



Guide on **Measuring Decent Jobs for Youth**

Monitoring, evaluation and learning in labour market programmes

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Overview



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

Overview

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. 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 to find the reproduction rights organization in your country.

Guide on Measuring Decent Jobs for Youth - Monitoring, evaluation and learning in labour market programmes, Overview/ International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131670-1 (print)
978-92-2-131671-8 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلم في برامج سوق العمل النشطة - نظرة عامة
ISBN 978-92-2-630801-5 (print), 978-92-2-630802-2 (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.

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

Printed in Switzerland

Acknowledgements

The Guide on Measuring Decent Jobs for Youth represents a collaborative effort involving expertise from the International Labour Organization (ILO) and its main constituents as well as from external partners. The Guide is a product of the ILO's Youth Employment Programme Unit (YEP) and is developed in close collaboration with the other branches and units of the Employment Policy Department including the Country Employment Policy Unit of the Employment and Labour Market Analysis Branch, the Skills and Employability Branch and the Development and Investment Branch. Other main contributors include the Evaluation Office, the Enterprise Department, the Research Department, Better Work, the Regional Office for the Arab States, the Decent Work team for North Africa and Country Office for Egypt and Eritrea, the Country Office for Algeria, Libya, Morocco, Tunisia and Mauritania, the Bureau for Workers' Activities, the Bureau for Employers' Activities and the International Training Centre of the ILO (ITCILO).

The drafting of the guide was made possible in-part through generous support from the International Fund for Agricultural Development (IFAD), under an IFAD-ILO project entitled "Strengthening gender monitoring and evaluation in rural employment in the Near East and North Africa." Through rigorous impact research, this capacity development and learning grant project aims to understand "what works" in the promotion of gender mainstreaming, with the ultimate goal of reaching gender equality in rural employment outcomes across the region.

The drafting process included review at an experts meeting in March 2017 where excellent

comments were received. The expert meeting included: Samuel Asfaha, Naomi Asukai, Nathalie Bavitch, Romulo Cabeza, Juan Chacaltana, Veronica Escudero, Sergio Iriarte Quezada, Jean-François Klein, Tobias Lechtenfeld, Pedro Martins, Nawel Marzouki, Rute Mendes, Mohammed Mwamadzingo, Niall O'Higgins, Aurelio Parisotto, Miquel Pellicer, Arianna Rossi, Roland Sarton, Olga Strietskallina, Steven Tobin, Sanchir Tugschimeg, Peter Van Rooij and Peter Wichmand.

Drew Gardiner coordinated the drafting of the guide and authored Notes 1, 3 and 7. Jonas Bausch authored Notes 4 and 5. Paul Berbée, Verena Bruer, Paul Dyer, Sonja Kovacevic, Susana Puerto Gonzalez and Felix Weidenkaff provided substantial technical inputs throughout. Matt Ripley authored several of the case studies and acted as penultimate reader. Copy-editing was done by Book-Now and typesetting by ITCILO.

Note 6, "A step-by-step guide to impact evaluation" is an adaptation of a chapter of the Global Partnership for Youth Employment's "Measuring success of youth livelihood interventions: A practical guide to monitoring and evaluation", authored by Kevin Hempel and Nathan Fiala.

Sangheon Lee, Director, Employment Policy Department, Sukti Dasgupta, Chief, Employment and Labour Markets Analysis Branch and Valter Nebuloni, Head, YEP, provided overall guidance to the production of this publication.

Questions or feedback on this guide can be submitted to YEP at youth@ilo.org



Contents

Overview



Note 1: Diagnosing, planning and designing youth employment interventions



Note 2: Concepts and definitions of employment indicators relevant for young people



Note 3: Establishing a monitoring system



Note 4: Enhancing youth programme learning through evaluation



Note 5: Impact evaluation methods for youth employment interventions



Note 6: A step-by-step guide to impact evaluation



Note 7: Evidence uptake in policy formulation

Appendix: Answer key for case studies *(available separately)*

Preface

In June 2012, the International Labour Conference of the ILO resolved to take urgent action to tackle the unprecedented youth employment crisis through a multipronged approach geared towards pro-employment growth and decent job creation. The resolution, entitled “The youth employment crisis: A call for action” (ILO, 2012a), contains a set of conclusions that constitute a blueprint for shaping national strategies for youth employment. The associated background report (ILO, 2012b) warns: “(M)ajor gaps in knowledge (on “what works” on youth employment) persist. There have been relatively few rigorous evaluations of youth employment policies and programmes, of their impact in the short and long term, and of their relative cost-benefit, including in developed countries. This needs to be remedied since lessons learned from evaluations can lead to greater programme effectiveness and better targeting of scarce resources. Continuous building of the knowledge base on country policies and programmes and the impact evaluations of the range of measures is a paramount priority.”

The ILO has responded to this call by making greater investments in efforts to develop the evidence base on youth employment. “What works in youth employment” is the ILO’s offer to constituents to assist them in the rigorous monitoring and evaluation of their youth employment programmes and policies. The objective is twofold: (1) to ensure accurate measurement of youth employment outcomes and (2) to promote evidence-based youth employment interventions and programmes through policy dialogue. By building capacities to measure results, the ILO contributes to tracking progress on Sustainable Development Goal (SDG) 8 on the promotion of sustained, inclusive and sustainable economic

growth, full and productive employment and decent work for all.

The ILO’s work on results measurement in youth employment began in 2010 with the Fund for Evaluation in Youth Employment. This action was complemented by the global “Youth employment crisis: A call for action” and its corresponding “Follow-up plan” (2012–2019), which appealed for improved assessment of interventions to support better youth employment outcomes. Then in 2013, the “Area of Critical Importance: What Works in Skills and Youth Employment” was set-up to provide financial and technical assistance for the rigorous assessment of youth employment. Regional approaches have since been established, including the Taqueem (meaning “evaluation” in Arabic) Initiative. Taqueem is a partnership between the ILO and the International Fund for Agricultural Development (IFAD) as part of an IFAD-financed project titled “Strengthening gender monitoring and evaluation in rural employment in the Near East and North Africa”. Through rigorous research, this capacity development and learning project aims to understand “what works” in the promotion of gender mainstreaming, with the ultimate goal of achieving gender equality in rural employment outcomes across the region.

It is in this context that we have developed the present guide. It offers a comprehensive and accessible introduction to the topics of results measurement and impact assessment, their practical application in the youth employment field and how evidence created via results measurement strategies can lead to improved programming. It is our sincere hope that this guide will help social partners and practitioners to make informed decisions

in choosing the evaluation frameworks that benefit their organizations and youth-centred programmes and will contribute to enhancing the youth employment sector in general. If, in the future, we are able to draw more robust evidence from all the good work being undertaken to support youth throughout the world and the wide experience of the various actors,

we will also have a stronger voice for convincing policy-makers to scale up interventions that have proven successful. We look forward to the continued collaborative work of policy-makers, development practitioners and other stakeholders in providing tomorrow's leaders with the economic opportunities they deserve.

Valter Nebuloni
Head
Youth Employment Programme Unit
Employment and Labour Market
Analysis Branch

Sukti Dasgupta
Chief
Employment and Labour Market
Analysis Branch
Employment Policy Department

Introduction

Investing today in the employment of young people means investing in the present and future of our societies.

Guy Ryder, ILO Director-General

Young people are estimated to account for over 35 per cent of the unemployed population worldwide (ILO, 2017a). While the global youth unemployment rate stabilized at 13 per cent in 2016, it rose slightly to 13.1 per cent in 2017. The estimated figure of 70.9 million unemployed youth in 2017 is an important improvement from the crisis peak of 76.7 million in 2009, but this number is expected to rise by a further 200,000 in 2018, reaching a total of 71.1 million. More importantly, 39 per cent of young workers in the emerging and developing world – 160.8 million youth – are living close to or in extreme poverty, i.e. on less than US\$3.10 a day (at 2011 purchasing power parity (PPP)). More than two in every five young people in today's workforce are unemployed or are working but remain poor, a striking reality that is impacting societies across the world.

Despite this challenging situation, the adoption of the Sustainable Development Goals (SDGs) in 2015, and their explicit focus on decent work, offers hope for young people making the transition to the world of work. The ILO and its constituents – governments, workers' and employers' organizations – are working towards global targets on decent work and inclusive growth. The interests of young people are strongly represented in the SDGs, specifically in Goal 1 on fighting poverty, Goal 4 on providing quality education and Goal 8 on decent work and economic growth. The well-being of the most vulnerable young people

is explicitly addressed in the Goal 8.6 target, “By 2020, substantially reduce the proportion of youth not in employment, education or training”. For youth employment practitioners, this group of excluded and marginalized young people is a focus of primary concern. The trajectory of this growing young generation will depend on their social and economic integration, their ability to live fulfilling lives and their participation in and contribution to society.

Taking effective action towards realizing these SDGs requires a systematic approach to measurement and evaluation. The United Nations (UN) community has laid out a results-based reporting plan, designed to assess and monitor the realization of the SDGs and the contributions of specific UN agencies and member States. In monitoring and evaluating SDG implementation, the Agenda 2030 document provides guidance for systematic and rigorous follow-up processes (UN, 2015). The Agenda 2030 document recognizes that the review process for the SDGs requires “enhanced capacity-building support for developing countries, including the strengthening of national data systems and evaluation programmes”.

This guide contributes to our collective efforts in moving towards fulfilling the SDGs and securing better youth employment prospects by promoting the implementation of effective results measurement systems in

youth employment programmes and interventions. With an extensive section on attributing changes to youth employment interventions, this guide fosters evidence-based policy and programming choices by equipping ILO partners and stakeholders with the tools to attribute their actions to positive youth related outcomes. Guidance on results measurement of decent jobs for youth is provided with the premise that, for jobs to be created, employment conditions to be improved and SDGs to be realized, we need more evidence on “what works”. And, more importantly, can we say “how” and “why” interventions work in order to inform future programme design?

Although knowledge gaps remain, we can safely say that the evidence base is stronger than it was 10 years ago and it continues to grow at an impressive rate. The recently released systematic review on “Interventions to improve the labour market outcomes of youth” (Kluve et al., 2017) identified 113 rigorous evaluation reports to include in its meta-analysis, 74 of which have been released since 2010. The main findings of the review were also encouraging: Overall, youth employment interventions can produce positive job and income effects for young people. As global evidence increases, so will the effectiveness

of youth employment interventions, resulting in the creation of more decent jobs for young people.

This guide proposes a framework that links practical, implementation-focused measurement and monitoring to research-oriented impact evaluation. This is achieved in a series of seven “notes”, which lead readers through the key steps in diagnosing and formulating youth employment interventions, setting up a results measurement system and defining corresponding key youth employment indicators. It then moves on to providing an overview of the types of evaluation approaches available to measure youth employment outcomes, including non-experimental, quasi-experimental and experimental methods that encourage reflection on the challenge of attribution. The menu of evaluation methods available is then discussed and, finally, guidance is provided on how to ensure that the findings from evidence-based evaluations are taken up in dialogue and policy formulation processes related to employment and young people. The technical notes are supplemented by interactive case studies supplied at the end of each note, to be used as a complement to classroom based learning on this topic.

ABOUT THIS GUIDE

Objective

With this guide, we aim to equip readers with the basic set of concepts and tools needed to make informed decisions about how best to measure and evaluate the results of youth employment interventions. We review the entire life cycle of a youth employment intervention, starting with diagnostics and programme design, then moving on to setting up a monitoring system and measuring results. We seek to provide a clear understanding of the variety

of evaluation options available and the issues to consider in choosing the most appropriate method, given the learning objectives and operational context. Additionally, the guide describes how to manage an impact evaluation, if that is the assessment method of choice.

Our overarching goal is to strengthen the foundation of sound programming and policy-making by increasing the number of quality evaluations in the youth employment field, thereby facilitating the process of scaling up

and replicating successful interventions. We want to promote a perception of evaluation as a tool for both internal and external learning rather than as simply an accountability exercise. Using evaluation to learn about the performance of an intervention can facilitate adaptive management where problems can be recognized early and projects adjusted accordingly. Learning that is shared publicly can help others to select the most effective intervention models, ensuring that resources spent on youth employment are optimally used.

Learning tool

The guide can also be used to complement learners taking training courses on results measurement and what works in youth employment. From 2010 to 2017, the ILO has trained over 2,000 constituents on results measurement in youth employment using an experiential learning approach called “Evaluation Clinics”. This guide closely follows the learning modules and curriculum of the ILO Evaluation Clinics but provides learners with more in-depth information and tools to build on the knowledge gained in the Clinics.

Audience

While this guide is primarily a reference tool for ILO constituents, it can also be used by all those involved in the implementation of youth employment programmes.

The guide has therefore been drafted with the following organizations in mind.

