.



### Keywords:

Attrition, data mining, evaluation report, impact heterogeneity, population, institutional review boards, power calculations, research bias, research ethics, regression analysis, sample, sampling frame, survey team.

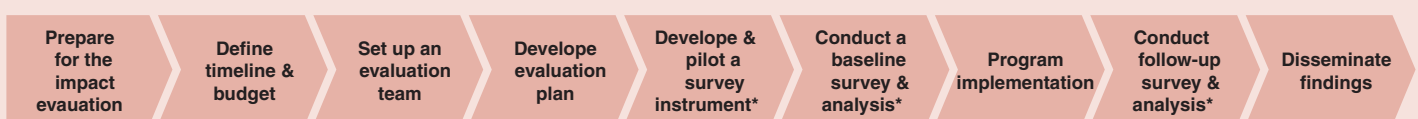
This note takes the form of a step-by-step guide. It is designed as a basic introduction to the impact evaluation process (see figure 6.1) from a programme perspective, intended for evaluation managers, commissioners and stakeholders. The information in this note is not intended to replace an impact evaluation specialist, who will always be needed to run the evaluation. Instead, the aim is to demystify what it means to carry out an impact evaluation and therefore make it easier for each organization or programme to consider undertaking an impact evaluation.

## Step 1. Prepare for the impact evaluation

Before initiating an impact evaluation, the following questions need to be asked:

- ▶ **Have I clearly defined my programme objective?** The programme objective represents what an intervention seeks to accomplish. The more concrete the objective in terms of target population, magnitude and timing of the expected changes, the easier it will be to track progress and carry out an evaluation. For instance: “By 2019, contribute to the promotion of more and better jobs for 1,000 young people between the ages of 18-29 in Serbia” (see Note 1).
- ▶ **Have I prepared a results chain?** The results chain provides stakeholders with a logical, plausible outline of how the resources and activities of the programme can lead to the desired results and fulfil the programme’s objective. Every programme should put its results chain in writing as it is the basis for monitoring as well as for defining evaluation questions (see Note 3).
- ▶ **Have I set up a monitoring system with appropriate indicators and data-collection mechanisms?** Every intervention should have a monitoring system in place before starting an impact evaluation. A monitoring system requires defined indicators and data collection techniques along all levels of the results chain in order to track implementation and results. Without effective monitoring, the results of an impact evaluation may be of limited usefulness, since it will be impossible to determine whether potentially negative

FIGURE 6.1: STEPS TO CONDUCTING AN IMPACT EVALUATION



Note: \* This step applies only to methods that require data collection by the organization.



results are due to programme design or the quality of implementation (see Notes 2 and 3).

- ▶ **Have I written down the learning objectives and evaluation questions?** Impact evaluations should be based on our information needs. Impact evaluations answer cause-and-effect questions; that is, they determine whether specific programme outcomes (usually a subset of those defined in the results chain) are the result of the intervention. Since the type of questions that we want answered may vary, we may need to think of other evaluation tools beyond impact evaluation to answer all our questions (see Note 4).
- ▶ **Have I identified an array of impact evaluation methods?** Before getting started, we should have a basic understanding of the

general mechanics of an impact evaluation and the major methodologies that can be used. Knowing the programme to be evaluated, we can identify which methodology would best suit our operational context. It is essential to have, at least, this level of understanding to inform subsequent discussions with evaluation experts and to facilitate planning (see Note 5).

In practice, misunderstandings can arise between programme managers and impact evaluation experts because the context of the evaluation has not been clearly defined at the outset. Having a clear idea about how the intervention is intended to work and what should be learned from an evaluation will make the subsequent steps more efficient, saving both time and money.

## TROUBLESHOOTING: PREPARING FOR THE EVALUATION

- ▶ **Selecting the wrong programme to evaluate:** A lot of money can be wasted on impact evaluations whose benefit and contribution are unclear. Given limited resources, it is important to target impact evaluations at strategic and untested interventions which offer the potential for replication and scaling up.
  - ▶ **Unrealistic objectives:** Many interventions suffer from “mission drift”, whereby the expressed objective of a programme changes as time goes on. It is difficult to establish useful evaluation indicators under such circumstances. Similarly, stating unrealistic objectives in terms of intended outcomes is likely to result in evaluation findings that show no impact on these outcomes. It is
- important to be realistic when defining the desired outcomes and learning objectives of the evaluation.
- ▶ **External influences:** Even after agreeing to a specific evaluation design, political factors may impede the process of moving ahead with the selected evaluation strategy. Additionally, external factors can rush or delay implementation, affecting the delivery of services and the evaluation, for example through delayed or inconsistent treatment, or the contamination of treatment and comparison groups. One possible way to reduce the influence from third parties is to firmly agree on an implementation and evaluation plan (ideally a memorandum of understanding) and to revise it periodically.

## Step 2: Define timeline and budget

### TIMELINE

By definition, the timing of an impact evaluation is highly dependent on the time frame established by the rest of the programme. It is therefore advisable to design an impact evaluation before the start of an intervention. It is also important to know when the results of the evaluation are needed. If clear deadlines for obtaining the results exist – for example, to inform decisions about programme scale-up or policy reforms – we can plan backwards from these milestones to see whether we have enough time to conduct the impact evaluation method we are considering.

Some methods require more time to implement than others. Prospective evaluations (evaluations planned in advance), such as all randomized evaluations, naturally have a longer time horizon than retrospective techniques,

such as propensity score matching. As a general rule, prospective evaluations are likely to take between 12 and 18 months, and retrospective impact evaluations will take at least six months.

In practice, the longer lead time for prospective evaluations is less problematic than it may at first appear. When new programmes are first set up, they usually take several months to become fully operational. Preparation for the impact evaluation can be carried out during the programme planning and feasibility pilot phases, allowing the evaluation to be ready by the time the programme is about to start. Even if a programme is already up and running, should the programme be organized in phases, a prospective impact evaluation can be planned for the next programme phase.

### BUDGET

Impact evaluations can be expensive, which is why many organizations are reluctant to finance them. The reality is that costs vary widely from country to country and across the methodologies and the specific programmes evaluated. Evaluations generally cost from US\$50,000 to well over US\$500,000. In some very specific circumstances, such as when all data are readily available, impact evaluations can cost as little as US\$30,000. If original data collection is needed, it is unlikely that the design and implementation of an impact evaluation will cost less than US\$50,000.

#### Cost drivers

The two major expenses in an impact evaluation are always associated with consultant and staff time and data collection (see table 6.1).

**Consultant/staff time:** The time needed to choose an appropriate evaluation methodology and design should not be discounted. Often, the monitoring and evaluation team can design the evaluation in conjunction with an evaluation consultant. The specialist's rates

will range according to experience and can be US\$200–US\$1,000 per day, for up to 20 days. More time is needed for data analysis, which can be done by the same consultant who was involved in designing the evaluation. Moreover, additional consultants may be needed to support specific elements of the evaluation, such as survey design. (Step 3: Set up an evaluation team will provide more details about the roles and responsibilities of different evaluation team members.)

**Data collection:** The main cost component for any impact evaluation is primary data collection. Hiring a survey firm is more expensive than using programme staff to collect data but normally ensures better data quality. A benchmark cost per interviewee for a baseline survey depends on the size of the

questionnaire and how easily interviewees can be found. In some cases, a short questionnaire, conducted by a survey firm with people that are easily identified with the help of the programme staff will cost \$20–\$40 per interviewee. In places where transport is difficult or where interviewees are not easily found, costs can be \$50–\$80 per interviewee. This cost includes all aspects of the survey, including hiring and training interviewers, conducting the survey and presenting the data. Follow-up surveys often present special issues with tracking participants and are likely cost to about 1.5 times more than the baseline survey. On the other hand, if tracking is not an issue, because the sample population is relatively stable and easy to find, then the follow-up survey may be less expensive than the baseline.

## TROUBLESHOOTING: DEFINING TIMELINE AND BUDGET

► **Unrealistic planning:** When developing the timeline and budget, the main risk is to underestimate the time and resources needed to carry out an impact evaluation properly. It is common to experience delays in programme design and implementation,

which, in turn, will also increase the duration – and probably the cost – of the evaluation. For example, delays can result in key staff and consultants no longer being available. Conservative budgeting and forward planning for staffing levels is essential.

### Box 6.1: Impact evaluations: Are they worth it?

For most youth employment interventions, it is probably fair to assume that the total cost of an impact evaluation will be US\$100,000–US\$500,000. This is a lot of money for many small- or medium-sized programmes and it raises the question of whether the cost is justified.

Answering this question mainly depends on (1) the time horizon of the programme, and (2) current and future funding expectations. For example, if the time horizon for even a relatively small programme with an annual budget of US\$200,000 is five years or more, or if there is potential for the programme to be scaled up to, say, US\$2 million per year, then spending US\$250,000 on an impact evaluation that informs the design of the larger programme is a practical use of money. In fact, not conducting an impact evaluation and scaling up an ineffective programme would be much more costly. On the other hand, if it is clear that the same programme will run for only two years, then the cost of an impact evaluation may be disproportionate, even though the wider youth employment community would benefit from the knowledge generated by that study. In such a case, the decision may be dependent on the availability of external funds to share the costs.