Primary audiences

Governments: Particularly ministries and agencies which focus on employment, youth, labour market services and training. The information may be more relevant to agencies

and institutions which are directly responsible for delivering and monitoring different programmes, such as youth employment agencies, public employment service agencies and technical and vocational training institutes. It can be used to assist in designing new public employment programmes.

Employers’ organizations: Organizations representing the interests of private employers who offer youth vocational education, employment and entrepreneurship opportunities. They will be able to use the guide both to inform their advocacy on important legislative and regulatory issues and to offer value-added services and advice to their members, based on the latest research and evidence. Companies have an increasingly large stake in issues involving youth and making sure they are prepared with the requisite skills and attitudes to become the employers and employees of the future.

Workers’ organizations: Workers’ councils and trade unionists who are interested in the potential of youth employment interventions for realizing decent work for youth. The guide will be particularly useful in formulating programmes and interventions for evidence-based advocacy around key policy issues. Workers’ organizations have an interest in learning about the conditions under which employment, and in particular decent employment, can be increased.

Youth-led organizations: not-for-profit organizations working in the youth employment field whose staff and members are predominantly made up of young people. Most organizations use the UN’s definition of youth which is a person between the ages of 15 to 24.

Secondary audiences

Civil society: NGOs are one of the main implementers of youth employment programmes

in many countries. As the guide covers the full life cycle of a youth employment intervention, it can be used at every stage of a project. The focus on evaluation and evaluation methods allows intervention results to be translated into lessons to inform both follow-up projects and other practitioners.

Donors: Multilateral and bilateral development agencies provide a considerable share of the funding for youth employment programmes. Donors, who have a fiduciary interest in maximizing the impact of their grant funding and ensuring accountability in the use of resources, will be able to use the guide to set appropriate evaluation schedules, as well as to design and commission evidence-based youth programmes.

Researchers: Researchers conducting field research and liaising with programme managers and implementing organizations will find the information presented in this guide relevant for adapting standard research methods to local programme conditions. A context-sensitive approach to impact evaluation is presented that facilitates research design.

Key references

This guide complements existing materials on results measurement, applying them to the specific area of youth employment. It builds on the following works:

- ▶ *Measuring success of youth livelihood interventions: A practical guide to monitoring and evaluation* (Hempel and Fiala, 2011).

Authored as a contribution to the Global Partnership for Youth Employment, this work was used as the key reference document.

- ▶ *ILO Policy Guidelines for Evaluation* (ILO, 2017b) and *ILO Development Cooperation Manual* (ILO, 2015), which set out the overall framework for project monitoring, reporting and evaluation in an ILO context.
- ▶ *Monitoring and evaluation of youth employment programmes: A learning package* (ILO, 2013). This guidebook provides advice on monitoring the performance of youth employment programmes and measuring both short- and long-term outcomes.
- ▶ Practical toolkits that emphasize general monitoring and evaluation (e.g. Gosparini et al., 2003; Kellogg Foundation, 2004; Kellogg Foundation, 2017, 2004) and other publications that focus specifically on impact evaluation (e.g. Baker, 2000; Duflo et al., 2006; Gertler et al., 2016; Khandker et al., 2010; Ravallion, 2008).

Considering similar resources on the topic of monitoring and evaluation, our Guide is unique in that it is the only guidance tool that orients all elements of the results measurement life cycle to the topic of youth employment. These unique elements include a chapter advising practitioners on the appropriate youth employment indicators to select (Note 2); a list of challenges, and their solutions, that are specific to youth employment focused impact evaluations (Note 5); and a reflection on the existing body of youth employment evidence and advice on how this evidence can lead to policy change (Note 7).

OVERVIEW OF THE GUIDE AND HOW TO USE IT

While the guide leads the reader through all stages involved in formulating youth employment programmes, starting with the diagnostic phase and closing with evidence uptake in youth employment policy formulation, the main thrust of the guide is on monitoring, results measurement and evaluation. Figure 0.1 sets out a simple results measurement cycle and the relating key elements of the seven notes.

The first note focuses on designing a youth employment intervention and establishing a solid theory of change as a basis for quality results

measurement. The second note covers labour market indicators with particular relevance for youth employment interventions, while Note 3 concentrates on setting up the monitoring system. Note 4 introduces readers to the importance of evaluation before presenting different approaches, including performance and impact evaluations. In order to address the challenge of attribution, Note 5 presents a number of different methods which can help decision-makers weigh the desired level of rigour with the feasibility of conducting the research. Note 6 guides readers through the step-by-step

FIGURE 0.1 RESULTS MEASUREMENT CYCLE



process of implementing a youth employment focused impact evaluation. The guide ends by offering practical advice on ensuring that evaluation evidence is taken up in policy formulation processes.

Although it is important to be familiar with all parts of the measurement process, it is not necessary to read the guide from beginning to end. Instead, each note is conceived as a standalone entity that can be read independently of the others, according to each reader's needs. Table 0.1 indicates which notes are most relevant to different types of readers.

Case studies

At the end of each of the seven notes, we present case studies detailing how different youth employment interventions apply results measurement strategies. The case studies illustrate the main points in each note and ask readers to apply the concepts they have learned to “real world” situations. The case studies are designed to complement classroom-based learning, as pedagogic exercises to be discussed in small groups with the assistance of an expert or facilitator in results measurement in youth employment.

While all of the case studies relate to youth labour market interventions, some of them are derived from experiences of the ILO in supporting organizations which are seeking to improve their results measurement systems and implement impact evaluation projects. Several of these case studies are drawn from organizations that were offered support under ILO's Fund for Evaluation in Employment, a technical and financial support programme for youth employment researchers and organizations in the Arab States and Africa regions.

The case studies are accompanied by an appendix “Answer key for case studies”, which is available separately from the seven note volume of Guide. This answer key appendix

is intended to be used by facilitators to assist in small group discussions about the case studies.

Overview of key terms

The guide addresses the effective monitoring and measurement of outcomes of youth employment interventions with a specific focus on impact evaluation. It is important here to make reference to four key terms which are used extensively throughout the guide: *results measurement system*, *monitoring*, *evaluation* and *impact evaluation*.

Results measurement system refers to the overall processes, plans, tools and resources that are used to determine whether a programme has been implemented according to the plan (monitoring) and is having the desired result (evaluation).¹ A results measurement system specifies:

- ▶ indicators to be tracked
- ▶ milestones (mid-stream) and end targets
- ▶ data collection tools
- ▶ the personnel who will gather, record and analyse the data and
- ▶ the types of reports that will be prepared, including for whom, why and how often.

The key activities of a results measurement system are *monitoring* and *evaluation*:

Monitoring tracks the implementation and progress of an intervention in order to support programme management. Monitoring:

- ▶ involves the collection of data on specific implementation and results indicators
- ▶ assesses compliance with work plans and budgets
- ▶ uses information for project management and decision-making
- ▶ is ongoing

¹ Also known as a monitoring and evaluation (M&E) system.

Table 0.1 Reader's guide

Note	Title	Description	Case study	Governments	Employers/workers	Civil society	Researchers	Donors
1	Diagnosing, planning and designing youth employment interventions	Guides readers through an employment diagnostic analysis as the basis for developing the theory of change then proceeds to programme design	Labour market diagnostics for the promotion of rural youth livelihoods in Zambia	✓	✓	✓	✓	✓
2	Concepts and definitions of employment indicators relevant for young people	Reviews concepts and definitions of employment indicators relevant for young people in the areas of employment opportunities, employment quality, employment access and employment skills	Selecting indicators for the Northern Uganda Youth Entrepreneurship Programme	✓	✓	✓	✓	
3	Establishing a monitoring System	Presents the main steps in developing a results measurement system, including how to collect and analyse data.	Establishing a monitoring system for the Jordan Economic Growth and Employment Project	✓	✓	✓		
4	Enhancing youth employment learning through evaluation	Asks which type of evaluation best suits an individual programme. The answer depends on evaluation questions, the context and characteristics of the project, and available resources	Developing Terms of Reference for a mid-term evaluation of a youth employment project in Egypt	✓	✓	✓		
5	Impact evaluation methods for youth employment interventions	Presents the main features of an impact evaluation and focuses on finding a good comparison group to reliably demonstrate impact. Presents different (quasi-)experimental methods for conducting an impact evaluation	Assessing rural microenterprise growth through different evaluation methods	✓	✓	✓	✓	
6	A step-by-step guide to impact evaluation	Moves from the conceptual to the practical level, describing the major steps involved in carrying out an impact evaluation and providing practical resources. These steps cover the entire process, from initial preparations to the dissemination of results	Survey design and implementation for Negdar Nesharek in Egypt	✓	✓	✓		
7	Evidence uptake in policy formulation	Helps readers to plan how evaluation results can be used to influence policy and improve programming. Communication of results and stakeholder engagement strategies are discussed	Uptake of evidence on the effects of skills training on young people's financial behaviour and employability in Morocco	✓	✓		✓	✓

- ▶ seeks to answer the question, “Are we doing things right?”.

Evaluation assesses the relevancy, efficiency, effectiveness and sustainability of an intervention to drive both accountability and learning. Evaluation:

- ▶ involves the collection of data on the design, implementation and results of a project
- ▶ focuses on the achieved results of the project relative to objectives (performance evaluation) or on the process of implementation (process evaluation)
- ▶ is periodic, usually conducted annually and at the completion of a project, and includes recommendations for follow up steps
- ▶ answers the question, “Are we doing the right thing?”.

An *impact evaluation* is a particular type of evaluation that attempts to attribute measurable outcomes on a specific population to a particular intervention. An impact evaluation:

- ▶ Answers the question, “What happened to the beneficiaries as a result of the intervention (as opposed to other contributing factors)”. This compares outcomes to the “counterfactual” – “What would have happened had there been no intervention?”.
- ▶ This causal link is established by comparing the outcomes experienced by participants with those experienced by a comparison group of non-participants.
- ▶ It is neither a requirement nor suited to all types of interventions.

Programme vs. project. Throughout the guide, we use the terms, “programme”, “project” and “intervention.” A *project* is a set of tasks executed over a given time period, with a defined scope and budget. The term *intervention* is used interchangeably with project. A *programme* usually comprises several

projects (focusing on different sectors, target groups or using different intervention combinations), all set up with the same overarching objective in mind. Throughout this guide, we focus primarily on youth-centred active labour market programmes but also draw on experiences from value chain and market development interventions.

Active labour market programmes

Active labour market programmes (ALMPs) are policy interventions aimed at increasing the employment probability of jobseekers in order to reduce aggregate unemployment, raise productivity and boost individual earnings (Auer et al., 2008, pp. 18–20). Unlike passive labour market policies (PLMPs), ALMPs exclusively target increased employment rather than other relief options, such as wage replacement. Both supply-side and demand-side measures fall under this category. Typical programmes on the labour demand side are public employment programmes and interventions designed to promote self-employment and entrepreneurship. Concerning labour supply, skills training is most prominent, this being the most common ALMP overall. Finally, ALMPs can also address market frictions by providing labour market information, offering employment services and registering vacancies (ILO, 2003, pp. 6–8). Good ALMPs are additionally characterized as those that encourage and allow for social dialogue; since collective bargaining mechanisms should not be compromised by incentives provided by ALMPs (ibid., pp. 3–5). Youth employment intervention types will be more extensively presented and discussed in Note 1.

References

- ▶ Auer, P.; Efendioğlu, Ü.; Leschke, J. 2008. *Active labour market policies around the world: Coping with the consequences of globalization* (Geneva, International Labour Office).
- ▶ Baker, J.L. 2000. *Evaluating the impact of development projects on poverty: A handbook for practitioners* (Washington DC, World Bank).
- ▶ Duflo, E.; Glennerster, R.; Kremer, M. 2006. *Using randomization in development economics research: A toolkit*. BREAD Working Paper No. 136.
- ▶ 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).
- ▶ Gosparini, P.; Russo, L.; Sirtori, M.; Valmerana, C. 2003. *The monitoring and evaluation manual of the NGOs of the Forum Solint* (Development Researchers' Network, Rome).
- ▶ Hempel, K.; Fiala, N. 2011. *Measuring success of youth livelihood interventions: A practical guide to monitoring and evaluation* (Washington DC, Global Partnership for Youth Employment).
- ▶ ILO. 2003. *Active labour market policies*, Proceedings of the 288th session of the Governing Body, Nov., GB 288/ESP/2 (Geneva).
- ▶ ILO. 2012a. *The youth employment crisis: A call for action*, Resolution and conclusions of the 101st session of the International Labour Conference (Geneva).
- ▶ ILO. 2012b. *The youth employment crisis: Time for action*, Report V, International Labour Conference, 101st session, Geneva, 2012 (Geneva).
- ▶ ILO. 2013. *Monitoring and evaluation of youth employment programmes: A learning package* (Geneva).
- ▶ ILO. 2015. *Development cooperation manual* (Geneva).
- ▶ ILO. 2017a. *Global Employment Trends for Youth 2017: Paths to a better working future* (Geneva).
- ▶ ILO. 2017b. *ILO policy guidelines for evaluation: Principles, rationale, planning and managing for evaluations*, 3rd edition (Geneva).
- ▶ W.K. Kellogg Foundation. 2004. *Logic model development guide* (Michigan).
- ▶ W.K. Kellogg Foundation. 2017. *The step-by-step guide to evaluation* (Michigan).
- ▶ Khandker, S.R.; Koolwal, G.B.; Samad, H.A. 2010. *Handbook on impact evaluation: Quantitative methods and practices* (Washington DC, World Bank).
- ▶ Kluge, J.; Puerto, S.; Robalino, D.; Romero, J.M.; Rother, F.; Stöterau, J.; Weidenkaff, F.; Witte, M. 2017. *Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services, and subsidized employment interventions*, Campbell Systematic Reviews 2017:12.
- ▶ Ravallion, M. 2008. "Evaluating anti-poverty programs" in T.P. Schultz and J. Strauss (eds): *Handbook of Development Economics*, Vol. 4 (Amsterdam, Elsevier), pp. 3787–3846.
- ▶ UN General Assembly. 2015. *Transforming our world: The 2030 Agenda for Sustainable Development*, 21 October 2015, A/RES/70/1.