TABLE 6.1: SAMPLE IMPACT EVALUATION BUDGET

	Design stage				Baseline stage				Follow-up stage			
	Unit	Cost per unit (US\$)	No. of units	Total cost (US\$)	Unit	Cost per unit (US\$)	No. of units	Total cost (US\$)	Unit	Cost per unit (US\$)	No. of units	Total cost (US\$)
<b>A. Staff salaries</b>												
Program Manager	Weeks	2,000	2	4,000	Weeks	2,000	1	2,000	Weeks	2,000	1	2,000
M&E Officer	Weeks	1,000	3	3,000	Weeks	1,000	3	3,000	Weeks	1,000	3	3,000
<b>A. Consultant fees</b>												
Principal investigator	Days	400	10	4,000	Days	400	5	2,000	Days	400	10	4,000
Survey specialist	Days	300	5	1,500	Days	300	0	0	Days	300	5	1,500
Field coordinator/Research assistant					Days	100	80	8,000	Days	100	100	10,000
<b>C. Travel and subsistence</b>												
Staff airfare	Trips	3,000	2	6,000	Trips	3,000	2	6,000	Trips	3,000	2	6,000
Staff hotel & per diem	Days	150	5	750	Days	150	5	750	Days	150	5	750
Consultant airfare	Trips	3,000	2	6,000	Trips	3,000	2	6,000	Trips	3,000	2	6,000
Consultant hotel & per diem	Days	150	20	3,000	Days	150	20	3,000	Days	150	20	3,000
<b>D. Data Collection*</b>												
Surveying					Youth	40	2,000	80,000	Youth	60	2,000	120,000
<b>E. Dissemination</b>												
Report, printing										5,000	1	5,000
Workshop (s)										5,000	1	5,000
<b>Total cost per stage</b>				<b>28,250</b>				<b>110,750</b>				<b>166,250</b>
<b>Total evaluation cost</b>												<b>305,250</b>

Note: \* Includes training, piloting, survey material, field staff (interviewers, supervisors), transportation, etc.

Source: Adapted from [Gertler et al., 2016](#).

## Step 3: Set up an evaluation team

Impact evaluations require a range of skills, which, in turn, usually requires a big evaluation team. On the one side, there are those responsible for the programme, who will determine whether an impact evaluation is needed, formulate evaluation questions and supervise the overall evaluation effort. On the other side, there are evaluation experts, usually consultants, who are responsible for the technical aspects of the evaluation, including choosing the right methodology, planning data collection and carrying out the analysis.

The core team consists of the programme manager and the monitoring and evaluation (M&E) officer (both internal), a lead evaluation expert (often called the principal investigator or PI), a research assistant working with the principal investigator and, for evaluation designs involving new data collection, a survey expert, a field coordinator and fieldwork team (such as a data collection firm), as well as data managers and processors. Table 6.2 details the roles and responsibilities of each team member. Depending on the size of the programme and evaluation, and the skill level of the team members, multiple tasks can be assigned to one person.

After the initial evaluation design and baseline data collection, and once the programme

begins, there will be little direct work for the programme manager and the M&E officer. It is a good idea to keep one of them, perhaps the M&E officer, working on the evaluation part time during this period to ensure a connection between programme monitoring and evaluation. If any major issues relating to the implementation of the programme arise, they will need to be documented and, in some cases, reported to the wider team.

Not all outside experts need to be hired at the same time. The first priority is to select the principal investigator, who should be retained for the entirety of the evaluation, from designing the evaluation to writing the final report, to ensure continuity (although he or she will probably not be working on the evaluation during the implementation of the programme). Together with the lead evaluator, other external team members can be selected as necessary. For instance, the survey development expert is normally contracted for short tasks and may be involved in the evaluation for only a few weeks, depending on the size of the evaluation. The data collection firm is hired to conduct the baseline and **endline** surveys and is ideally the same firm for both data collections, although this is not always necessary or, indeed, feasible.

### DEFINITION

**Endline:** An endline survey is run after a programme's benefits are expected to have materialized. Comparing outcomes at baseline and endline allows changes that occurred during the intervention to be identified.

**Table 6.2: Impact evaluation team and responsibilities**

Who	Major tasks	Profile/skills required
<b>Programme manager</b>	<ul style="list-style-type: none"> <li>Define learning objectives</li> <li>Estimate resource requirements</li> <li>Prepare terms of reference for PI</li> <li>Hire evaluation consultants</li> </ul>	<ul style="list-style-type: none"> <li>Experience with designing and implementing youth employment programmes</li> <li>Experience with managing a team</li> <li>Ability to develop budgets</li> <li>Ability to work closely with programme and evaluation teams</li> </ul>
<b>Internal M&amp;E officer/unit</b>	<ul style="list-style-type: none"> <li>Define programme theory model (results chain)</li> <li>Define indicators and measurement tools</li> <li>Manage the monitoring system once the programme begins</li> </ul>	<ul style="list-style-type: none"> <li>Undergraduate or graduate degree in economics, public policy or related field</li> <li>Ability to work closely with programme and evaluation teams</li> <li>Ability to multitask monitoring and impact evaluation responsibilities</li> </ul>
<b>Principal investigator (local or international university, think tank, specialized consultancy)</b>	<ul style="list-style-type: none"> <li>Select evaluation design</li> <li>Adapt theoretically sound designs to real-world budget, time, data and political constraints</li> <li>Develop mixed-method approaches</li> <li>Identify evaluation team and prepare terms of reference</li> <li>Supervise staff</li> <li>Determine sampling and power requirements</li> <li>Analyse data and write report</li> </ul>	<ul style="list-style-type: none"> <li>Graduate degree in economics, public policy or related field</li> <li>Knowledge of the programme or similar types of programmes</li> <li>Experience in research methods and econometric analysis</li> <li>Some experience in the country or region</li> <li>Demonstrated ability to work effectively in multidisciplinary teams</li> <li>High-level written and oral communications skills</li> </ul>
<b>Survey expert (may be the same person as the PI)</b>	<ul style="list-style-type: none"> <li>Design survey instrument</li> <li>Prepare accompanying manuals and codebooks</li> <li>Train the data collection firm</li> <li>Support piloting and revision of questionnaires</li> </ul>	<ul style="list-style-type: none"> <li>Graduate degree in economics, public policy or related field</li> <li>Experience in surveying children and youth</li> <li>Experience in carrying out fieldwork in the country or region of interest</li> <li>Ability to interact effectively with research and programme counterparts</li> </ul>
<b>Field coordinator and fieldwork team</b>	<ul style="list-style-type: none"> <li>Assist in the development of the questionnaire</li> <li>Hire and train interviewers</li> <li>Form and schedule fieldwork teams</li> <li>Oversee data collection</li> <li>Clean the data so it can be shared with the evaluation specialist</li> </ul>	<ul style="list-style-type: none"> <li>Legal status, business licences recognized by the Government of the country where work is to be performed</li> <li>Good network of experienced interviewers, supervisors and data-entry clerks</li> <li>Demonstrated 5+ years' experience with organizing surveys on the scale of this programme</li> <li>Strong capacity and experience in planning and organizing survey logistics</li> <li>Strong capacity in data management and statistics</li> <li>Ability to travel and work in difficult conditions</li> </ul>
<b>Research assistant</b>	<ul style="list-style-type: none"> <li>Analyse data</li> <li>Support the PI in writing the evaluation reports</li> </ul>	<ul style="list-style-type: none"> <li>Undergraduate or graduate degree in economics, public policy or related field</li> </ul>
<b>Data managers and processors</b>	<ul style="list-style-type: none"> <li>Clean the data so the research assistant and PI can use it</li> <li>Manage data team</li> </ul>	<ul style="list-style-type: none"> <li>Experience with data software and management of data team</li> </ul>

## TROUBLESHOOTING: SETTING UP AN EVALUATION TEAM

- ▶ **Recruitment:** Recruiting a good impact evaluation team, from writing the terms of reference to identifying qualified experts and firms, can be a challenge. Underestimating the expertise needed at different stages and hiring the wrong people can lead to significant delays and cost overruns, and ultimately impair the results of the evaluation. It is necessary to ensure that the requirements for each role are clearly defined from the outset and fulfilled by the respective expert or firm. Working with established institutions (such as universities and think tanks) that have a proven track record in conducting quality research studies can help to build local support and ensure that the final results are widely accepted.
- ▶ **Changing staff:** Firms that win evaluation contracts sometimes replace key staff with less experienced personnel. This can be prevented through clear contractual clauses with penalties or remedial actions in case of default.
- ▶ **Survey team management:** Managing an internal survey team can become complicated very quickly. When organizing data collection with programme staff, ensure

### TIP



Partnering with academic institutions is often a powerful strategy for NGOs and governments to develop their impact evaluation capacities. For example:

- Save the Children is partnering with Universidad de los Andes in Colombia to evaluate the YouthSave initiative
- the ILO, through the Taqueem Initiative, partnered with the Population Council Egypt and with researchers from the American University of Cairo and the Institute of Labor Economics (IZA) in Bonn, Germany, to evaluate a business and vocational training programme for young women in rural Egypt
- the Turkish Ministry of Labour is partnering with the Middle East Technical University on the evaluation of the Turkish Public Employment Agency (ISKUR).

that you have a clear understanding of the full staff needs and that the oversight and management structures in place are suitable for directing the team.

## Step 4: Develop an evaluation plan

Once the principal investigator is on board, he or she will usually prepare an impact evaluation plan (also called a concept note) in coordination with programme leaders. That plan will describe the objectives, design, sampling and data collection strategies for the evaluation. In essence, the impact evaluation plan (see the sample outline in box 6.2) will provide the basis for choosing the impact evaluation methodology and will guide all subsequent steps in the process of implementing the evaluation.

Developing the evaluation design (point 4) should not be carried out by the evaluation

expert in isolation; instead, the process should closely involve the programme staff to make sure that the evaluation method fits the learning objectives and operational context of the programme. In addition, although the principal investigator will certainly approach the programme staff and make suggestions for defining the sample for the evaluation (point 5) and planning the data collection (point 6), it is still useful for the implementing organization to have a basic understanding of how these aspects are relevant to the overall evaluation and the programme itself. Therefore, we explore these two points in more detail in the following sections.

### Box 6.2: Outline of an impact evaluation plan

- 1 Introduction
- 2 Background
- 3 The intervention
- 4 The evaluation design
  - 4.1 Objective of the evaluation
  - 4.2 Hypotheses and research questions
  - 4.3 Evaluation methodology
- 5 Sampling strategy and power
- 6 Data collection plan
- 7 Data analysis plan
  - 7.1 Measuring impacts
  - 7.2 Examining differential treatment effects
  - 7.3 Measuring the return to the programme (cost-benefit analysis)
- 8 Risks and proposed mitigation
- 9 Audience and dissemination
- 10 Timeline and activities
- 11 Budget
- 12 Annexes



## DEFINING THE SAMPLE FOR THE EVALUATION

We do not need to assess every programme participant in order to evaluate an intervention. We just need to choose a group of people – a **sample** – that is big enough for the purposes of our evaluation. If our sample is representative of all eligible youth, we can generalize the results of the evaluation to cover the total eligible population. To obtain a representative sample, we need a sampling strategy.

We also want the sample to be large enough to be able to generate a reliable comparison of outcomes between those in the treatment group and those in the comparison group. If the sample is too small, we may not be able to

see a statistically significant impact of the programme, even if there is one. To know how big is big enough, we need **power** calculations. These concepts are discussed below.

### Creating a sampling strategy

A sampling strategy involves the following three steps:

1. **Determine the population of interest:**  
First, we need to have a very clear idea about who we want to target and who will

#### DEFINITION

**Power** is the probability of detecting an impact, if one has occurred. There is always a risk that we will not detect an impact with confidence even if it does exist. However, if the risk of not detecting an existing impact is very low, we say that the study is sufficiently powered.

#### DEFINITION

**Sample:** A sample is a subset of a population. Since it is usually impossible or impractical to collect information on the entire population, we can instead collect information on a subset of manageable size. If the subset is well chosen, then it is possible to make inferences or extrapolations that apply to the entire population.

### Box 6.3: The importance of close collaboration between programme staff and evaluators

The example of a planned impact evaluation of youth microfinance in Yemen shows the importance of programme staff and evaluators collaborating closely from the beginning of a programme to develop a mutual understanding of the operational context. In this case, evaluators independently designed a randomized control trial to assess the impact of lending and other financial services for youth on employment creation, business expansion and other outcomes. When the evaluation design was presented, the CEO of the bank involved made it very clear that such a design would be unacceptable in the context of a recently founded financial institution that could not afford to exclude potential clients for the purposes of an evaluation. The evaluation team then had to start again and finally chose a randomized promotion evaluation design that was more suitable for an intervention with universal coverage.

be eligible for the programme. For example, age, gender, income level, employment status and location could determine eligibility. Those who are not eligible will not be included in the study.

2. **Identify a sampling frame:** A **sampling frame** is the most comprehensive list of units in the population of interest that we can possibly obtain. It tells us how our sample relates to the general population of interest, for which we want to extract the lessons of the evaluation. Ideally, then, the sampling frame corresponds exactly to the population of interest, indicating that it would be fully representative. We would try to get a list of eligible youth from a population census, school or voter registration, or city registry that includes as many of the eligible youth as possible. In reality, however, it is not always easy to obtain a sampling frame that would fully cover the eligible population. In practice, studies often rely on sampling households and choosing youth in those households.
3. **Draw the desired number of units from the sampling frame using one of the available sampling methods:** Various methods can be used to draw samples from our frame, but the most commonly used are some form of probability sampling. With this method, participants are selected into the sample with a specific probability. In the case of random sampling, for instance, every participant in the sampling frame would have the same probability of being included.

**TIP**

Sometimes it proves impossible to obtain a sampling frame. In that case, there are other possible ways of obtaining a sample. A popular technique is snowball sampling, where a small set of initial research participants recruit other participants into the study (usually family, friends or colleagues). Over time, as the ball gets rolling it collects more “snow” (more respondents) and the sample becomes larger and larger.

In the absence of a comprehensive list and if we don’t know how our study population represents the general population of interest, we should be careful when generalizing lessons learned to the study population. It is tempting to draw general lessons beyond the sample, and many studies do, but we must be modest and careful when interpreting the results. Similar caution about generalizing conclusions is needed when a programme is scaled up, since a larger programme may reach youth who differ significantly from those who took part in the original study.

### “How large does my sample need to be?”

It is crucial to know the ideal size of our sample, that is, how many individuals we should draw from the sample frame. There is an important trade-off involved in choosing sample sizes. On the one hand, more observations allow for more precise estimates of effects: if the sample is very small, there will be lots of uncertainty surrounding our estimates of the impacts of interest. It might become very hard, or impossible, to find out whether the intervention had any effects on beneficiaries at all. On the other hand, sample size is an important cost

**DEFINITION**

**Sampling frame:** The most comprehensive list of units in the population of interest that we can possibly obtain. Drawing from this list allows us to obtain the (representative) sample.

driver. However, the relative added value of one additional observation declines as the sample grows. Thus, if we already have a big sample, adding some more observations will add to our costs without being much help in terms of precision. Power calculations help us to find the right size of the sample by indicating the smallest sample with which it is still possible to measure the impact of our programme with a reasonable level of confidence.

Although appropriate sample sizes for evaluations vary, in general we should aim to include 1,000–3,000 youth in our evaluation to ensure that both the treatment and comparison groups are sufficiently large. In some very specific cases, a sample size of fewer than 1,000 youth may be acceptable in some cases. It is almost never advisable to have fewer than 500 respondents (250 in the treatment group and 250 for comparison).

## PLANNING THE DATA COLLECTION

The evaluation plan will need to establish the basic data collection strategy. Data collection can be a very complicated task, which is best handled by a team of outside experts. Key issues include the timing of data collection, whether new data must be collected, who is going to collect the data and how the data will be managed. These issues are discussed below.

### Timing of data collection

The timing of data collection is very important and depends on the nature of the programme. When a baseline survey will be used, it should be completed before the programme starts and *before participants know if they are going to be enrolled in the programme* to ensure that their answers are consistent across the treatment and comparison groups. This is critical, as youth may give different answers if they know that they will be receiving the programme.

#### TIP



Evaluation professionals will be able to calculate the appropriate sample size for your particular evaluation. The Abdul Latif Jameel Poverty Action Lab (J-PAL) provides a range of online resources to help conduct power calculations (see <https://www.povertyactionlab.org/research-resources/software-and-tools>).

#### TIP



Since some young people may drop out of the programme during implementation, and hence drop out of the evaluation, it is wise to choose a sample size that is bigger than the minimum sample indicated by the power calculation.

The timing of the follow-up survey should take into account the programme needs and programme effects. If a follow-up survey is conducted too soon, no effect will be found; if too late, the programme may not benefit from the knowledge gained.

### Existing versus new data

It is not always necessary to collect new data. In some cases, the data required for an evaluation already exist (box 6.4 offers suggestions for where to find it). Two types of data commonly exist and should be explored before deciding to collect new data.

First, depending on the questions that the evaluation wants to answer, **the necessary data may already have been collected in the form of monitoring data**. For example, many employment programmes already ask for information on employment status at the start of the programme, thus reducing the need for a baseline. However, this information is normally only collected for those in the programme. For the purposes of an impact

## Box 6.4: Potential sources of data

**Monitoring data:** Administrative data are usually collected by an implementing programme for monitoring and reporting purposes.

**Household survey data:** National household surveys are periodically conducted in many developing countries. These include multi-topic surveys, such as the Living Standards Measurement Survey and the Demographic and Health Survey, which can cover a wide range of information on housing characteristics, household consumption and wealth, individual employment, education, and health indicators. Other surveys, such as labour force surveys, are more restricted in scope and sometimes cover only urban areas.

Where to look:

- statistical institutes in the respective country
- International Household Survey Network (<http://www.ihsn.org>)
- demographic and health surveys (<http://www.measuredhs.com>)
- Living Standards Measurement Study (<http://iresearch.worldbank.org/lsms/lsmssurveyfinder.htm>).

**Census data:** Most countries conduct a population and housing census every ten years, and many conduct additional surveys. The advantage of census data is that they cover the entire population, so there are data for virtually every potential treatment and comparison observation. The drawback of census data is that they are infrequently collected and typically contain only a limited number of indicators, which limits their value for an impact evaluation.

Where to look:

- ILO Microdata Repository (<http://www.ilo.org/surveydata>)
- International Household Survey Network (<http://www.ihsn.org>)
- IPUMS International (<https://international.ipums.org>).

**Facility survey data.** Facility surveys collect data at the level of service provision, such as at a school or vocational training centre. National ministries, state entities, or even local authorities may compile the information. In many cases, facility-level surveys will provide control variables (such as teacher–student ratio), while others may capture outcomes of interest, such as attendance rates.

Where to look: Relevant national ministries and local representatives.

**Specialized survey data.** A specialized survey is one that is collected for a specific purpose, often for research on a particular topic. Many take modules from the existing national household survey and add questions on topics of interest. Coverage of specialized surveys can be quite limited, sometimes resulting in little or no overlap with programme areas. Nevertheless, if the evaluation team can find existing data from a specialized survey on a topic related to the evaluation, these datasets can provide a rich collection of relevant indicators.

Where to look:

- local officials, donors and NGOs in the area of interest
- ILO school-to-work transition surveys (<http://www.ilo.org/w4y>)
- World Bank STEP Skills Measurement Program (<http://microdata.worldbank.org/index.php/catalog/step/about>).

Source: World Bank, 2007, pp. 8–11.

evaluation, data must also be collected on individuals in the comparison group. To avoid inadvertently introducing biases through inconsistent data collection, it is important that any system designed for data collection is as consistent and objective as possible for both the treatment and comparison groups. This is often difficult to achieve through monitoring data. Unless such a system is already an integral part of the programme, it is best to use a dedicated team to collect new data on both the treatment and comparison groups.