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Note

1

Diagnosing, planning and designing youth employment interventions



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

NOTE 1
**Diagnosing, planning and designing
youth employment interventions**

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. 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 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 1: Diagnosing, planning and designing youth employment interventions / International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131656-5 (print)
978-92-2-131657-2 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلم في برامج سوق العمل النشطة - الملاحظة 1: تشخيص التدخلات التي تستهدف عمالة الشباب وتخطيطها وتصميمها، ISBN 978-92-2-630787-2 (print), 978-92-2-630788-9 (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.

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

Printed in Switzerland

Contents

Diagnosing, planning and designing youth employment interventions	1
Diagnostics: Understanding the labour market constraints faced by young people	2
Assessing the local labour market	2
Stakeholder consultation	4
Labour market status of young people	4
Labour demand and supply	8
Objectives and design: What do we want to achieve and how can we best achieve it?	14
Strategic and institutional framework	14
Setting the higher-level goal	17
Effective participation in objective setting	18
Setting outcomes	18
Choosing the youth population	19
Defining the intervention	20
Key points	25
Key resources	25
References	26
Case study: Labour market diagnostics for the promotion of rural youth livelihoods in Zambia	27
Figures	
1.1: An overview of youth labour markets indicators	5
1.2: Youth environments	11
1.3: How to develop higher-level goals	17
Tables	
1.1: Descriptions of labour market indicators for young people	6
1.2: Basic typology of youth skills level	19
1.3: Youth-focused ALMPs	22
1.4: Example of a menu of evidence-based interventions, by constraint	24





Diagnosing, planning and designing youth employment interventions



Prerequisites:

This note requires no prior knowledge. It guides readers through the processes and tools required to diagnose the state of labour markets and design an intervention focused on youth employment needs.



Learning objectives:

At the end of this note, readers will:

- ▶ understand the key steps for planning and designing a youth employment programme, from diagnosing needs, through setting objectives, to selecting interventions.
- ▶ be able to deploy appropriate labour market assessment approaches to identify supply and demand constraints facing youth in the labour market, including school-to-work transition surveys, value chain analysis and enabling environment for sustainable enterprises.
- ▶ know how to use the current evidence base to link the youth employment situation to the appropriate intervention type.



Keywords:

Labour market analysis, employment diagnostics, participatory design, youth unemployment, youth underemployment, working poor, informality, vulnerable employment, labour demand, Sustainable Development Goals, Decent Work Country Programmes, skills training, public employment programmes, employment services, entrepreneurship promotion.

Youth employment practitioners are required to make difficult decisions about how to spend limited resources to assist young people in their transition into the world of work. In order to find appropriate solutions, it is important to understand the specific context and constraints that are affecting young people. Before crafting a monitoring and evaluation system, we should make sure that our intervention itself is carefully planned:

- ▶ Do we have good knowledge about the needs of the people we are trying to support?
- ▶ Do we understand why certain conditions, such as youth unemployment, exist?
- ▶ Do we have a clear objective of what we want to achieve?
- ▶ Are we building on existing experience and evidence when designing our intervention to fulfil this objective?

The decision to formulate new strategies to support young jobseekers implies that there is a discrepancy between the status quo and what could be or should be. The gap between existing conditions and the desired condition is commonly referred to as the need. We must understand the need before designing an intervention.

Diagnosics: Understanding the labour market constraints faced by young people

ASSESSING THE LOCAL LABOUR MARKET

In order to systematically assess both the need and the economic and social environment in which youth are transitioning into employment, the labour market should be analysed. This includes diagnosing the employment situation, based on the overall state of the economy, as well as a set of contextual factors, such as demographics, educational systems, the human resource base, the regulatory framework and social norms. In this context, the International Labour Organization and the German development agency,

GIZ, provide two labour market analysis tools that can help guide youth employment practitioners in developing a labour market assessment: the Employment Diagnostic Tool (EDA) (ILO, 2012a) and the Employment and Labour Market Analysis (ELMA) (GIZ, 2014).

The focal point of an employment assessment is the labour market, where both labour demand and labour supply should be carefully scrutinized. Assessment can be used to identify the needs of youth, which will inform

the decisions on the appropriate intervention types and target groups.

At the outset, it is important to understand that the functioning of the labour market is heavily influenced by wider economic factors, including institutions and macroeconomic conditions. Importantly, the opportunities for and returns to productive employment depend to a large extent on the ability of the economy to generate employment opportunities. However, the efficiency with which economic growth translates into productive employment can vary widely and economic growth can be either more or less job-rich, with consequent negative or positive effects on young people.

TIP



International institutions like the ILO have the capacity to undertake detailed and comprehensive assessments before launching nationwide interventions. For smaller entities, non-governmental organizations and independent research teams, or for smaller teams within international organizations, such an assessment can present challenges in terms of time and financial costs. For such organizations, full labour market diagnostics are not recommended. Instead, they can utilize the data provided by existing analyses carried out by research institutions or international organizations, including national labour market studies and youth reports.

Box 1.1: Examples of youth-focused employment diagnostics

Somalia: Youth employment and livelihood survey on skills and market opportunities

“Youth Employment and Livelihood Survey on Skills and Market Opportunities” is a report prepared by ILO and UNICEF to support the sustainable economic reintegration of children associated with armed forces and armed groups (CAAFAG) in South Central Somalia. The purpose of the study is to provide information to youth on promising opportunities for starting businesses and obtaining jobs in Mogadishu, and advise them on the skills needed to do so (Forcier et al, 2013).

Selecting the road to more and better jobs: Sector selection report of the Road to Jobs project in Northern Afghanistan

Road to Jobs (R2J) is a three-year ILO project that aims to create more and better jobs in the provinces of Samangan and Balkh in Northern Afghanistan. The project follows a market systems approach to address underlying constraints inhibiting better growth and employment outcomes, which in turn contribute to improving livelihoods and poverty reduction. The report documents the sector selection process the project team implemented during the project design phase. The selection process involved three main elements: (1) Participatory Appraisal of Competitive Advantage (PACA®) (2) Rapid Market Assessments (RMA) and (3) internal assessment of the findings (ILOa, 2015).

Bangladesh: Looking Beyond Garments, Employment Diagnostic Study

This Employment Diagnostic Study, a collaboration between ILO and the Asian Development Bank, highlights key labour market trends and challenges in Bangladesh, analyzes in depth the major issues relating to employment, and makes recommendations for government and stakeholder consideration. The study team undertook a series of consultations with key stakeholders, including government, employers’ and workers’ organizations, leading academicians, and development partners on the study framework, approach, and findings. Through the consultations a number of key employment challenges were identified and later became the thematic areas the study covered, including (i) diversification of the economy, (ii) skills development, (iii) women at work, and (iv) overseas employment (ILO, ADB, 2015).

STAKEHOLDER CONSULTATION

One important element during the diagnostic phase, regardless of the depth of analysis, is wide consultation with stakeholder groups. A central component of this process should be social dialogue, consulting with trade unions and employers' organizations involved in the policy process. The breadth of these consultations and the fact that employers' and workers' organizations together represent the interests of a sizeable fraction of the population can generate sustainable and readily accepted policies and interventions. The inclusion of civil society actors can also further their legitimacy and effectiveness.

Listening to the youth voice is essential. It is important to bear in mind that the views

provided by those organizations described above – particularly employers' and workers' organizations – reflect the perspectives of economic actors who are already firmly established in the economy. As prospective entrants, youth are struggling to transition into the economy, to identify new employment opportunities or to establish themselves in quality employment. As such, their interests and needs are not necessarily reflected in the views expressed by older generations, who are often represented by such organizations. Engaging youth perspectives is an important factor in ensuring that programmes are aligned with the needs and expectations of young people.

LABOUR MARKET STATUS OF YOUNG PEOPLE

While it is beyond the scope of this note to give comprehensive guidance on how to determine the status of young people in the labour market, it is helpful for youth employment practitioners to understand a few basic concepts. Further guidance on labour market indicators, as a basis for monitoring project outcomes in the results chain, can be found in Note 2.

Youth policy and programmes often focus on youth unemployment. For young people in low- and middle-income countries, the problems of underemployment, vulnerable employment and working poverty present a much greater challenge than unemployment itself. Many young people simply cannot afford not to work and often end up working in low-quality or hazardous forms of employment. Therefore, an assessment of youth employment status should concentrate on dimensions of employment associated with poverty and vulnerability in addition to unemployment.

The working-age population of a given country can be categorized according to whether individuals are in or out of the labour force, time use, employment status, job quality and the degree of formality of employment. Underemployment describes non-voluntary, part-time work and informal employment captures, for example, contributing family workers or own-account workers with no formal work arrangements and no access to benefits or social protection programmes. As informal employment relations are sometimes conducted under precarious conditions, strong connections between informal employment and vulnerability exist, with informal workers at greater risk from the effect of economic cycles. Also, vulnerably employed people usually make up a significant proportion of the working poor. These are defined as employed persons, either working for wages, on their own account or as unpaid family helpers, whose income is insufficient to bring them and their dependants over a defined poverty threshold (ILO, 2012a).

Further dimensions of employment problems can be found by looking more closely at the characteristics of people who are out of the labour force. While not being actively engaged in the labour force can be a choice, young people and women often remain outside the labour force because of perceptions that there are no jobs available to them or because they are discouraged in their efforts to secure work. Because their status of inactivity or engaging solely in unpaid work in the home is the result of having given up the search for employment rather than a voluntary decision, this status should be considered as disguised unemployment, even though differentiating

disguised unemployment from voluntary inactivity can be difficult.

The NEET rate (which captures people not in employment, education or training), differentiated by youth and gender, can also highlight needs to be taken into account. It can serve as a narrower proxy of marginalization than labour market inactivity. Moreover, it reflects a status in which potential workers are experiencing skills degradation and erosion. By not being actively engaged in work or preparation for work, these individuals are most at risk of not making the transition to work successfully without targeted support.