Second, the local statistics office may already have collected data on many of the programme participants and comparison groups. For smaller programmes, it is unlikely that enough people in the programme would have been part of an existing survey. For larger programmes, however, it is likely at least some would have been. It is also important to understand what data was collected and how that collection was carried out. Ensure that the questions asked pertain to the programme that we have in mind and that the sample size was large enough to warrant drawing conclusions. Check with the local statistics office to confirm that the data exist and can be used.

If insufficient existing information can be obtained, new data will have to be collected.

### Internal versus external data collection team

The collection of data is the most expensive part of an evaluation for good reason. The collection of high-quality data that can be easily analysed is key to a successful evaluation. Without high-quality data, all of the work put into designing the evaluation may go to waste. When deciding between hiring a survey firm or using internal staff to collect data, the programme must choose the method that both fits its budget and ensures quality and systematic data collection. Some programmes

#### TIP



In some cases, programmes attempt to use partner implementing organizations to collect data through their programme staff. It is not advisable to have people who are dependent on funding conducting the data collection because there is a greater chance that the results will be biased in favour of the programme. If it is decided that data collection will be handled internally, it is best to employ a separate team that is focused solely on data collection and is not associated with the programme.

prefer to conduct data collection on their own since it can save money. This may work well for short, simple surveys, but it has some important drawbacks, especially for extensive data collection processes. Due to the complexity of collecting data and ensuring the proper logistics, it is normally not advisable to collect use programme staff to collect data. While hiring a survey firm is typically more expensive than handling the data collection internally, it does mean the data can be collected more quickly and with less input from the programme office. It also ensures that the team doing the data collection is well qualified. (Additional guidance on quality assurance is included under the sections Training the fieldwork team and Supervising the data collection, in Steps 5 and 6, respectively). Moreover, hiring an outside firm helps to establish neutrality and increases the credibility of the evaluation results.

### Data collection process and techniques

Generally, surveys should be administered by trained personnel; self-administered questionnaires should be used only in certain circumstances. When individuals fill out surveys on their own, they often interpret questions differently from what was intended by

the survey team. Trained interviewers ensure greater consistency of interpretation. Also, in many contexts, participants are not as literate as we might expect or hope, so they may require guided interviews.

There are several ways to collect and record survey responses. Paper surveys are traditional. If available, interviewers can also use mobile phones (to which surveying software can be downloaded), computers or personal digital assistants. It may also be possible to

tape interviewee responses. Although technology-based tools may require some initial training (usually relatively minor), they can reduce the time needed for each interview, cut the time required for data entry and minimize data errors that arise from traditional data entry and processing. They can therefore save time and money, especially in larger surveys. However, one also needs to consider the appropriateness of using expensive equipment in poor households and neighbourhoods.

## TROUBLESHOOTING: DEVELOPING AN EVALUATION PLAN

- ▶ **Limitations of existing data:** When working with secondary data, it is important to ensure its availability and quality. Existing surveys may not ask the questions relevant to our particular evaluation or address our population of interest, or they may have a sample size which is too small to adequately power our study. Before committing to using only existing data, it is important to fully understand its limitations.
- ▶ **Disconnect between programme and evaluation:** Insufficient communication and coordination between the implementing organization and the lead evaluator can result in choosing an evaluation design that

will not be feasible in practice. Keeping key programme staff involved in the evaluation planning can help to ensure that the evaluation suits the operational context. If a disconnect does arise and it is caught in time, the best solution is to find a more realistic evaluation method.

- ▶ **Selection bias:** Carefully identifying the sample and randomizing study participants is the simplest and most robust way to eliminate selection bias. If selection bias is present in the data, one imperfect solution is to compare the outcomes among the treated group to a matched sample drawn from a different data set.



## Step 5: Develop and pilot a survey instrument

If the evaluation plan calls for new data collection, it is important to choose the right data collection tool. In most cases, some sort of survey will be used, often in combination with other qualitative methods, such as focus groups or key informant interviews.

Because the survey will be the basis for collecting data about participants and the

comparison group, the survey design is crucial. Although designing questionnaires may seem trivial, coming up with a high-quality survey that yields reliable results is a science and an art. Surveying adolescents and youth poses additional challenges compared with surveying adults, so it may be wise to seek support from an expert consultant for this step (see box 6.5).

### Box 6.5: Factors affecting data reliability when surveying youth

Any evaluation depends on reliable information. While research indicates that young people are generally reliable respondents, there are a number of reasons why youth may be more likely than adults to misreport or even falsify answer questions:

- **Comprehension:** Young people may have less education and relatively limited cognitive ability. Does the respondent understand the question? Is the question asked using age-appropriate language? Some questions are subtle and may be difficult for youth to understand even when asked in a simple and straightforward manner.
- **Recall:** How likely is it that the respondent remembers the events or information? This has partly to do with the reference period: how long ago the event occurred or how frequently the event occurs. In general, shorter recall periods are more accurate than longer ones.
- **Confidentiality:** Does the respondent have any reason to fear reprisal or other consequences arising from the answers he or she gives? Is the interview really being conducted in private? The interviewer must be able to convince the respondent that the information is confidential.
- **Social desirability:** Does the respondent believe that the interviewer is expecting one specific response or another? Can one answer be perceived as “correct”? This factor is particularly pertinent to behaviours that are illegal, stigmatized or subject to moral strictures. [Brener et al. \(2003\)](#) report studies showing that adolescents are more likely to report recent alcohol consumption in self-administered questionnaires than in interviews, whereas there is no difference in the responses of adults.
- **Exhaustion:** Although surveys among adults can take many hours to complete, young people are more likely to lose patience with long interviews. For example, the NGO Save the Children created the Youth Livelihoods Development Index, which comprises three self-administered surveys for young people aged 11–24 to elicit information about assets and competencies. The pilot test found that youth “got bored with the long questionnaire and fabricated answers” ([Bertrand et al., 2009, p. 5](#)).

## DESIGNING AND TESTING THE SURVEY

Before the survey can begin in the field, the questionnaire must be developed. This is done through an iterative process, which usually takes one to two months.

### Step 1: Design

The questionnaire is based on the outcomes and indicators previously developed. Local language, dialects and youth slang are important aspects to incorporate, and a translator may be needed to do this effectively. If sensitive topics are included in the questionnaire, such as questions about mental health or violence, questions must be formulated thoughtfully and in line with local norms and customs. The first draft will usually contain questions that will eventually be cut or changed.

### Step 2: Internal review

Once a questionnaire has been drafted, other team members and stakeholders, such as the programme manager, M&E officer, principal investigator and fieldwork team, should review it to confirm that the questionnaire collects all the information needed.

### Step 3: Piloting

The draft questionnaire is then taken into the field. The importance of this step is often overlooked, but it is critical for the production of a quality evaluation. Field-testing is crucial to confirm that the survey's length, format and phrasing are all appropriate, and to make sure that the survey can yield consistent and

#### TIP



Good practices for surveying youth include the following:

- obtain informed consent from both the young person and the parent (see the section below on human subjects protection)
- use familiar local language or slang, if appropriate
- be mindful of the young person's attention span; keep surveys short and interesting
- use probing questions to improve the quality of responses; refer to the recent past to help with memory and recall
- as with all respondents, be cautious about the timing and phrasing of sensitive questions
- to help with finding youth again later, gather a lot of information on family, friends and neighbourhood contacts
- if information about the household is needed, include a separate survey module targeted at parents or guardians.

reliable results. The questionnaire should be tested on a selection of individuals who are similar to those who will be part of the programme, but who will not be in the final sample. This will ensure that those people who receive the final questionnaire are not influenced by having already been exposed to the questions. It is also important to pre-test the procedures that will be used for locating interviewees, to ensure that they can easily be found.

### Step 4: Revision

The draft questionnaire is revised to address the issues raised in the field. If necessary, the steps can be repeated until all issues have been resolved.



## TRAINING THE FIELDWORK TEAM

When the questionnaire is ready, the fieldwork team must be trained to administer it. The survey expert or data collection firm should develop a manual to be used as a training tool and reference guide for interviewers. As a minimum, the manual should discuss the survey objectives and procedures, including procedures for dealing with difficulties in the field. Each survey question should be explained, so that interviewers understand the rationale for the question's inclusion in the survey. In addition, the manual should provide interviewers with specific instructions on how to ask each question and obtain usable information. The principal investigator and programme manager should review the manual. Box 6.6 presents a sample outline of a survey manual.

Training interviewers can take a few days or over a week, depending on the complexity of the survey. Training should begin by going through the entire survey, question by question. Then, each interviewer should practise

on another interviewer. Interviewers should be encouraged to ask questions during this process to ensure that everyone understands each of the questions. This process should continue until all interviewers are completely familiar with all the questions. After the training is complete, interviewers should be taken to a site where they can practise the questionnaire on at least five people who resemble the sample respondents.

Interviewer training is both a training process and a job interview. Invite at least 20 per cent more interviewers to the training than you expect to need, and retain, and accept only the best.

If a survey firm is contracted, they will be in charge of the training process. It is often a good idea to have someone from the programme attend the first few days of the training to answer any questions that arise. This is the last chance to eliminate errors in the questionnaire.

## HUMAN SUBJECTS PROTECTION

Research that involves human beings can sometimes create a dilemma. When our research is intended to generate new

knowledge for the benefit of a specific programme or an entire field, for example by measuring the impact of a youth employment

### Box 6.6: Sample outline of a survey manual

1. Objectives of the survey
2. Duties, roles and expectations of interviewers, supervisors and other survey personnel
3. Procedures for checking data accuracy
4. Detailed survey and interview procedures (including procedures for identifying, locating and contacting respondents, as well as guidance on surveyor conduct, confidentiality, objectivity, interview pace, bias and probing)
5. General instructions for filling out the questionnaire and coding
6. Simple explanations of each question
7. Instructions for finishing and checking the survey and thanking respondents
8. Instructions for filling out the field report and notifying supervisors of any difficulties encountered

intervention, we may be inclined to consider the outcomes of our evaluations as more important than protecting individual research participants. Clearly, we should not use young people solely as means to an end, and there are procedures in place to help us assess our evaluation's ability to protect participants. See table 6.3 for an overview of the ethical considerations to bear in mind when conducting surveys involving young people.

Basically, three main principles protect the interests of research participants (NIH, 2008, pp. 17–20):

- ▶ **Respect for persons:** This principle refers to making sure that potential participants comprehend the potential risks and benefits of participating in the evaluation. In practice, this means that a process must be in place to ensure **informed consent**, the explicit willingness of young research participants to answer the survey questions in light of their clear understanding of the nature of the survey.
- ▶ **Beneficence:** This principle refers to doing no harm and maximizing the possible benefits of the research.
- ▶ **Justice:** This principle requires individuals and groups to be treated fairly and equitably in terms of bearing the burdens and receiving the benefits of research.

In order to ensure the highest ethical standards in an evaluation, many researchers will

#### TIP



Be mindful of cultural norms and local customs when recruiting and assigning interviewers. For example, it is usually a good idea to use female enumerators to interview female respondents, particularly when sensitive questions are being asked. If respondents (or their guardians) do not feel comfortable with an enumerator, it is more likely that they will not participate in the survey, or, if they do, that the information provided will be incomplete, inaccurate and, therefore, unreliable.

be required to submit their impact evaluation plan for review by an **institutional review board** (IRB) in the donor country, the host country or both. These reviews are mandated by law for anyone engaging in research supported by the US Government and many other governments, as well as most universities throughout the world. Even if they are not legally required, conducting ethics reviews is a good idea for anyone working with human participants. Ideally, the IRB would review the survey before it is piloted, but certainly before the final survey is implemented at large. IRBs can be found in any US-based university (the best option when working with a US-based researcher) or through a local ethics review board. Other institutions, such as the US National Institutes of Health or Innovations for Poverty Action, also conduct ethics reviews on request.

#### DEFINITION

An **institutional review board**, also known as an independent ethics committee, is a committee that has been formally designated to approve, monitor and review research involving human participants with the aim of protecting the rights and well-being of these individuals.

#### DEFINITION

**Informed consent** refers to the explicit willingness, preferably expressed in writing, of a person (and, when necessary, his or her parent or guardian) to participate in the research. Informed consent requires full information about all features of the research that may affect a young person's willingness to participate.

**Table 6.3: Overview of ethical considerations when conducting research on children and youth**

Issues	Why it matters	What to do
Information about risks and benefits of participation	Young people and adults may have different abilities to accurately assess the benefits and risks associated with participating in a particular programme or research initiative. Young people may also be more risk-taking in general, making them more vulnerable to the potential negative consequences of participation	<ul style="list-style-type: none"> <li>• Anticipate possible consequences for the children and youth involved. Do not proceed unless potentially harmful consequences can be prevented or mitigated</li> <li>• Provide young participants with an explanation of the proposed research objective and procedures in a language and format appropriate to their age, maturity, experience and condition</li> <li>• Provide explicit information on any inconveniences or risks the young person may experience if she or he agrees to take part in the programme or evaluation</li> <li>• State clearly that there is no obligation to participate in the study and that the decision to participate in the study will have no effect on eligibility for the programme</li> <li>• Do not raise unrealistic expectations about the benefits or rewards of participation</li> <li>• If any, provide only modest rewards or incentives to participate that are in line with local living standards</li> </ul>
Consent	Young people may not have reached the age of legal maturity; their parents or guardians need to be asked for consent prior to engaging the youth themselves. Moreover, obtaining young people's truthful opinions can be difficult because they are often exposed to social pressure to comply with adult opinions, regardless of whether or not they agree	<ul style="list-style-type: none"> <li>• Determine the age of majority in the country and consult locally to determine who must give permission to work with the young people (parents, teachers, local authorities, community leaders, etc.)</li> <li>• When working with minors, always seek informed consent from parents or guardians,</li> <li>• If the age, maturity and situation of the young participants allow, also obtain informed consent from the youth, in addition to that of their parents</li> </ul>
Data collection	The collection of information on sensitive topics (e.g. drug use, sexual activity, involvement in crime) or distressing experiences (abuse, loss of parents, deprivation) is more delicate when dealing with children and youth compared to adults. Their emotional and physical vulnerabilities have to be protected	<ul style="list-style-type: none"> <li>• Prior to interviewing young people, try to collect as much information as possible from alternative indirect sources (adults, administrative records, etc.)</li> <li>• Consult locally and design questionnaires, focus group guidelines and other materials according to the characteristics of the specific target group (e.g. make sure that survey instruments are age-appropriate and comprehensible)</li> <li>• When necessary, acknowledge that questions can be sensitive, and anticipate and address the concerns of parents and participants</li> <li>• State clearly that the young participant can refuse to answer any or all questions, and that this will have no effect on eligibility for the programme. Such disclaimers should be repeated before asking sensitive questions</li> </ul>
Confidentiality and protection	Protection of privacy is always crucial, and even more so when dealing with young respondents and sensitive topics. Given the involvement of parents or other guardians during the consent process and as legal representatives, there may be trade-offs between confidentiality and the ethical obligation to protect the safety of the respondents. For example, the presence of parents in the interview may undermine the privacy of the youth. At the same time, there may be a responsibility to inform guardians if the young person is at risk of harm	<ul style="list-style-type: none"> <li>• Always ensure the privacy and confidentiality of responses from parents and young participants, which will also strengthen the reliability of the information provided</li> <li>• Never release information about the respondent without the express approval of the respondent and his or her parent</li> <li>• Plan how to intervene if the respondent provides information suggesting they or others may be at risk of harm (from domestic abuse, neglect, crime and violence), or may require medical, legal or other services</li> <li>• At the beginning of each interview, and regardless of the apparent conditions of the respondent, inform <i>all</i> participants of the resources available for referral</li> </ul>

## TROUBLESHOOTING: DEVELOPING AND PILOTING A SURVEY INSTRUMENT

- ▶ **Measuremania:** Targeting too many outcomes, and thus including too many questions in the survey instrument, often extends the cost of the survey beyond the survey budget. Too many questions increase the burden on survey participants and may reduce response rate and the quality of responses. Cutting questions that related to indirect outcomes is a good way of limiting this issue.
- ▶ **Insufficient testing:** The step that is often skipped in the interest of saving time is piloting the evaluation tools. However, piloting is a critical step in the process that cannot be eliminated, especially because surveying youth poses additional challenges that may not be immediately understood. If the tool isn't validated, the results could be inaccurate, incomplete or misleading. Take the time necessary during the field testing phase of a survey to ensure that the information collected is of the highest quality.
- ▶ **Discounting ethics:** Administering a survey that hasn't been approved by an IRB or local ethics committee may lead to massive push-back from stakeholders and may disqualify the entire evaluation. Basic ethics training for all parties involved in the evaluation is a minimum requirement.

### Box 6.7: Managing the research protocol approval process

To conduct a survey for the job training programme “Mi Primer Empleo”, targeting urban youth in Honduras, the World Bank contracted the National Opinion Research Center (NORC) at the University of Chicago to design the questionnaire and manage the data collection process. Even though Honduras does not have any statutory requirements for dealing with sensitive survey data involving human participants, the terms of reference for the evaluation required US IRB approval for the research design and data collection plan, as well as data security procedures that meet international standards. NORC therefore submitted all research protocols and questionnaires to its university IRB for approval prior to beginning fieldwork.

Given the nature of the research, field interviewers and supervisors were screened to determine their experience with youth-related surveys. During the programme registration process, applicants were informed that they would be asked to participate in a voluntary survey but that their decision to participate in the survey would in no way influence their selection for the training programme. Given that the legal age of consent in Honduras is 18 years old, the data collection team sought written consent from respondents aged 17 or younger, and oral or written consent from the minor's parent or guardian for programme registration, as well as a separate consent from both the minor and the guardian to participate in the evaluation survey.

To ensure confidentiality, personal information was strictly separated from interview forms, and the latter contained only a numeric identifier. Thus, personal registration information (names, address, etc.) was available exclusively to the implementing organization (Ministry of Labour and Social Security) for the purpose of contacting youth who had registered, while survey response data (without personal information) was delivered only to the World Bank for analysis.

Source: [National Opinion Research Center \(NORC\), 2007](#).

## Step 6: Conduct a baseline survey and analysis

The baseline survey is the first set of data collected on the treatment and comparison groups. Collecting baseline data provides an early indication of whether the chosen impact

evaluation design is valid in practice, while also gathering useful information about beneficiary characteristics that can inform the programme.

### TIMING

Baseline data should be collected shortly before the programme begins. Conducting the survey after programme initiation runs the risk that the programme might already have influenced the characteristics to be measured. Equally, conducting the baseline survey too far in advance of the programme could result in the information collected failing to accurately reflect the situation of participants at the beginning of the intervention.

If a prospective evaluation is being conducted, individuals will need to be assigned to treatment and comparison group before the programme begins. However, that assignment decision should not be communicated to the survey participants until after the baseline data have been collected.

### SUPERVISING THE DATA COLLECTION

Quality assurance is key to ensuring that the data collected is of the highest quality. First, it is important to conduct validity testing to ensure that interviewers are meeting the requisite standards of their job and achieving the target number of surveys per day. It is customary to establish an independent team to audit 10–15 per cent of the surveys to verify that respondents actually exist and that data was collected accurately. Incentives may help to ensure that interviewers keep a positive attitude in a difficult job. In addition to wages, interviewers often receive a per diem allowance to cover food and housing while travelling, as well as other incentives.

Second, steps should be taken to protect the data collected. Information can be lost if completed questionnaires are misplaced or computers are stolen or malfunction. To avoid the loss of data, surveys should be collected as soon as possible from interviewers and stored safely. Computer data must always be backed up.