FIGURE 1.1 AN OVERVIEW OF YOUTH LABOUR MARKET INDICATORS

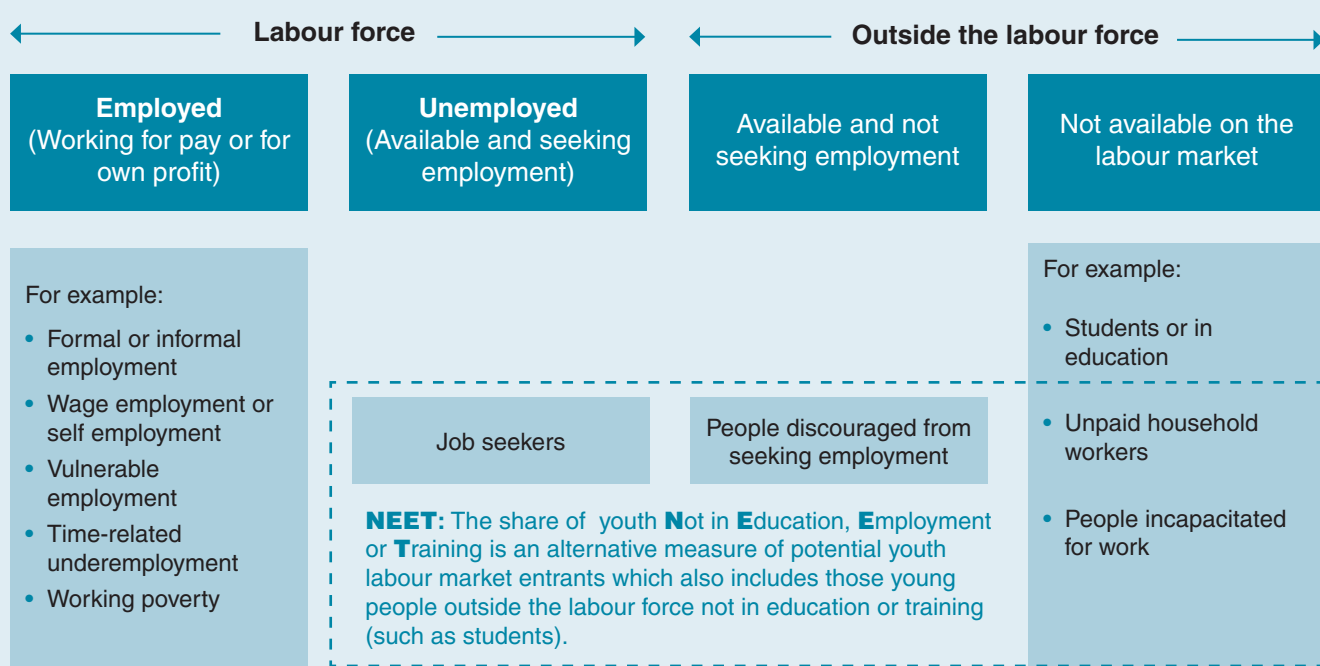


Table 1.1: Descriptions of labour market indicators for young people

Employment	Persons in employment are defined as all those of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit. They comprise employed persons "at work", i.e. who worked in a job for at least one hour; and employed persons "not at work" due to temporary absence from a job, or to working-time arrangements (such as shift work, flexitime and compensatory leave for overtime).
Unemployment	Persons in unemployment are defined as all those of working age who were not in employment, carried out activities to seek employment during a specified recent period and were currently available to take up employment given a job opportunity.
Labour Force	The labour force comprises all persons of working age who furnish the supply of labour for the production of goods and services during a specified time-reference period. It refers to the sum of all persons of working age who are employed and those who are unemployed.
Underemployment	Persons in time-related underemployment are defined as all persons in employment who, during a short reference period, wanted to work additional hours, whose working time in all jobs was less than a specified hours threshold, and who were available to work additional hours given an opportunity for more work.
Vulnerable employment	Vulnerable employment comprises own-account workers and contributing family workers. Workers in these forms of employment are less likely to have formal work arrangements, and are therefore more likely to lack decent working conditions, adequate social security provisions and "voice" through effective representation by trade unions.
Working poverty	The working poor are persons who, in spite of being employed, still live in a household classified as poor (i.e. household that has income or consumption levels below the poverty line).
Informal employment	Informal employment comprises persons who in their main job were: (a) own-account workers, employers or members of producers cooperatives employed in their own informal sector enterprises; (b) own-account workers engaged in the production of goods exclusively for own final use by their household; (c) contributing family workers, irrespective of whether they work in formal or informal sector enterprises; or (d) employees holding informal jobs, whether employed by formal sector enterprises, informal sector enterprises, or as paid domestic workers by households.
NEET	The share of youth not in education, employment or training (NEET rate) conveys the number of young persons not in education, employment or training as a percentage of the total youth population. provides a measure of youth who are outside the educational system, not in training and not in employment, and thus serves as a broader measure of potential youth labour market entrants than youth unemployment, since it also includes young persons outside the labour force not in education or training. This indicator is also a better measure of the current universe of potential youth labour market entrants compared to the youth inactivity rate, as the latter includes those youth who are not in the labour force and are in education, and thus cannot be considered currently available for work.

Source: www.ilo.org/ilostat/faces/ilostat-home/metadata

In diagnosing the employment situation of young people, using adequate data sources is crucial. Consulting the latest national labour force survey (LFS) is a first step to assessing the youth employment situation. Labour force surveys cover core labour force variables, such as current activity status (employed, unemployed, not economically active), hours of work, wages, etc. Useful indicators based on LFS results can often be found on the websites of national statistics agencies and through the ILO.¹

However, if the intention is to assess the situation of young people specifically, labour force surveys are often too general. They do not provide substantial information on the transition to work. To address this issue, the ILO has developed a diagnostic survey methodology called the school-to-work transition survey (SWTS, see box 1.2 for details) to collect detailed information on the labour market situation of young people as they leave the education system, their job search and work experience over time, as well as a number of other relevant factors.

¹ See ILOSTAT (available at: <http://www.ilo.org/ilostat>) Microdata sets for many labour force surveys can be found in the online ILO's Microdata Repository (available at: <http://www.ilo.org/surveydata/>).

TIP



A simple analytical framework that can be used by practitioners to examine youth employment data could comprise:

- comparison of youth (aged 15–24 or 15–29 depending on the defined age range) indicators against the labour market figures for adults (aged 25–64 or 30–64)
- examination of core labour force variables for different groups of young people: young adults, young men and young women, rural and urban youth, ethnic majorities and minorities, individuals with low and high educational attainment
- assessment of performance of youth employment indicators over time and benchmarking of youth indicators against those of neighbouring countries and/or regional and world aggregates
- determination of productive employment of youth: counting young people in both unemployment and working poverty relative to the employed gives an estimate of the extent of the productive employment deficit, where both underemployment and low wages can be sources of working poverty.

Box 1.2: School-to-work transition surveys

The SWTS is a survey instrument that generates relevant labour market information on young people aged 15 to 29 years old, including longitudinal information on transitions within the labour market. The SWTS allows for an exploration of the increasingly indirect paths to decent and productive employment that today's young men and women are facing.

The decisive advantage of the SWTS data is that it applies a stricter definition of “stable employment” than is typically used in the genre. By starting from the premise that a person has not “transited” until settled in a job that meets basic criteria of “decency” – namely a position that provides the worker with a sense of security (e.g. a permanent contract) or a job with which the worker feels personally satisfied – the ILO is introducing a new qualitative element to the standard definition of labour market transition.

In the context of diagnosing barriers to youth employment, this qualitative factor is particularly helpful because it overcomes the binary distinction between unemployment and employment. Decent work deficits can be identified, so that interventions can be designed to address these problems directly.

Between 2012 and 2016, the SWTS was deployed in more than 30 countries. A full list of the countries covered and access to the microdata files can be found online at: <http://www.ilo.org/w4y>

LABOUR DEMAND AND SUPPLY

Labour demand and labour supply each play a role in creating deficits to productive employment for youth. Both aspects should be analysed in order to identify the main constraints to youth employment.

Labour demand

To assess the level of demand for youth labour, the two sources to consider are: the private sector, including self-employment, as well as the public sector.

Private sector and self-employment: Elements to consider in assessing labour demand in the private sector include business and labour regulations, capital and access to finance, the state of infrastructure, including information technology (IT) as well as other features of the enabling environment for sustainable enterprise (see box 1.3). It is also important to understand investment policy, which

TIP



For a detailed analysis of the labour market, the ILO guide *Understanding deficits of productive employment and setting targets* (ILO, 2012b) offers step-by-step guidance on calculating past and present deficits in productive employment.

presents opportunities for business expansion through export opportunities and tax incentives. When looking at small and medium-sized enterprises (SMEs) as a source of employment, the availability of affordable finance requires careful analysis. Although capital may be available for larger businesses, the lending terms governing collaterals and other securities can be challenging for smaller enterprises, limiting their growth prospects and therefore their ability to generate new jobs. The relative size of

Box 1.3: Enabling environment for sustainable enterprises (ESEE)

The ILO's ESEE tool aims to provide an enabling environment for sustainable enterprises, including an assessment of bureaucratic red tape and the procedural difficulties associated with enterprise formalization. Getting the enabling environment right is of key importance as there is limited value in promoting enterprise development for the creation of more and better jobs in an environment that is hostile to firm start-ups. This is especially relevant for SMEs, which are hit harder by the costs associated with burdensome procedures and regulations than larger enterprises, which are better placed to absorb these costs. To stimulate SME development and, in turn, job creation, it is essential to level the playing field between SMEs and larger firms.

ESEE is a comprehensive methodology developed to assess, advocate and reform the environment in which enterprises start up and grow. ESEE:

- (1) allows stakeholders to identify the major constraints hampering business development
- (2) fosters dialogue between workers, employers and government to reach shared policy recommendations
- (3) supports the adoption of effective reforms
- (4) unlocks entrepreneurial potential and boosts investments to generate overall economic growth, create better jobs and reduce poverty.

For more information see www.ilo.org/eese.

the formal and informal sectors must also be taken into account, which growth sectors or industries and geographic areas have the most potential to boost demand for labour, industry trends and projections and the number of jobs expected to be created in each area.

Public sector: The public sector differs from the private sector in terms of wage levels and wage compression, employment protection, social security and unionization. Wages in the public sector tend to be lower, and the spread of wages usually narrower than in the private sector (meaning that workers have fewer chances to increase their income over time). However, public sector jobs often provide high levels of non-wage benefits, including more attractive social security plans and greater employment security. Pay schemes and wage structures are also more transparent and predictable within the public sector. These factors combine to make the public sector potentially more attractive than private sector work, despite lower wage levels.



TIP

Establishment surveys: Data from establishment surveys (also known as firm surveys) are more appropriate for the analysis of the demand side of labour than data from labour force surveys. These surveys focus on the characteristics and operations of businesses. Specific information in establishment surveys includes, for instance, data on production, exports and imports, labour costs, employment and average earnings (by occupation and skill level), vacancies, hiring and firing practices and future employment prospects.

Depending on local context, public sector employment can have different impacts on employment outcomes for youth. On the one hand, benefits offered in the public sector might lead to higher job quality overall, from decent working times to occupational safety, as private employers have to compete for qualified workers.

Box 1.4: Diagnosing youth employment using a value chain approach

To foster the creation of decent jobs, it is important to acknowledge that labour markets are not uniform across all sectors. Growth and investment patterns across industries offer different opportunities for young people to gain productive employment. The ILO guide *Value chain development for decent work (2015)* provides an overview of criteria for selecting the sectors with the greatest potential for decent job creation, and a framework for analysing sectoral constraints to inclusive growth.

Before designing an intervention it is crucial to identify the sector(s) in which the greatest impact can be achieved. Three criteria can be used:

- (1) potential relevance of the sector to youth target groups
- (2) relative importance of the sector to help youth “step up” (improve their productivity or job quality), “step in” (access new jobs or enter new markets) or “hang in” (maintain their existing livelihood and income streams)
- (3) feasibility of stimulating change, given the time and resources available to the intervention, and the level of ambition.

A value chain analysis can help to identify specific underperforming “functions” (supporting services, such as access to information, technology or finance) and “rules of the game” (regulations, laws and informal norms). Interventions can then be designed to help overcome the public and private market failures and improve employment outcomes for youth.

Further information on the approach can be found in [Nutz and Sievers \(2015\)](#).

On the other hand, public sector employment can result in the private sector being crowded out of the labour market, thus preventing economic expansion. The lure of secure public sector work also plays a role in signalling the skills in which youth should invest: where the attraction is too strong, youth may make educational investments that aim to secure public sector jobs rather than acquiring the skills sought by private employers.

Labour supply

Labour supply can be defined as the attributes that determine young people's employability – education, skills, health, cognitive abilities, etc. – and their ability to access productive employment. An analysis of labour supply should examine the three aspects of skills, geographical distribution and the social context of youth livelihoods.

Skills: It is essential to consider the distribution of education and skill levels in the total

population and the labour force, as well as making relative comparisons between young people and adults. A baseline assessment can include the share of the population that has completed primary, secondary and tertiary education, disaggregated by age and gender. In applying educational data to population or labour force categories, it is vital to consider both rates and distributions. For example, unemployment rates might be highest among university graduates, but the majority of the unemployed might be secondary graduates if this population is significantly higher than the population of university graduates. An important aspect of this analysis can be enrolment versus completion rates in education, as these two rates can differ sharply depending on drop-out rates. Also, the level of public expenditure and the calibre of teaching personnel can provide insights into the quality of education.

In many countries, technical and vocational education and training (TVET) systems are an important alternative to secondary and tertiary

Box 1.5: Top websites for business and macroeconomic data



The Atlas of Economic Complexity: A powerful interactive tool that enables users to visualize a country's total trade, track how these dynamics change over time, and explore growth opportunities for more than 100 countries.



IMF data: A key source of information on major economic indicators for every country in the world. The site includes data on debt and capital flows, as well as qualitative information about the key political and economic developments.



World Bank Doing Business database: An easy-to-use source of information on a broad range of business environment indicators, as well as excellent comparable data on the actual costs of trade. The site includes tools that allow users to evaluate how changes in particular regulations could improve the overall business environment.



World Bank's World Development Indicators (WDI): The World Bank's central database, the WDI includes hundreds of indicators, from the number of mobile phones per 1,000 people to the number of children out of school. The WDI provides socio-economic data from a range of national and international sources reported both at the national level and as regional and international aggregates.

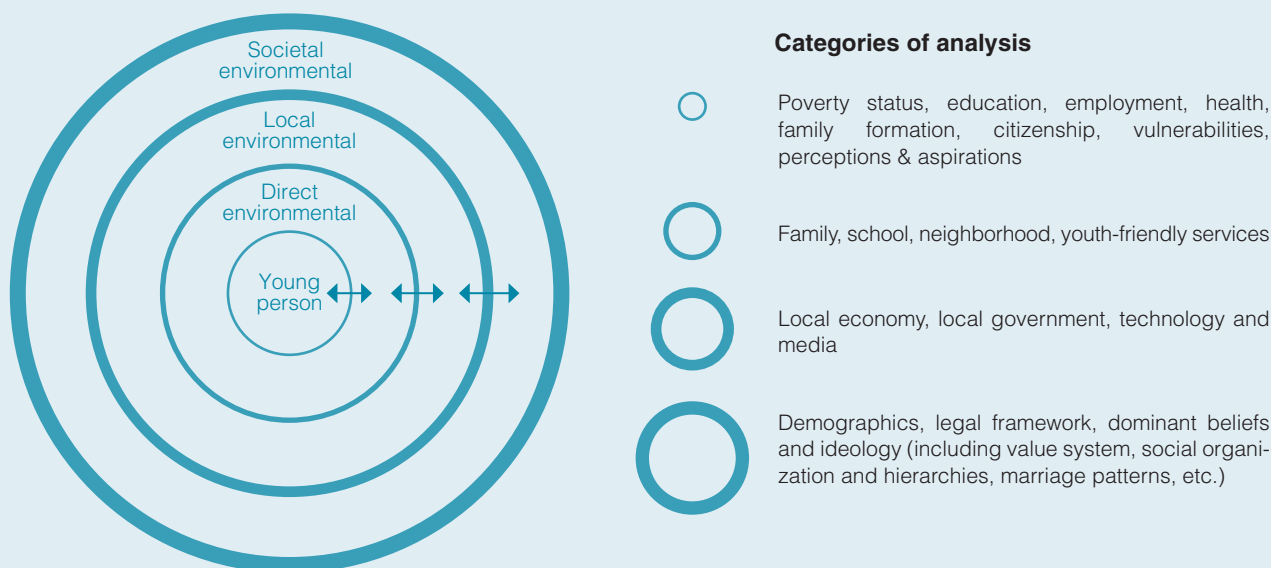
education and therefore the quality and prevalence of TVET should be assessed. TVET systems require special attention due to the different organizational forms that they can take across countries. While formalized and school-based in some countries, TVET education can be almost indistinguishable from a normal work situation in others (i.e., apprenticeships and internships). Considering this aspect is crucial in determining how TVET systems affect the skill level of the population and therefore the structure of the country's labour supply.

If the skill sets of potential workers do not match the economic structure and demand for labour, unemployment in some sectors can run parallel to a large number of vacancies in other sectors. This qualification-driven gap between labour supply and demand is addressed primarily through educational and institutional reforms, which ensure that labour market signals of skills demand within the private sector reach students, teachers, education policy-makers and parents. As in the case of public sector employment, detailed above,

existing institutional norms can fundamentally affect this balance, leading to skewed outcomes in terms of skills alignment.

Geographical distribution: If work is available mainly in the urban centres of a country but excess labour supply is still confined to rural areas, unemployment and vacancies, even on the same skill level, can coexist. Often, such imbalances create incentives for internal migration. Although migration can help to address the geographical mismatch of labour demand and labour supply, it can also create challenges and vulnerabilities. This is especially true for young people, who make up the majority of economic migrants. Although youth face fewer obstacles to migrating than established adults, migration does come at a cost. Without effective safety nets and family support, relocating to the city can be a disruptive move for young people, and one that does not always guarantee access to employment opportunities or a brighter economic future.

FIGURE 1.2: YOUTH ENVIRONMENTS



Source: Hempel and Fiala (2011).

The social context of youth livelihoods:

The potential social dislocation of migrating youth underlines the importance of including the broader social context within a comprehensive assessment of the labour market. Since young people's labour market transitions are influenced by a wide range of factors, including family, peers, community, local and national institutions and social norms, an effective employment assessment should also analyse the direct, local and societal environments in which young people live (see Bronfenbrenner, 1979). This includes issues such as health (physical and mental), family formation and citizenship, as well as issues that are of particular importance to sub-groups within the youth demographic (persons with disabilities, young women, marginalized communities, etc.). A holistic assessment will provide a rich picture of the needs and challenges that youth are facing and will therefore allow interventions to be adapted to fit local realities more precisely (see figure 1.2).

TIP**Assessing the potential for creation of decent work**

An assessment of the needs of young people and the barriers facing them should not only determine whether they have access to employment but also whether the employment obtained constitutes decent work. Decent work is defined as “productive work for women and men in conditions of freedom, equity, security and human dignity” (ILO, 1999). Decent work is a concept with multiple dimensions, the indicators used to measure it will be discussed in more detail in Note 2.

Box 1.6: Websites for labour supply data

ILOSTAT: Labour market information database managed by the ILO.



UNESCO Global Education Digest: A detailed source of internationally comparable data on education, science, culture and communication.



UNDP Human Development Index: This index brings together gross national income (GNI) per capita, life expectancy and high-level education indicators into a useful tool for international comparisons.



World Bank STEP Skills Measurement Program: This is the first initiative to measure skills in low- and middle-income countries (<http://microdata.worldbank.org/index.php/catalog/step/>).



Programme for the International Assessment of Adult Competencies (PIAAC): This instrument measures adults' proficiency in key information-processing skills: literacy, numeracy and problem solving.



Education GPS: Internationally comparable data and analysis on education policies and practices, opportunities and outcomes including the Programme for International Student Assessment (PISA).

In looking at young people's broader environment, it may become apparent that the issue of limited job opportunities is only one among many challenges that young people in a specific location are facing, which may suggest ways to build or adapt our intervention so that it can address more than one issue. This approach can help to assess whether there are, for example, legal or social barriers facing particular groups of youth, based on ethnicity, gender or sexual orientation, excluding them from (segments of) the labour market.

Equally importantly, the assessment will help us to specify our target group and ensure that our approach is more closely aligned with this targeting. Are we interested in all youth, or only those who are out of school? What age range do we want to focus on? Are there gender or ethnic considerations we would like to prioritize? Which geographic areas will we target? Given our resource constraints, we are rarely able to serve every young person.

Box 1.7: Market assessment for decent jobs in Afghanistan

Road to Jobs (R2J) is a three-year project that aims to create more and better jobs in the provinces Samangan and Balkh in Northern Afghanistan. The project follows a market systems approach to address important underlying constraints that are inhibiting better growth and employment outcomes, which would otherwise contribute to improving livelihoods and poverty reduction.

R2J positioned itself for sustainable jobs impact by conducting a thorough sector selection process. This was to target those sectors that are labour intensive and have a high reliance on wage labour inputs. A set of six rapid market assessments (RMAs) were conducted to give programme developers a "first look" at high-potential markets to determine their likely relevance to the target groups and capacity for positive employment change.

The RMAs scored and ranked potential sectors based on standardized sector selection criteria:

- (a) Relevance to target groups
 - Criterion 1: Number of target group members active in the sector
 - Criterion 2: Nature of the target group's participation in the sector
 - Criterion 3: Decent work deficits faced by target groups in the sector
- (b) Opportunity for growth
 - Criterion 4: Likelihood of sector growth
 - Criterion 5: Scope for improving target group employment in the sector
- (c) Feasibility of stimulating market system change
 - Criterion 6: Capacity of market actors
 - Criterion 7: Willingness of market actors to change
 - Criterion 8: Likelihood of distortion

Based on this analysis, R2J selected an initial set of three sectors in which to concentrate programmatic efforts: cotton, poultry and grapes and raisins.

The full set of tips, tricks and tools for value chain selection – drawing on the R2J experience – can be found at: http://www.ilo.org/empent/Projects/the-lab/WCMS_563289/lang--en/index.htm

Objectives and design: What do we want to achieve and how can we best achieve it?

Once the barriers facing young people and the opportunities for creating decent work have been identified, we can begin to formulate **programme objectives** and select possible strategies. Clearly defining what we want to

achieve will help us identify the desired results of our programme, communicate with donors and stakeholders, manage the interventions and monitor and evaluate our work.

STRATEGIC AND INSTITUTIONAL FRAMEWORK

In all project formulation processes, the high-level strategic and institutional framework has to be considered. By strategic and institutional framework, we mean an organization's position in relation to global, national and local strategies concerning youth employment. There are three main development frameworks that must be considered when formulating youth employment interventions: Sustainable Development Goals, development cooperation frameworks and ILO Decent Work Country Programmes.

Sustainable Development Goals (SDGs):

The SDGs set out objectives for development community actors to achieve collectively by 2030. The 2030 Agenda for Sustainable Development has people and the planet as its primary focus and gives the international community the impetus it needs to work together to tackle the formidable challenges

confronting humanity, including the decent work deficits experienced by many young people.

The 2030 Agenda places decent work for all at the heart of policies for sustainable and inclusive growth and development. Goal 8, “Decent work and economic growth,” calls for sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all, and sets as Target 8.6, “By 2020, substantially reduce the proportion of youth not in employment, education or training”.

Approaches to address these decent work deficits in youth employment include:

- ▶ developing strategies to promote youth employment that balance an integrated strategy for growth and job creation with

DEFINITION

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.

targeted interventions, such as job-search assistance or measures to support young entrepreneurs

- ▶ addressing skills mismatch by ensuring that training programmes meet labour market needs and by introducing work experience components in technical and vocational education and training (TVET)
- ▶ investing in innovative forms of social protection to improve income security for workers in vulnerable employment.

The 2030 Agenda has mobilized a range of actors at the global, national and local level to work collaboratively to achieve Goal 8. A multi-stakeholder partnership – the Global Initiative on Decent Jobs for Youth – has also been initiated as a United Nations system-wide response to the youth employment challenge (see box 1.8).

Development cooperation frameworks:

Youth employment strategies should also be anchored in national development strategies and national and international cooperation frameworks, such as the Poverty Reduction Strategy and the United Nations Development Assistance Framework (UNDAF). Recent UNDAF guidance places special emphasis on an integrated approach to programming that brings together the development,

humanitarian, human rights and peace and security pillars of the UN, building on a common strategy of effective and coherent implementation support for the 2030 Agenda. This is known under the acronym MAPS (Mainstreaming, Acceleration and Policy Support). MAPS frame the UN's engagement with national counterparts on the SDGs, including the support process for national SDG planning. Bearing these strategies in mind when planning an intervention will allow synergies to be leveraged.

ILO Decent Work Country Programmes:

Another crucial element of the institutional framework for young people's employment are the Decent Work Country Programmes (DWCPs). They constitute the primary framework for ILO support to constituents, including governments and employers' and workers' organizations, at the country level. This country programming mechanism, based on results-based programming, a solid diagnostic of the national labour market and close consultation with country stakeholders, needs to be taken into account when formulating youth employment strategies and interventions, and aligning them with the activities undertaken by other members of the development community.

Box 1.8: Global Initiative on Decent Jobs for Youth



In October 2014, the UN High-level Committee on Programmes selected youth employment as a prototype for an issue-based initiative that would mobilize the capacity of the UN and other global actors committed to more and better jobs for youth. Positive collaboration of numerous UN entities led to the development of the Global Initiative on Decent Jobs for Youth, which was endorsed by the UN system Chief Executives Board for Coordination and subsequently launched by the ILO in February 2016.

The Global Initiative is the overarching and inclusive platform for the promotion of youth employment within the 2030 Agenda for Sustainable Development, as well as the implementing arm for youth employment action within the new UN Strategy on Youth developed by the Inter-Agency Network on Youth Development.

The initiative is framed by four interconnected strategic elements:

- **Alliance** – A strategic and inclusive multi-stakeholder alliance carries out advocacy, ensures policy convergence, stimulates innovative thinking and mobilizes resources for more and better investments in youth employment. It comprises government agencies, social partners, the UN system and other multilateral organizations, the private sector, parliamentarians, youth and civil society, foundations, academia and the media.
- **Action** – The Global Initiative scales up evidence-based action at regional and country levels, ensuring ownership and coherence with national development priorities. With the commitment of governments, social partners, regional institutions and the support of UN Country Teams, the Global Initiative engages a diverse set of national and local partners on a range of themes that include: digital skills, quality apprenticeships, green jobs for youth, young people's transition to the formal economy, youth in the rural economy, youth in fragile situations, youth entrepreneurship and self-employment, and youth aged 15–17 in hazardous occupations.
- **Knowledge** – A global knowledge facility will capture, analyse and widely share best practices and innovation, enhance capacity development and facilitate peer learning on what works to improve labour market outcomes for both young women and men.
- **Resources** – The Global Initiative advocates for high-level commitment of local and international actors to increase resources through present and future funding facilities to enable scaling-up activities in support of decent jobs for both young women and men in the most inclusive and transparent manner.