Finally, it is important to ensure quality data entry. Using electronic data entry tools, such as mobile phones or personal digital assistants, can help to avoid data entry errors, as can standard quality control measures, such as entering the same data twice.

## ANALYSIS AND REPORT WRITING

Once the baseline data have been collected, the lead evaluation expert and the research assistant should complete the baseline analysis and report. As there are no programme results to report at this stage, the baseline report will consist of descriptive statistics. The average values of the demographics of the treatment and comparison groups should be compared to ensure that the necessary similarities exist between the two groups, and any statistically significant differences should be noted. Any issues that arose during the data collection phase should also be presented in the baseline report.

### TIP



To ensure that final evaluation results are considered reliable later on, it is good practice to include external experts in the review process for the baseline and final report. Moreover, by disseminating the baseline report, programme and evaluation staff can create public interest in the ongoing research and strengthen the ownership and dialogue among internal and external stakeholders.

As we have seen in Note 5, the validity of each impact evaluation method rests on a number of assumptions. The baseline analysis can play an important role in verifying these assumptions to confirm that our evaluation method of choice can be used or, if problems are encountered, how to resolve the issue.

## TROUBLESHOOTING: CONDUCTING A BASELINE SURVEY AND ANALYSIS

- ▶ **Finding respondents:** It may be difficult to locate youth for the survey. In this case, it is advisable to involve local programme staff and other stakeholders in finding suitable participants.
- ▶ **Data quality.** Even professional survey firms may not always have a good understanding of impact evaluation and may not be as qualified and reliable as one might hope. Interviewers may falsify or incorrectly record information. Poor data collection methods should not be tolerated. If contrived or low-quality data is discovered, it is important to let the survey firm know that this is unacceptable and the data collection must be done again to ensure high standards. To reduce and detect these cases, make sure that an independent auditing team is in place to oversee the data collection. Randomly auditing a small percentage of surveys is customary to ensure good

practice. When problems are found, some enumerators may need to be retrained, or even fired.

- ▶ **Data loss:** This can happen if completed questionnaires are lost or computers are stolen or malfunction. Computer data should always be backed up. In the field, surveys should be collected from interviewers as soon as possible, two to three times per week, if possible, to protect against loss. In the event that data are completely lost, it is best to go back and recollect the data. This entails revisiting already surveyed individuals and explaining to them that we need to ask the questions again. This can be very annoying to the respondents and costly for the programme.
- ▶ **Data entry:** Data entry should be performed promptly as surveys are collected. This allows problems to be identified quickly and corrected in the field. In addition, errors

often occur during data entry. Most data entry computer packages allow for (but do not require) double entry, in which each value must be entered twice. Transcription errors are further minimized by the use of mobile phones, personal digital assistants, laptop computers or tablets in data entry.

- **Incorrect assumptions:** The main assumptions for the chosen evaluation design may not hold. By always using verification and

falsification tests, we can detect these cases during baseline analysis and take appropriate action, including modifying the evaluation strategy. To reduce the chances that our chosen design is invalidated, it is important that the evaluation and programme staff maintain close communication and cooperation, ensuring that programme registration and data collection are in line with the evaluation requirements.



## Step 7: Conduct follow-up survey and analysis

When an evaluation method relies on collecting new data, the follow-up or endline survey will provide the long-awaited data that will allow us to analyse whether our intervention

was successful or not. When an evaluation is based entirely on existing data, then its analysis will be conducted during this stage.

### TIMING

The programme manager and lead evaluator will jointly determine the timing of the follow-up survey. Not every programme benefit will be observable immediately after the intervention, so the follow-up survey must be conducted after enough time has passed for the impact to materialize. The time varies according to programme and depends very much on the specific outcomes of interest. For example, young people participating in a training programme may actually face a short-term disadvantage in terms of earnings compared

with their peers, since they cannot work during the training course. However, if our training provides relevant skills, we would expect them to have a relatively higher income over the medium to long term. The timing of the follow-up will be crucial to identifying the true effect of the intervention.

If we want to measure both short- and long-term outcomes, we will need to conduct several follow-up surveys. Although this will increase the cost of the evaluation, it may also

### Box 6.8: Example – Timing is everything

In Kenya, the ILO ran an impact evaluation of a women's entrepreneurship training package called "Get Ahead". Researchers took a sample of 3,537 baseline firms and randomly assigned them into treatment and control groups. Outcomes were measured one year and three years after training occurred.

One year after the training there were limited effects in terms of business performance or well-being. However, three years after participating in training, female entrepreneurs had 18 per cent higher sales and 15 per cent higher profits than their untrained peers. Trained women also had improved mental health and a better standard of living.

The fact that it took three years for significant impact to manifest itself has implications for both the timing of interventions and evaluation activities. Data collection that is not well-timed risks leading to premature – and possibly inaccurate – conclusions about programme effectiveness.

Further information can be found in [ILO, 2017](#).



drastically enhance its value. Impact evaluations that follow treatment and comparison groups over many years are relatively rare, and their results are all the more in demand and appreciated. Conducting more than one follow-up survey will also allow us to analyse how the programme outcomes change over time.

## TRACKING

One major difference between the baseline and endline surveys is the issue of **tracking** respondents. If the surveyed youth are not found at the time of the follow-up survey, it can introduce very serious biases to the analysis and reduce the value of findings. For instance, if the lowest performing participants drop out, the evaluation results are likely to overestimate the impact of the programme. Equally, the most able youth might drop out. Because we cannot be sure whether **attrition** will lead us to underestimate or overestimate impact, minimizing attrition is essential to conducting any good evaluation. Although it is almost never possible to find 100 per cent of the individuals who were previously surveyed, every effort must be made to find as many as possible. A generally acceptable rate of attrition is 5–15 per cent, meaning that at least 85 per cent of youth in both the treatment and comparison group should be located.

### TIP



It is often possible to identify intermediate indicators that are consistent with the anticipated long-term outcomes. For example, the impact of entrepreneurship education and promotion programmes on the probability of starting a business might not always materialize for a number of years (students leave school, get a job to gain relevant experience, and eventually consider starting their own business.) By measuring short- and medium-term outcome indicators, such as business skills, the preference for starting a business as a career choice and concrete steps taken towards starting a business, it is possible to obtain intermediate impact results without having to wait for several years.

Tracking people, especially highly mobile youth, can be difficult. The following are three common ways to reduce attrition:

- ▶ **Gather good contact information during the baseline survey:** The baseline survey should include various types of contact information (street address, email address, phone number, etc.) from the respondent and also from friends and family who can help to locate the youth for the follow-up survey. Using social media channels, such as Facebook, can also help to keep track of young people.

### DEFINITION

**Tracking:** Tracking respondents throughout the study is crucial because if those surveyed at the baseline cannot be found for the endline survey it can introduce very serious biases into the analysis and reduce the value of findings.

### DEFINITION

**Attrition** refers to the drop-out rate of participants or survey respondents. This represents a problem for the evaluation because the dropouts are likely to be systematically different from those who can be found, thus skewing our results. Attrition can occur for any number of reasons, such as loss of interest in the programme, migration or simply unwillingness to participate in the survey.

- ▶ **Motivate youth in treatment and comparison groups to be available for future surveys:** Incentives to participate in follow-up surveys can include small payments to compensate for lost time or lotteries for cash or prizes. Youth can be notified of these incentives through prearranged communication (perhaps during the baseline survey), through mass media, such as radio and newspaper advertisements as well as through social media channels.
- ▶ **Use a tracking survey:** For evaluations that have a significant length of time between the baseline and endline, such as two years or more, and especially for those that do not use a baseline, a short, fast tracking survey can be used to estimate the likely attrition rate and gather additional information. If the programme is budget-constrained, it might be worth considering conducting follow-up surveys by phone to get up-to-date contact

**TIP**

Additional ways to facilitate tracking include the following:

- ask the advice and help of local leaders, officials and residents: locals may know the best way to find someone
- involve field enumerators from the study location, since they are familiar with the area and local customs
- if participants are still cannot be found, select a random sample of those not found and conduct a thorough and diligent search for them. If random selection is used, those who are eventually found through a more intensive search can be considered representative of others who have not been found.

information from survey respondents, while limiting personal visits to those youth who cannot be reached by phone.

## FOLLOW-UP SURVEY DESIGN AND DATA COLLECTION

It is likely that the programme or evaluation team will want to add a few additional questions to the original survey. These may include

questions about attendance, dropout and motivations for both, since this information can be used to estimate how much treatment

### Box 6.9: Examples of effective tracking of youth

In the Middle East, a survey company provided mobile phone charge cards to motivate youth to participate in a survey. To save costs, the survey company asked mobile phone operators to provide these cards as in-kind donations. Mobile phone companies provided 10,000 cards at US\$2 each. For the phone companies, it was good publicity at minimal cost.

In Uganda, the Northern Uganda Social Action Fund programme hired a firm to conduct a ten-minute tracking survey of respondents one year after the baseline and one year before the endline. The questionnaire asked participants who could be located easily for their updated contact information. For those who could not be easily found, information was collected from friends and family on the likely whereabouts of the person. This information was then kept for the endline to aid the teams in finding survey respondents, as well as to give the teams an indication of how hard or easy it would be to find people.

individuals actually received. New questions will need to be piloted and revised as necessary. In general, it is best to keep follow-up questions and the order of questions as similar to the baseline survey as possible to ensure that they are comparable. Unless there was a major issue with a question in the baseline survey, it is best to leave the wording unchanged in follow-up surveys. The survey manual will also need to be updated to reflect any changes from the baseline. In particular, it should include specific protocols for tracking survey participants.