Note: More information about the Global Initiative on Decent Jobs for Youth is available at www.ilo.org/decentjobsforyouth or <https://www.decentjobsforyouth.org/>

SETTING THE HIGHER-LEVEL GOAL

The higher-level goal (also known as a development objective) of a youth employment intervention is expressed as the result we want to achieve. It describes the ultimate objective to which the intervention aims to contribute. The goal captures two concepts: (i) the impact on end-beneficiary youth as a result of the intervention or the type of services they are able to offer following the programme, and (ii) the intervention's contribution to the higher-level development context. The latter is determined by policies at national or regional level and by the strategies of the implementing agencies. For instance, a typical higher-level goal for

youth employment projects implemented by the ILO is “more decent and productive work for all young people”.

A common mistake when defining a goal is to focus on what we will do, instead of what we intend to achieve (see point 1 in figure 1.3). If the ultimate reason for our intervention is to improve the living conditions of young people, then that should be reflected in our higher-level goal. The way we achieve this – for example by providing psychosocial support, training, seed capital or other services – is the “how to” and not the actual objective.

FIGURE 1.3: HOW TO DEVELOP HIGHER-LEVEL GOALS

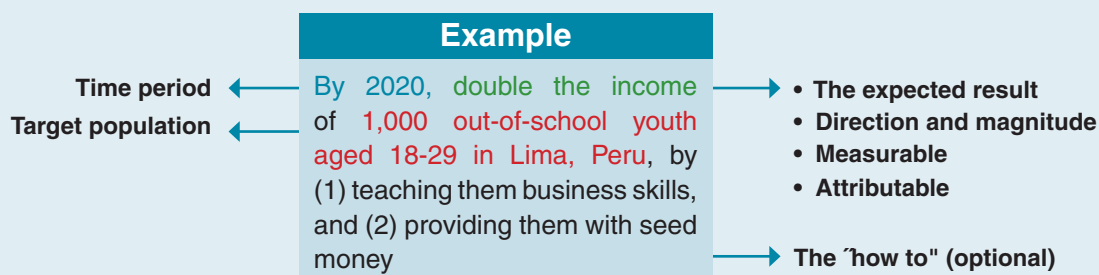


Based on the specific problems and constraints identified:

1. Specify the expected result – what we expect to achieve, not what we want to do.
2. Specify the *target population*.
3. Specify the *direction* and *magnitude* of the expected changes.
4. Specify the time period in which expected changes will occur.
5. Make sure the objective is *measurable*.
6. Make sure the objective can be attributable to the intervention.

Examples of higher-level goals in youth employment

- “By 2019, contribute to the promotion of more and better jobs for 1,000 young people between the ages of 18–29 in Serbia.”
- “To support national efforts to integrate policy priorities on youth employment and migration into national development and employment strategies in Albania by 2020’.”
- “To contribute to the easing of school-to-work transitions for 20,000 Bosnian and Herzegovinian young people.”



EFFECTIVE PARTICIPATION IN OBJECTIVE SETTING²

Programme design should be participatory in nature, as this approach creates local ownership for the initiative and at the same time enhances cooperation between different market actors.

Inputs into objective-setting can be generated using a number of research methods:

- ▶ interviews and consultations with market players, such as lead firms, government authorities, service providers, etc., but also selected smaller-scale market players whose position we are particularly interested in
- ▶ focus group discussions with groups of market players, which are easier to organize (e.g. small producers, SMEs or local service providers)
- ▶ observations of business practices and transactions (e.g. workplace conditions, labour productivity, etc.).

Once a programme has a draft objective in mind, a validation workshop can be organized, where stakeholders and end-beneficiaries can check whether the supply–demand balance is correct, and provide feedback on the programme scope. When planning and facilitating the workshop, it is important to create an environment in which participants feel comfortable identifying and articulating their concerns. However, a stakeholders’ workshop may not always be the most appropriate

option. Although the participatory approach increases ownership of the findings, there is the risk that increasing participation can result in a loss of analytical depth and/or a “dumbing down” of objectives.

A stakeholders’ workshop is not recommended when there is a risk that:

- ▶ actors do not possess the necessary knowledge and/or understanding of the market
- ▶ participants may become biased when tracing causality towards constraints in areas for which they are responsible, because they think that by adopting this approach will gain them project resources
- ▶ there is a highly charged political context
- ▶ discussions may focus on constraints that are beyond the scope of the project, which could either raise expectations that cannot be met, or create frustration among participants that their time has been wasted
- ▶ market actors regard youth employment, working conditions and gender equality as low-priority issues, so may not prioritize alleviating constraints that could lead to improvements in these areas
- ▶ market actors from some groups are not present, so those attending may not represent the interests of other groups in the market system.

SETTING OUTCOMES

The outcome (or immediate objective) describes the specific changes that a youth employment intervention is expected to bring about, by the end of its implementation period, to the quality and quantity of the services provided, and/or the way in which they are

delivered to the direct beneficiaries. It must be described in terms of a target to be achieved (result) rather than actions to be taken. The outcome is related to the problems identified during the labour market diagnostic phase and their cause–effect relationship. The problem analysis is mirrored in the objectives (i.e. the problem identified, which is the combined effect of a number of causes, is reversed and

² Adapted from *Value chain development for decent work* (ILO, 2015).

becomes the target to be achieved). For example, if the problem is a high share of youth in the informal economy, the outcome of the project becomes “Informality among young

workers reduced”. If the problem is lack of demand for youth labour, the outcome becomes “Labour demand for young workers increased”, and so on.

CHOOSING THE YOUTH POPULATION

When designing an intervention, we need to recognize that “one-size does not fit all”. Youth are different along age, gender, socio-economic status, location, skills, status (i.e. refugees), religious lines, among others. Some youth face distinct or more acute labour market disadvantages, such as disabled youth, that exacerbate their challenges of finding good quality jobs.

For example, by simply assessing youths’ skills levels, one can determine the characteristics and the extent to which youth are excluded from the labor market (see table 1.2). It is therefore critical to tailor the youth employment interventions to address the specific labor market challenges faced by different types of youth.

Table 1.2: Basic typology of youth skills level

Level of skill	Unskilled	Semi-skilled	Highly-skilled
Charateristics	<ul style="list-style-type: none"> • Temporary, informal and vulnerable jobs • Casual, low-productivity jobs • Minimal training, working in precarious conditions (domestic servants, gardeners, etc.) • Non-remunerated employment • Urban and peri-urban areas 	<ul style="list-style-type: none"> • Semi-educated/skilled youth (high school drop-outs) • Laid-off workers • Actively searching for stable or self employment • Urban areas 	<ul style="list-style-type: none"> • Enrolled in or graduated tertiary education, actively searching for full time employment or entrepreneurship • Urban areas
	<p>NEET</p> <ul style="list-style-type: none"> • NEET: Not in education, employment or training • Marginalized from society • Youth excluded from school system 		

DEFINING THE INTERVENTION

With a clear goal in mind, we can define the scope of an intervention (or set of multiple interventions, also known as a programme) that will result in achieving the stated objectives. Naturally, the choice of intervention should be driven by the specific barriers identified in the previous section; that is, we should choose an intervention that explicitly addresses the underlying causes hindering young people's abilities and opportunities to make a decent living for themselves and their families.

This guide focuses on interventions that promote youth employment and aim to improve access of young people to decent work. Therefore, guidance provided is targeted to active labour market programmes (ALMPs). ALMPs are interventions aimed at the employment of vulnerable groups in society. ALMPs tackle both supply-side and demand-side issues and can therefore offer solutions for the range of barriers to youth employment. We emphasize that good ALMPs are characterized as those that encourage and allow for social dialogue; collective bargaining mechanisms should be embedded within the programmes and should not be compromised by incentives provided by ALMPs (ILO, 2003, pp. 3–5).

Generally, ALMPs can be categorized into four types of programmes: skills training, entrepreneurship promotion, public employment programmes and employment services.

Skills training: Depending on the type of skill constraints faced by youth, four different types of skills trainings can be conducted: basic literacy and numeracy, trade- or job-specific skills, soft and life skills and business training (Kluve et al., 2017, p. 30). In general, this training does not aim to replace formal education but to fill knowledge and attitudinal gaps that prevent youth from transitioning from education to work. According to the

TIP



Intervention models and performance indicators of the ILO Small Enterprises Unit by Fiala and Pilgrim (2013) discusses possible types of entrepreneurship interventions with examples taken from ILO projects to illustrate each. This guidance focuses primarily on subsistence entrepreneurship interventions.

Youth Employment Inventory (YEI), this is the most common form of youth-focused ALMP.

Entrepreneurship promotion: These types of programmes have the twofold objective of generating work for the self-employed or business owners and creating further employment through job creation within the new businesses. There is a set of building blocks that are frequently used in entrepreneurship interventions: the first element is access to capital, which can be addressed by the provision of grants and loans either to individuals or through savings cooperatives. Second, business and management training often accompanies the financial part of the programme. Finally, mentoring and coaching for young entrepreneurs who have already set up their business help them to navigate challenges in the business world (Kluve et al., 2017, p. 35). Depending on the scope and scale of the programme, organizations should consider deeper, more sustained approaches to incubating start-ups and aiding new firms to leverage initial access to capital through efforts to secure outside investment – important factors in helping start-ups to become growth-oriented small businesses.

When discussing entrepreneurship programmes, it is important to distinguish between choice and necessity entrepreneurship. Those who are forced to become

self-employed because there are no other employment options available are also called subsistence entrepreneurs. Growth entrepreneurship, on the other hand, can have a more far-reaching employment and income effect and can boost innovation and competition in the private sector.

Public employment programmes: Public employment programmes can include public works programmes and government wage subsidies to support employment and on-the-job learning. Public works are a tool traditionally used by policy-makers to address a short-term mismatch of labour demand and supply in times of crisis. This instrument is recognized as a regular component of national employment strategies, especially when targeting particularly vulnerable groups (Lieuw-Kie-Song et al., 2010). Wage subsidies, on the other hand, are intended to foster longer term gains in private

sector job creation, lowering the barriers facing youth by reducing the costs associated with their employment. Youth can receive wage subsidies directly for employment to reduce the gap between offered wages and reservation wages. Wage subsidies can also be paid directly to employers, either as direct subsidies or indirectly through tax cuts. With wage subsidies, the aim is to lower the cost to the employer, making it financially feasible to employ young people, offer on-the-job training and keep them in their workforce after the end of the intervention. Wage subsidies paid directly to youth tend to be more effective when wage levels are a determining issue, while wage subsidies paid to employers are more effective when the focus is on subsidizing the costs of on-the-job learning and training.

Employment services: These programmes include job counselling, job placement

Box 1.9: Start and Improve Your Business (SIYB): An example of a youth entrepreneurship intervention

SIYB is a management training programme. Bearing in mind the need for more and better employment in developing and transitional economies, the focus of the programme is on starting and improving small businesses. The SIYB programme is a system of interrelated training packages and supporting materials for small-scale entrepreneurs. The programme is designed by the ILO and implemented with support from certified trainers in partner institutions in more than 100 countries with an estimated outreach of 6 million trainees. Initially developed in the 1980s, it has now been translated into more than 40 languages.

The Start Your Business (SYB) package provides a five-day training course for potential entrepreneurs with concrete and feasible business ideas and proposes a follow-up programme, including counselling sessions. SYB assists participants to develop a business plan with a marketing strategy, a staffing plan and a cost plan. Details of the training materials can be found at www.ilo.org/siyb.

The 2011 SIYB Global Tracer Study found that for each new business started after the training, on average, three jobs were generated. In Uganda, a randomized control trial (Fiala, 2015) providing mainly young business owners with loans, cash grants and the SYB training module, or a combination of these components, showed that, six and nine months after the interventions, men with access to loans and with business skills training reported 54 per cent greater profits relative to the comparison group.

For more information see [Majurin \(2014\)](#).

programmes and job-search assistance services. Job placement programmes acknowledge that there is a mismatch between labour demand and supply and provide young people with information about vacancies, and employers with information about eligible

unemployed young people. Job-search assistance services target primarily discouraged youth who are having difficulties connecting with existing opportunities in the labour market (ILO, 2003, pp. 6–8).

Table 1.3: Youth-focused ALMPs

Several studies of youth employment programmes have shown that some are successful while others fail to improve participants' chances of gaining a job. Some of the features of these programmes are summarized below.