Finally, interviewers will need the same level of training and oversight as for the baseline survey to ensure the best quality of data collection. If possible, select the best interviewers

#### TIP



#### Common areas for additional follow-up survey questions:

- Reasons for not participating or for dropping out
- Frequency of participant attendance or amount of benefits received
- Participant satisfaction with the programme
- Participant rating of the quality of the programme
- Participant self-assessed outcomes of the programme.

from the baseline staff to conduct the follow-up survey. Interviewers with high error rates or those who were less reliable should be replaced or given additional training.

## FINAL ANALYSIS AND EVALUATION REPORT

After follow-up data are collected, the final impact evaluation report can be produced, which represents the main product of the evaluation. The final report will repeat much of the information presented from the baseline survey and will add detailed information on the endline survey administration and final data analysis.

The analysis will be based on the outcomes and variables previously identified. In some rare cases, the analysis can be done by a simple comparison of the average values

between the treatment and comparison groups (usually in the case of lottery designs). In practice, however, some form of **regression analysis** will be applied to control for multiple key variables that might otherwise bias the results.

Box 6.10 presents a sample outline for sections of an evaluation report. All of this information is important to ensure that someone not involved in the evaluation can interpret the results correctly.

#### DEFINITION

In statistics, **regression analysis** includes any techniques for modelling and analysing several variables. In impact evaluation, regression analysis helps us to understand how the typical value of the outcome indicator changes when the assignment to treatment or comparison group is varied while the characteristics of the beneficiaries are held constant.

### Box 6.10: Example of the outline for evaluation reports

Evaluation reports that report on results from the baseline survey might be structured as follows:

1. Introduction
  - 1.1 Description of programme and evaluation
  - 1.2 The research team
  - 1.3 Report overview
2. Background
  - 2.1 Setting and location
  - 2.2 Historical background
  - 2.3 Scientific background
  - 2.4 Programme description and implementing partners
3. Intervention
  - 3.1 Group and participant selection
  - 3.2 Description of intervention
  - 3.3 Issues with implementation
4. Impact evaluation design
  - 4.1 Intervention objectives and hypothesized outcomes
  - 4.2 Research design and randomization
  - 4.3 Outcome measures
    - 4.3.1 Primary desired outcomes
    - 4.3.2 Secondary desired outcomes
    - 4.3.3 Adverse outcomes
    - 4.3.4 Other measures of interest
    - 4.3.5 Treatment heterogeneities
  - 4.4 Problems encountered
  - 4.5 Intervention and evaluation flow chart and timeline
5. Baseline survey administration
  - 5.1 Individual and group surveys
    - 5.1.1 Baseline survey development and pre-testing
    - 5.1.2 Enumerator/survey firm recruitment and training
    - 5.1.3 Baseline survey implementation
    - 5.1.4 Problems and concerns
  - 5.2 Other surveys
6. Baseline analysis
  - 6.1 Baseline characteristics of participants
  - 6.2 Power calculations and tests of balance on baseline data
  - 6.3 External validity
  - 6.4 Data quality issues
7. Conclusions
  - 7.1 Discussions
  - 7.2 Interpretation
  - 7.3 Generalizability

Evaluation reports that are written after the endline data collection should add the following sections:

- 7. Endline survey administration
  - 7.1 Endline individual and group survey
    - 7.1.1 Endline survey development and pre-testing
    - 7.1.2 Survey firm/interviewer recruitment and training
    - 7.1.3 Mobilization and tracking protocols
    - 7.1.4 Endline survey implementation
  - 7.2 Qualitative protocols
  - 7.3 Problems and delays
  - 7.4 Data quality issues
- 8. Data analysis
  - 8.1 Statistical methods used
  - 8.2 Levels of analysis
  - 8.3 Summary of outcomes
  - 8.4 Ancillary analyses
- 9. Conclusions
  - 9.1 Discussions
  - 9.2 Interpretation
  - 9.3 Generalizability
  - 9.4 Directions for future research

Source: Based on [Bose, 2010](#).

## Understanding heterogeneity

Not all programme beneficiaries may benefit from our intervention in the same way. Therefore, one important benefit of evaluation is to understand the variation in programme impacts. For instance, many programmes want to know whether boys or girls, younger or older youth, or those with higher or lower levels of education or experience perform better in the programme. In addition to looking at

gender, age or education, we may also want to assess whether outcomes differed according to participants' initial wealth (the value of participant assets), social capital (access to networks) or psychological traits (optimism, risk-taking attitudes, etc.). Understanding which participants have benefited the most and which the least from our programme can help us to achieve better design or target the intervention more effectively.

### DEFINITION

**Impact heterogeneity** refers to differences in impact by type of beneficiary; that is, how different subgroups benefit from an intervention to different extents.

For example, if our evaluation finds that an employment training programme had a greater impact on women, future iterations of the programme could focus more on women to increase the overall return of the programme. Alternatively, depending on priorities, we could explore ways to get men more involved so that they, too, benefit from the programme.

However, heterogeneities of interest should be specified in advance of any analysis and all results should be reported, not just those found to be statistically significant. We want to avoid data mining, which can be an especially serious problem with heterogeneity analysis.

## Interpretation of results

**Quality of implementation:** Results depend a great deal on how well an intervention was implemented. The final evaluation report should therefore discuss the quality of the implementation in detail. Having thorough

knowledge of how the programme was implemented is particularly important when evaluation results show a limited or negative impact since a deep understanding allows us to differentiate problems with implementation from problems with programme design. In order to be able to accurately interpret the evaluation results, it is necessary to embed the impact evaluation in a framework of strong monitoring, process evaluation and other qualitative tools.

**Generalizability of findings:** Ideally, our impact evaluation has external validity, which means that we can generalize our findings to cover similar populations in other contexts at other times. Whether this is the case largely depends on the sampling strategy chosen in the evaluation, and the nature of the outcomes in question. The more representative the sample, the more confident we can be that a programme would also work with different or larger groups of beneficiaries. This has important implications in terms of scalability and replication of the intervention. In general,

### Box 6.11: Impacts of business training and financial literacy

Bruhn and Zia (2011) studied the impact of a comprehensive business and financial literacy programme on firm outcomes of young entrepreneurs in an emerging post-conflict economy, Bosnia and Herzegovina. Although they did not find significant average treatment effects of the training programme on business performance, they identified high levels of heterogeneity among participants. Specifically, young entrepreneurs with relatively advanced financial literacy prior to the programme were found to exhibit improvements in sales due to the training programme. The effects on profits were also positive for this sub-group. The results suggest that training should not be the sole intervention to support young entrepreneurs and that the content of the specific course may have been appropriate for a very specific set of young entrepreneurs, but not for all.

it is prudent to assume that changes over time, different environments and different delivery mechanisms between one site and another have the potential to significantly affect the impact of the programme in either direction. We should therefore always be careful when translating evaluation lessons from one programme to another and be mindful that monitoring and evaluation will always be necessary for continuous learning and programme improvement.

#### TIP



Having good attendance data from programme monitoring is extremely useful as it tells us not only how many youth were enrolled but also the extent to which the services offered were used. This allows us to distinguish between regular and irregular participants and identify if someone drops out in the middle of the programme (possibly to be replaced by someone else). If this information is not collected and analysed, it is likely that an impact evaluation will underestimate programme effectiveness. Such information also helps us understand the effect of different dosages; for example, the difference in outcomes for someone who received 100 hours of training compared to someone who received only 50 hours.

## TROUBLESHOOTING: CONDUCTING A FOLLOW-UP SURVEY AND ANALYSIS

► **Attrition:** Attrition is a serious problem for studies and can greatly decrease the value of the findings. Clearly, prevention is better than mitigation. Obtaining good contact information during the baseline survey, providing incentives for youth to participate in the survey and using tracking surveys can all help to minimize attrition. If, despite prevention efforts, the programme experiences high levels of attrition, one mitigation technique is to select a random sample of individual who have not been located and to conduct a thorough and diligent search for them. These individuals, if found, may be considered to adequately represent those not tracked. Finally, since a degree of attrition is unavoidable, it is also possible to account for that attrition when defining the evaluation sample. Making the sample 10–20 per cent bigger than the minimum requirement allows for a large enough number of survey

responses to find statistically significant results even given the effects of attrition (although this approach does not offset the potential bias from attrition).

► **Non-compliance:** In addition to attrition, there may be other cases where people do not fully comply with a programme's selection criteria. For example, youth selected to participate in a training programme may not actually take part, while others who were assigned to the comparison group may, in fact, attend the training. A strict comparison of outcomes between the official treatment group and the comparison group will then misrepresent the actual impact of the programme. As long as the number of these cases is limited, and we can identify precisely which individuals were in the treatment and comparison groups and how much training they each received



(via programme records), it is possible to correct for non-compliance using statistical techniques, the “treatment-on-the-treated” estimate, which the evaluator will be able to calculate.

- **Black-box evaluation:** Another common problem at the follow-ups stage is the lack of knowledge about how well the

programme was implemented. This leads to evaluations that cannot attribute observed changes (or the lack thereof) to programme design or implementation. A common solution is to integrate findings from the monitoring system and to complement the impact evaluation with a process evaluation (see also Note 5 on the strengths of mixed-method designs).

### Box 6.12: Data mining

Data mining is a serious problem within statistics. It is especially problematic in the case of very long surveys that ask a large number of questions, often in different ways.

In data mining, a person seeks out results that confirm specific beliefs about a programme and ignores results that do not confirm these beliefs. For instance, a programme officer may strongly believe that a training programme has a positive impact on youth. Once the officer receives the data from the evaluation, she finds that there is a statistically significant increase in time spent working, but the youths’ average income is not statistically higher. Reporting only the increase in time spent working and not the fact that there is no change in income is one form of data mining.

Data mining can happen in two ways. The first is when we ignore evidence that is counter to our beliefs and report only those that confirm our beliefs. The second is a statistical anomaly. In statistics, there is always a chance that a variable will be found to be significant. In fact, at least 5 per cent of the time, something will be identified as significant which is in fact not significant. If an evaluator collects 100 pieces of information, at least five will be incorrectly attributed to be significant, when they are not. If the researcher looks for these five, and reports only these five, then the results are, factually incorrect.

An evaluation may find no statistically significant impact from a programme. However, by exploring every possible heterogeneity it is very likely that, due to statistical randomness, researchers will find some impact on a group. To avoid data mining, we should identify all of the outcomes of interest before conducting the analysis, and report all of these outcomes without fail, including those where no impact was found. In this way, the whole picture can be understood.



## Step 8: Disseminate findings

Once the results of the impact evaluation have been obtained, the final step is to disseminate the results to programme staff as well as to

those outside the programme who may be interested in the findings.

### INTERNAL DISSEMINATION

Internal dissemination of an evaluation provides the basis for organizational learning. Sharing results with the programme staff and the rest of the organization fulfils one of the main motivations for conducting an evaluation in the first place: enhanced programme management (see Note 1). In order to generate interest and ownership, the process of internal dissemination should ideally start immediately

after the baseline survey is completed; for example, by sharing and presenting baseline findings. The results of the evaluation should then be disseminated to executives and management in country offices and headquarters, where applicable. The report could include a discussion about how the results can affect the design of future or current initiatives.

### EXTERNAL DISSEMINATION

Dissemination should also target external stakeholders, such as local authorities, national ministries, local and international NGOs, universities (especially the development, economics and public health departments), multilateral organizations (such as the UN, World Bank and regional development banks) or bilateral donors (e.g., USAID, GIZ, DFID).

Indeed, impact evaluation findings are generally in high demand, especially in the youth employment field, where rigorous evidence on what works and what doesn't is still scarce.

External dissemination is covered in more detail in Note 7: Evidence uptake in policy formulation.

### TROUBLESHOOTING: DISSEMINATING FINDINGS

► **Limited use of the evaluation findings:** If the results of the evaluation are not shared sufficiently widely with internal and external stakeholders, then the evaluation's main objectives of facilitating learning for the programme and the youth employment sector as a whole are compromised. One way to overcome this issue is to define a

dissemination strategy (see Note 7) from the outset of the evaluation and to insist that at least one programme staff member works closely with the evaluation team. This ensures that at least one key person in the programme understands the evaluation and is well positioned to implement some of the report's findings.



## KEY POINTS

1. **Conducting an impact evaluation can be an expensive and time-consuming task**, with many potential pitfalls. It is therefore essential to assemble a high-quality team that can work on the evaluation over an extended period of time.
2. **The evaluation plan is the first major product of an impact evaluation**. It lays out the strategy for how to evaluate the intervention, including the research methodology, the sample size, the data collection plan and other elements.
3. **Interviewing children and youth poses particular challenges**, from obtaining parental consent to using appropriate language, so hiring a survey expert is advisable. Moreover, evaluations can raise ethical questions, so IRB approval should be sought for the evaluation design and the survey, once drafted.
4. **Conducting a baseline survey is highly recommended** as it provides valuable information to inform the programme design and allows us to verify the feasibility of the chosen evaluation design.
5. **The timing of the follow-up data collection has to be carefully thought through** to capture the outcomes of interest, some of which may occur in the short term, while others may need years to materialize.
6. **It is crucial that evaluation findings, whether positive or negative, are widely disseminated**. Sharing findings with internal, local and international stakeholders provides the basis for learning and feedback.



## KEY RESOURCES



- Bose, R. 2010. *A checklist for the reporting of randomized control trials of social and economic policy interventions in developing countries: CEDE Version 1.0*, Working Paper No. 6 (New Delhi, International Initiative for Impact Evaluation).



- ILO. 2017. *Policy guidelines for evaluation: Principles, rationale, planning and managing for evaluations*, 3<sup>rd</sup> edn (Geneva), see Chapter 4: Conducting the evaluation.



- Gertler, P.J.; Martinez, S.; Premand, P.; Rawlings, L.B.; Vermeersch, C.M., 2016. *Impact evaluation in practice*, Second Edition (Washington DC, Inter-American Development Bank and World Bank).



## REFERENCES

- ▶ Bertrand, A.; Beauvy-Sany, M.; Cilimkovic, S.; Conklin, S.; Jahic, S. 2009. *Monitoring and evaluation for youth workforce development projects* (Washington, DC, SEEP Network).
- ▶ Bose, R. 2010. *A checklist for the reporting of randomized control trials of social and economic policy interventions in developing countries: CEDE Version 1.0*, Working Paper No. 6 (New Delhi, International Initiative for Impact Evaluation).
- ▶ Brener, N.D.; Billy, J.O.G.; Grady, W.R. 2003. "Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence From the scientific literature", *Journal of Adolescent Health*, Vol. 33, pp. 436–457.
- ▶ Bruhn, M.; Zia, B. 2011. *Stimulating managerial capital in emerging markets: The impact of business and financial literacy for young entrepreneurs*, Policy Research Working Paper 5642 (Washington, DC, World Bank).
- ▶ Gertler, P.J.; Martinez, S.; Premand, P.; Rawlings, L.B.; Vermeersch, C.M., 2016. *Impact evaluation in practice*, Second Edition (Washington DC, Inter-American Development Bank and World Bank).
- ▶ International Labour Organization (ILO). 2017. *Developing markets: How training female entrepreneurs can improve business practices and profits: Evidence from Kenya*, What Works in SME Development Issue Brief No. 4 (Geneva).
- ▶ National Institutes of Health (NIH). 2008. *Protecting human research participants* (Bethesda, NIH, Office of Extramural Research).
- ▶ National Opinion Research Center (NORC). 2007. *Operational plan for the baseline survey of "Mi Primer Empleo" Youth Employment Program, Honduras* (Chicago, NORC).
- ▶ Powell, M.A.; Taylor, N.; Fitzgerald, R.; Graham, A.; Anderson, D. 2013. *Ethical research involving children* (Florence, UNICEF).
- ▶ Schenk, K.; Williamson, J. 2005. *Ethical approaches to gathering information from children and adolescents in international settings: Guidelines and resources* (Washington, DC, Population Council).
- ▶ World Bank. 2007. *Data for impact evaluation*, Doing Impact Evaluation No. 6 (Washington, DC), pp. 8–11.



© ILO / M. Crozet

## Case study

# SURVEY DESIGN AND IMPLEMENTATION FOR NEQDAR NESHAREK IN EGYPT

This case study is based on the questionnaire developed for the impact evaluation of the Neqdar Nesharek programme in Egypt.



## Learning objectives

By the end of this case study, readers will be able to demonstrate the following learning outcomes:

- ▶ an understanding of the key considerations when conceptualizing the design of a survey by considering the target population and the size and location of the survey
- ▶ knowledge of how to design labour market focused survey questions, building on guidance provided in Note 2 on key indicators relevant for youth employment
- ▶ a clearer understanding of how to supervise a data collection assignment by hiring a quality data collection firm and designing terms of reference to outline key deliverables.

## Introduction and case study context

The Neqdar Nesharek (NN) programme targets 4,500 young women aged 16–29 years old in 30 rural villages in the Fayoum, Qena and Suhag governates in Upper Egypt. The programme aims to empower young rural women in Upper Egypt by providing them with business and vocational skills training and supporting them in starting a business or seeking employment. It also aims to increase social empowerment for young women, while emphasizing the importance of involving women's "gatekeepers" (husbands and fathers) and community leaders. The business

skills curriculum is delivered over 12 weeks, at three sessions of two hours each week (a total of 72 hours).

An impact evaluation is being designed to accompany the programme to provide a rigorous assessment of the programme's impacts. The evaluation relies on a quasi-experimental approach that combines a difference-in-differences design with propensity score matching and will make use of data from a midline survey and a follow-up survey.

## Part I: Survey conceptualization

The midline survey will include 7,028 young women and should be conducted over a three-month period. The survey respondents will mainly be young women with a basic level of

education and a basic level of Arabic reading and writing skills. As they are young women, a parent, peer or other chaperone should be present when conducting the survey.



## DISCUSSION TOPICS

Given the large sample size, the rural nature and the specific cultural and social barriers of the sample population, what important design

elements should be considered when conceptualizing the survey?

## Part II: Designing labour market related survey questions

As the M&E officer for the NN programme, you have been asked to supervise the survey design and data collection process. In addition to modules on education, health, social networks and mobility, the survey requires the measurement of economic and labour market related indicators, including:

- ▶ employment status: whether the respondent is currently employed (self-employed or in wage employment), unemployed or not participating in the labour market (inactive)
- ▶ earnings: remuneration that the project beneficiary obtains from his or her work, in cash or in kind
- ▶ working hours: the number of hours worked in the reference period (e.g. per week)
- ▶ job satisfaction: level of satisfaction with their current job
- ▶ income-generating activity: whether a woman was involved in any economic activity with the goal of generating income during the three months prior to the survey interview
- ▶ economic aspirations: whether the woman plans to (a) set up or continue a business project or (b) obtain wage employment.



### DISCUSSION TOPICS

1. What criteria does a survey respondent have to meet to be considered employed, unemployed or inactive? What questions need to be asked to determine a person's employment status?
2. What are the important considerations in determining how many hours a survey respondent has worked?

## Part III: Terms of reference for a data collection firm

An external firm, the Egyptian Demographic Association (EDA) will be contracted to collect the midline data of the NN programme evaluation. As the M&E officer for the programme,

you are responsible for developing the terms of reference for the external data collection firm, EDA.



### DISCUSSION TOPICS

1. What are the key areas of responsibility and deliverables which you would require from the survey firm?







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

[youth@ilo.org](mailto:youth@ilo.org)  
[www.ilo.org/yep](http://www.ilo.org/yep)  
[www.ilo.org/taqeem](http://www.ilo.org/taqeem)

978-92-2-131666-4