Type of programme	Constraint addressed	Advantages	Disadvantages
Skills training	Deficits in labour supply and/or the quality of labour supply	Works better with broader vocational and employability skills that are in demand and includes work experience as well as employment services	May produce temporary, rather than sustainable solutions and, if not well targeted, may benefit those who are already “better off”. Training alone may not be sufficient to increase youth employment prospects
Employment services	Mismatch of labour supply and labour demand	Helps to link youth in search of employment to existing vacancies and job offers. If successful, this is a highly cost-efficient type of intervention. Can point youth towards desirable but not immediately obvious career paths	Might push youths towards jobs and occupations that do not match their original aspirations. Might lead to overly high or low expectations
Entrepreneurship promotion	Low business skills, high failure rate of new businesses	Can have high employment potential and may meet young people's aspirations (e.g. for flexibility, independence). More effective if combined with financial and other services, including mentoring	The jobs created are likely to have substantial decent work gaps. May have a high failure rate, which limits its capacity to create sustainable employment. Take-up is often difficult for disadvantaged youth, owing to their lack of networks, know-how and collateral
Public employment programmes	Employment subsidies: Low demand for youth labour from private sector/high costs of labour and training	Can create employment if targeting specific needs (e.g. to compensate for initial lower productivity and training level) and groups of disadvantaged young people	High deadweight losses and substitution effects (if not carefully targeted); employment may last only as long as the subsidy
	Employment intensive public works and community services: Low demand for youth labour/low levels of community investment	Help young people gain labour market attachment and, at the same time, improve physical and social infrastructure and the environment and enhance employability, if combined with training	Low capacity for labour market integration; young workers may become trapped in a carousel of public works programmes; often gender-biased; can result in displacement of private sector companies

Source: Based on the classification of interventions in Kluge et al. (2017).

Evidence-based programming

A crucial element in developing an intervention is reviewing the existing knowledge about various programme alternatives and their potential effectiveness. For example, to address business start-up constraints facing young people, we may want to implement a programme to promote youth enterprises. But what exactly should the intervention look like? Assume that we confirm that financial constraints are the major obstacle to starting a business. Should the programme provide grants or loans? Should it target younger or older youth? The less well educated or the better educated? And will financial support be enough, or should it be combined with other support services, such as training, mentoring and business development support?

TIP



The What Works in Youth Employment (www.wwinye.org) platform is a dynamic webpage that offers an accessible service allowing users to gain an understanding of the global evidence base for the effectiveness of youth employment interventions. The platform offers advice, videos featuring researchers and policy-makers and an evidence gap map by type of ALMP and labour market outcomes.

To answer these difficult questions, programme managers will benefit from looking at the existing evidence base. Very often, implementers tend to favour certain types of projects based on predispositions and prior experience. Yet, to develop innovative

Box 1.10: Sources of evidence on youth employment

Interventions to improve labour market outcomes of youth: Systematic review of training, entrepreneurship promotion, employment services and subsidized employment interventions (Kluve et al., 2017). Drawing on a pool of 1,114 records selected for screening, 107 youth employment interventions were analysed to systematically assess the employment, earnings and (for entrepreneurship interventions) business outcomes.

Do interventions targeted at micro-entrepreneurs and small and medium-sized firms create jobs? A systematic review of the evidence and alternative evaluation designs (Grimm and Paffhausen, 2015). Grimm and Paffhausen synthesize the existing evidence on the employment impact of 60 interventions targeted at micro, small and medium-size enterprises.

The Independent Evaluation Group (IEG, 2013) conducted an evaluation of all World Bank and IFC youth employment programmes between 2010 and 2011. The study included a systematic review of 38 impact evaluations. The study found that there was a lack of data on the impact of youth employment projects. The majority of World Bank projects do not disaggregate data by age, and the seven impact evaluations of World Bank projects that are specific to youth are not generalizable. The IEG called for a more evidence-based approach to youth employment operations.

Olenik, Fawcett and Boyson (2013) produced a state-of-the-field report on youth workforce development. The study was based on 54 publicly available studies published from 2001 to 2012, interviews with key informants from the field and staff experts from USAID, and a review of 23 major donor organizations in the field. The study reviewed the impacts on specific development outcomes, as well as the target populations most affected. The report also identified gaps in the existing evidence, such as the need for better information on cost-effectiveness of intervention components.

high-quality projects, it is important to consider the existing theoretical and empirical knowledge about youth employment programming. If the available evidence confirms our inclination, we can make a strong case for

a specific design. If, however, existing knowledge points to serious limitations of an intervention, then it will save time and money to incorporate the lessons learned into the new initiative.

Table 1.4: Example of a menu of evidence-based interventions, by constraint		
Constraint	Interventions with strong evidence	Interventions with mixed evidence
Skills development		
Insufficient basic skills	<ul style="list-style-type: none"> Information about the value of education 	<ul style="list-style-type: none"> “Second chance” education programmes
Technical skills mismatch	<ul style="list-style-type: none"> Training “plus”/comprehensive programmes Information on returns to technical specializations 	<ul style="list-style-type: none"> On-the-job training
Behavioural skills mismatch	n/a	<ul style="list-style-type: none"> Behavioural/life skills training
Signalling competencies	n/a	<ul style="list-style-type: none"> Skills certification Training centre accreditation
Entrepreneurship promotion		
Insufficient entrepreneurial skills	n/a	<ul style="list-style-type: none"> Entrepreneurial training
Lack of access to financial, natural or social capital	<ul style="list-style-type: none"> Comprehensive entrepreneurship programmes 	<ul style="list-style-type: none"> Microfinance
Employment services		
Employer discrimination	<ul style="list-style-type: none"> Affirmative action programmes 	<ul style="list-style-type: none"> Subsidies to employers who hire target groups Employee mentoring
Job matching	<ul style="list-style-type: none"> Employment services 	<ul style="list-style-type: none"> Technology-based information sharing
Public employment programmes		
Slow job-growth economy	<ul style="list-style-type: none"> Wage or training subsidies 	<ul style="list-style-type: none"> Public service programmes Labour-intensive public works
Excluded group constraints (ethnicity, gender etc.)	<ul style="list-style-type: none"> Target excluded group’s participation in programmes Non-traditional skills training Safe training/employment spaces for specific groups 	<ul style="list-style-type: none"> Adjusted programme content/design to account for excluded group’s specific needs

Source: Adapted from [Cunningham, Sanchez-Puerta and Wuermlı \(2010\)](#).

A thin or missing evidence base does not mean that a proposed intervention is doomed to failure. In fact, innovative approaches will, by definition, lack a track record. However, when we carry out interventions that lack a good evidence base, we should always be

aware of their probationary nature and not take positive results for granted. For these programmes, external learning – making findings accessible to other programme managers and practitioners – is particularly important.

KEY POINTS

1. **Start by understanding youth needs.** This requires an integrated supply-and-demand market assessment that captures the complex combination of social, economic and even environmental factors that influence young people's well-being and the constraints facing young people attempting to access decent jobs. Look beyond unemployment to consider underemployment, vulnerable employment and the working poor. Unless we know all the factors that limit youth opportunities, it will be impossible to design an intervention that truly addresses their needs.
2. **Anchor the project within the high-level institutional frameworks** – at the international level in the Sustainable Development Goals, and at the national level in the Decent Work Country Programmes. When formulating objectives, ensure that they are strategically aligned and technically robust. Be sure to consult widely among stakeholders, including intervention beneficiaries (young people).
3. **ALMPs can be categorized into four types of programmes: public employment, skills development, entrepreneurship promotion and employment services.** Select intervention combinations by reviewing the existing evidence about various programme alternatives and their potential effectiveness. If the available evidence confirms our choice, we can make a strong case for a specific design. If, however, existing knowledge points to an intervention having serious limitations, then it will save time and money to incorporate the lessons learned into our initiative.

KEY RESOURCES



- ▶ Kluge, J.; Puerto, S.; Robalino, D.; Romero, J.M.; Rother, F.; Stöterau, J.; Weidenkaff, F.; Witte, M. 2017. *Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services, and subsidized employment interventions*, Campbell Systematic Reviews 2017:12.



- ▶ ILO. 2012a. *Employment diagnostic analysis: A methodological guide* (Employment Sector, Geneva).



- ▶ Nutz, N.; Sievers, M. 2015. *A rough guide to value chain development: a short guide for development practitioners, government and private sector initiatives* (Geneva, International Labour Office).

REFERENCES

- ▶ Bronfenbrenner, U. 1979. *The ecology of human development: Experiments by nature and design* (Cambridge, MA, Harvard University Press).
- ▶ Cunningham, W.; Sanchez-Puerta, M.L.; Wuermli, A. 2010. *Active labor market programs for youth: A framework to guide youth employment interventions*, World Bank Employment Policy Primer, No. 16, Nov. (Washington, DC).
- ▶ Fiala, N.; Pilgrim, M. 2013. *Intervention models and performance indicators of the ILO Small Enterprises Unit* (Berlin and Geneva, DIW and ILO).
- ▶ Fiala, N. 2015. *Access to Finance and Enterprise growth: Evidence from an experiment in Uganda*, Employment Working Paper No. 190 (Geneva, International Labour Office).
- ▶ Forcier, N.; Adan, U.; Gutsell, A. 2013. *Youth employment and livelihood survey on skills and market opportunities* (Mogadishu, ILO).
- ▶ GIZ. 2014. *Guidelines for an Employment and Labour Market Analysis* (ELMA) (Bonn).
- ▶ Grimm, M.; Paffhausen, A. L. (2015). "Do interventions targeted at micro-entrepreneurs and small and medium-sized firms create jobs? A systematic review of the evidence for low and middle income countries", *Labour Economics*, 32, 67-85.
- ▶ Hempel, K.; Fiala, N., 2012. *Measuring success of youth livelihood interventions*.
- ▶ International Labour Organization (ILO). 1999. *Decent work*, Report of the Director-General. International Labour Conference, 87th Session (Geneva).
- ▶ ILO. 2003. *Active Labour Market Policies*, Proceedings of the 288th session of the Governing Body, Geneva, Nov., GB.288/ESP/2.
- ▶ ILO. 2012a. *Employment diagnostic analysis: A methodological guide* (Employment Sector, Geneva).
- ▶ ILO. 2012b. *Understanding deficits of productive employment and setting targets: A methodological guide* (Geneva).
- ▶ International Labour Organization and Asian Development Bank. 2015. *Bangladesh: Looking beyond garments*, Employment Diagnostic Study.
- ▶ ILO. 2015a. *Selecting the road to more and better jobs: Sector selection report of the Road to Jobs project in Northern Afghanistan* (Geneva).
- ▶ ILO. 2015b. *Value chain development for decent work: How to create employment and improve working conditions in targeted sectors*, 2nd edn (Geneva).
- ▶ Independent Evaluation Group (IEG). 2013. *World Bank and IFC Support for Youth Employment Programs* (Washington DC, Independent Evaluation Group, World Bank).
- ▶ Kluge, J.; Puerto, S.; Robalino, D.; Romero, J.M.; Rother, F.; Stöterau, J.; Weidenkaff, F.; Witte, M. 2017. *Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services, and subsidized employment interventions*, *Campbell Systematic Reviews* 2017:12.
- ▶ Lieu-Kie-Song, M; Philip, K; Tsukamoto, M.; Van Imschoot, M. 2010. *Towards the right to work: Innovations in Public Employment Programmes (IPEP)*, Employment Working Paper No. 69 (Geneva, International Labour Office).
- ▶ Majurin, E. 2014. *Start and improve your business (SIYB): Implementation guide* (Geneva, ILO).
- ▶ Nutz, N.; Sievers, M. 2015. *A rough guide to value chain development: a short guide for development practitioners, government and private sector initiatives* (Geneva, International Labour Office).
- ▶ Olenik, C.; Fawcett, C.; Boyson, J. 2013. *State of the field report: Examining the evidence in youth workforce development, Final report, USAID Youth Research, Evaluation, and Learning Project*.
- ▶ Olenik, C.; Fawcett, C.; Boyson, J. 2013. *State of the field report: Examining the evidence in youth workforce development, Final report, USAID Youth Research, Evaluation, and Learning Project*.



Case study:

LABOUR MARKET DIAGNOSTICS FOR THE PROMOTION OF RURAL YOUTH LIVELIHOODS IN ZAMBIA

Learning objectives

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

- ▶ apply appropriate assessment tools to identify the supply and demand constraints facing youth in the labour market
- ▶ understand the benefits of stakeholder consultation and the practical considerations that need to be taken into account when consulting stakeholders
- ▶ set programme objectives that are aligned with national strategic priorities.

Introduction and case study context

Rural Youth Enterprise for Food Security (or “Yapasa”) is a UN Joint Programme aiming to promote sustainable livelihoods for young women and men in rural areas of Zambia. The programme is implemented by the International Labour Organization (ILO) and the Food and Agriculture Organization (FAO)¹.

This case study focuses on the process that Yapasa followed to diagnose the youth

employment situation and design an intervention focused on needs. During a 12-month inception phase, Yapasa undertook a series of analyses and stakeholder consultations to strengthen programme design and arrive at a set of development interventions that are responsive to Zambia’s development challenges and priorities.

Part I: Do we understand youth needs?

Zambia has a relatively young population, with about 58.5 per cent aged between 15 and 35 years old. This demographic structure has a bearing on efforts to achieve the nation’s economic development objectives. This is reflected in the fact that, although Zambia has registered impressive economic growth trends in the recent past, these have not translated into significant gains in job creation, equality and poverty reduction. In rural Zambia, the poverty rate stands at around 80 per cent, compared to 34 per cent in urban areas, and the country’s income Gini coefficient is 0.65, which places

Zambia among the countries with the highest income inequality.²

Young people are severely affected by rural poverty. Registered youth unemployment in Zambia runs at 28 per cent in the 20–24-year-old age group and 16 per cent in the 25–29-year-old age group and, while data for rural areas is scarce, evidence indicates that the unemployment figures for rural areas are higher. The situation is aggravated by high levels of underemployment. Indications are that youth unemployment rates will continue to rise – 280,000 young women and men enter the labour market each year in search

¹ The programme has a budget of US\$7 million and a four-year implementation period, which started in 2013.

² See : UNDP. 2016. *Zambia Human Development Report 2016* (Lusaka).

of work. They are severely disadvantaged in the competition for the approximately 700,000 jobs in the formal economy, since they typically lack the skills, work experience and social networks of their older peers. As a result, many rural youth move to search for work in the capital city, Lusaka, where the few jobs available for youth are often underpaid and are performed in hazardous environments.

Within the labour market, there is perceived to be a mismatch between the skills needed to foster economic development and increased productivity, and the skills currently supplied by the tertiary education system. This has resulted in low levels of competency and employability and a weak entrepreneurial culture among young people, who are consequently unable to benefit from the country's economic growth.

The phenomenon of jobless growth in rural areas can be explained by a number of factors, prominent among them being the one-sided focus of investors on extractive industries, where financial returns are high, due to global demand for copper, but labour absorption rates are low. By direct comparison, in agriculture, the second main pillar of the Zambian rural economy where production methods are far more labour-intensive, economic activities have expanded at a much slower pace. While maize output has soared thanks to a combination of favourable weather conditions, input subsidy financing and guaranteed prices, commercial-scale production in other agricultural subsectors has picked up more slowly.

Yapasa wants to focus on creating decent jobs for youth in rural areas to support the Government in combating these trends.

Discussion topics



1. Which types of labour market status for young people should the Yapasa analysis focus on?
2. What diagnostic tools could help Yapasa to identify the “root causes” of the labour market constraints facing youth?
3. What type(s) of youth employment interventions do you think would be most effective in responding to the need?

Part II: How do we run stakeholder consultation?

Yapasa began with a 12-month programme inception phase. In addition to conducting analytical research, the objectives of the inception phase were to:

- ▶ put in place programme administration and implementation mechanisms
- ▶ conduct wider consultations in order to secure broader local programme ownership and alignment to key stakeholder priorities and development goals, and

- ▶ secure stakeholder consensus in the selection of the agricultural value chains for programme intervention.

The inception phase was designed to allow sufficient time to consult all stakeholders who have a bearing on the programme.

Yapasa formulated a consultative process, consisting of:

- ▶ an inception workshop to introduce the programme and raise awareness about its existence among stakeholders
- ▶ a high-level programme launch by the ILO Director General to raise the profile of the programme among the social partners
- ▶ a consultative sector selection workshop to seek stakeholder consensus in selecting the programme's target sectors
- ▶ a value chain development training course to orient stakeholders on the value chain development approach to youth entrepreneurship
- ▶ a programme review workshop to share with stakeholders the programme objectives and the plan for the remaining implementation period.



Discussion topics

1. Who are the main groups of stakeholders that Yapasa needs to engage with?
2. What do you think the stakeholder priorities will be for the programme? Do you think stakeholders will have similar or different expectations of the programme?
3. What would you change about Yapasa's consultative process?

Part III: Can we formulate a development objective in line with strategic priorities?

Yapasa is a contribution to Zambia's long-term objective, as articulated in the National Vision 2030, of becoming a "prosperous middle-income country by the year 2030". Programme results will contribute directly to the attainment of the Revised Sixth National Development Plan (R-SNDP) objective of "Promoting employment and job creation and rural development, through targeted and strategic investments in sectors such as science and technology, agriculture and energy

development". It will also contribute to the Government's vision, as elaborated in the National Industrialization and Job Creation Strategy document, which aims to create a total of one million new jobs.

Yapasa needs to define a development objective for the programmes which is consistent with these national priorities, as well as the priorities of its stakeholder groups:

#	Government ministry	National policy/ strategy	Goal/strategic objective
1	Ministry of Commerce, Trade and Industry	Commerce, Trade and Industry Policy	To support the effective development and utilization of domestic productive capacities as a means of increasing output and expanding employment opportunities
		Micro, Small and Medium Enterprise (MSME) Development Policy	To facilitate the creation and development of viable MSMEs that will contribute 30 per cent towards annual employment and 20 per cent towards GDP by 2018
		Industrialization and Job Creation Strategy	To create 1,000,000 jobs, targeting the unemployed, unpaid family workers and underemployed
2	Ministry of Agriculture and Livestock	National Agriculture Investment Plan	To increase fish production, productivity and value addition through sustainable and efficient management of aquaculture
			To increase sustainable crop production, productivity and value addition for a diversified range of competitive crops in addition to maize
3	Ministry of Youth and Sport	National Action Plan for Youth Empowerment and Employment	Boosting employment opportunities for young women and men
4	Ministry of Labour and Social Security	Decent Work Country Programme	More and better employment opportunities created, focusing on targeted groups
5	Citizens' Economic Empowerment Commission	2013–2016 Implementation Strategy	Value chain cluster development

At the institutional level, Yapasa also needs to contribute to the achievement of UNDAF Outcome 2 – “Targeted populations in rural and urban areas attain sustainable livelihoods by 2015” and, more specifically, Country Programme Outcome 2.2 – “Government and Partners provide targeted groups (including youths) with opportunities for gainful and decent employment by 2015”. The programme

will also contribute directly to UNDAF Outcome 2, on Sustainable Livelihoods and Food Security and, in particular, to Country Programme Outcome 2.2, on Decent employment – aimed at promoting gainful and decent employment and income generation with a focus on MSMEs, particularly women and youth, and promoting, adapting and implementing a rights-based approach to employment.



Discussion topics

1. Is Yapasa missing any key strategic or institutional frameworks that they need to align with?
2. Based on the above, can you formulate a suitable development objective for the programme?

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Note

2

Concepts and definitions of employment indicators relevant to young people



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

NOTE 2
**Concepts and definitions of employment indicators
relevant to young people**

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. 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 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 2: Concepts and definitions of employment indicators relevant for young people / International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131658-9 (print)
978-92-2-131659-6 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلم في برامج سوق العمل النشطة - الملاحظة 2: مفاهيم وتعريفات مؤشرات العمالة المرتبطة بالشباب, ISBN 978-92-2-630789-6 (print), 978-92-2-630790-2 (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.

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

Printed in Switzerland

Contents

Concepts and definitions of employment indicators relevant to young people	1
A menu of youth employment indicators	2
Employment opportunities	7
Employment quality	10
Employment access	13
Employment skills	16
Measuring basic skills	17
Measuring technical skills	18
Measuring core skills	20
Key points	24
Key resources	24
References	25
Case study: Selecting indicators for the northern Uganda Youth Entrepreneurship Programme	27
Figure	
2.1: Overview of employment indicators relevant for youth employment interventions	4
Tables	
2.1: A menu of labour market indicators relevant to young people	5
2.2: Indicators for employment opportunities	8
2.3 Indicators for employment quality	12
2.4: Indicators for employment access	15
2.5: Indicators for basic skills	17
2.6: Indicators for technical skills	19
2.7: Indicators for core skills	21





Concepts and definitions of employment indicators relevant to young people



Prerequisites:

This chapter requires no prior knowledge. It helps readers to measure outcomes by providing a “menu” of indicators to select from, depending on the objectives and design of the specific youth employment intervention.



Learning objectives:

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

- ▶ understand the challenges involved in measuring higher-level goals, such as jobs, employment and income, and the different options available to define and collect data on these concepts
- ▶ select from a menu of youth employment indicators across four key results areas: employment opportunities, employment quality, employment access and employment skills
- ▶ disaggregate indicators by age, gender, rural/urban location
- ▶ align indicators on youth employment outcomes with Decent Work Statistical Indicators and the 2030 Development Agenda.



Keywords:

Decent work statistical indicators, disaggregation, key indicators of the labour market, national labour force surveys, wages and income, full-time equivalent, job quality, labour force participation, employability, technical and soft skills.

Measuring youth employment outcomes is no simple task. The calculation of certain labour market indicators requires specialist knowledge in labour market econometrics, labour force surveys and data analysis. But the calculation of many labour market indicators is straightforward and, if applied correctly, can offer an accurate and statistically sound method for assessing the progress of a youth employment intervention.

The purpose of this note, therefore, is to provide guidance on the labour market indicators relevant to young people that can be used in the results measurement framework of youth employment interventions. We offer a menu to pick and choose from, depending on the context and objectives of the particular project. The official definitions for key labour market indicators are provided to inform our understanding of how to set and define measures for monitoring and evaluating our interventions, as well as offering an insight into the complexities and issues involved in determining the quality and quantity of a job.

A menu of youth employment indicators

For the purposes of this guide, we break down our menu of youth employment indicators into four key results areas: employment opportunities, employment quality, employment access and employment skills.

The proposed indicators concentrate mainly on the level of the individual beneficiary, focusing on participants who are seeking a job or who aspire to improve their employability, earnings or business and self-employment outcomes. The indicators are primarily useful for interventions that target individual beneficiaries, including active labour market programmes, such as public employment services, wage subsidies and public works, as well as interventions that promote sustainable livelihoods and income-generating activities.

The indicators proposed have been selected because of their relevance and applicability in youth employment programming. We only propose indicators that can be measured at the project level and which require no more than a basic understanding of youth labour markets. That being said, the definition and calculation of most of the indicators has been set through an international standard-setting process. Therefore, measuring these youth employment indicators will provide credibility and evidence to the monitoring and evaluation of the job outcomes of the youth beneficiaries, as well as to the donors, government partners and other stakeholders.

The name, definition, description and source of each indicator is provided in the tables that

Box 2.1: Framework on the measurement of decent work

The International Conference of Labour Statisticians presents a measurement framework that covers ten substantive elements, corresponding to the four strategic pillars of the Decent Work Agenda, namely full and productive employment, rights at work, social protection and the promotion of social dialogue. The ten elements are:

1. employment opportunities
2. adequate earnings and productive work
3. decent working time
4. combining work, family and personal life
5. work that should be abolished
6. stability and security of work
7. equal opportunity and treatment in employment
8. safe work environment
9. social security
10. social dialogue, employers' and workers' representation.

The proposed framework of indicators for project results measurement purposes builds on this decent work measurement framework.

Box 2.2: Measuring youth employment in the 2030 Agenda for Sustainable Development

Within each of the 17 Sustainable Development Goals (SDGs) that make up the 2030 Agenda are a total of 169 targets, which provide the basis of a roadmap for action. Progress towards these targets will be measured through a set of globally harmonized indicators for monitoring performance.

A number of the indicators are directly related to the youth employment challenge, ranging across three of the 17 SDGs. Youth-related indicators are useful for those interventions which align with national development goals and poverty reduction measures. However, although they provide a helpful starting point, it should be noted that SDG indicators do not offer a detailed insight into youth in labour markets, including issues such as status in employment, hours of work, skills mismatch, social protection and social dialogue. SDG Indicators related to youth development include:

Goal 1: End poverty in all its forms everywhere

- 1.1.1 Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)

Goal 4 : Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

- 4.3.1 Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex
- 4.4.1 Proportion of youth and adults with information and communications technology (ICT) skills, by type of skill

Goal 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

- 8.5.1 Average hourly earnings of female and male employees, by occupation, age and persons with disabilities
- 8.5.2 Unemployment rate, by sex, age and persons with disabilities
- 8.6.1 Proportion of youth (aged 15–24 years) not in education, employment or training
- 8.7.1 Proportion and number of children aged 5–17 years engaged in child labour, by sex and age
- 8.b.1 Existence of a developed and operationalized national strategy for youth employment, as a distinct strategy or as part of a national employment strategy

Source: UN (2017)

follow each section. However, our guidance stops at defining the indicators and does not provide direct instructions on how to tabulate or calculate each indicator. The ILO and its partners offer a wealth of guidance on the coding, computation and tabulation of all labour market indicators; you will find the source for further information next to each indicator in the relevant table.

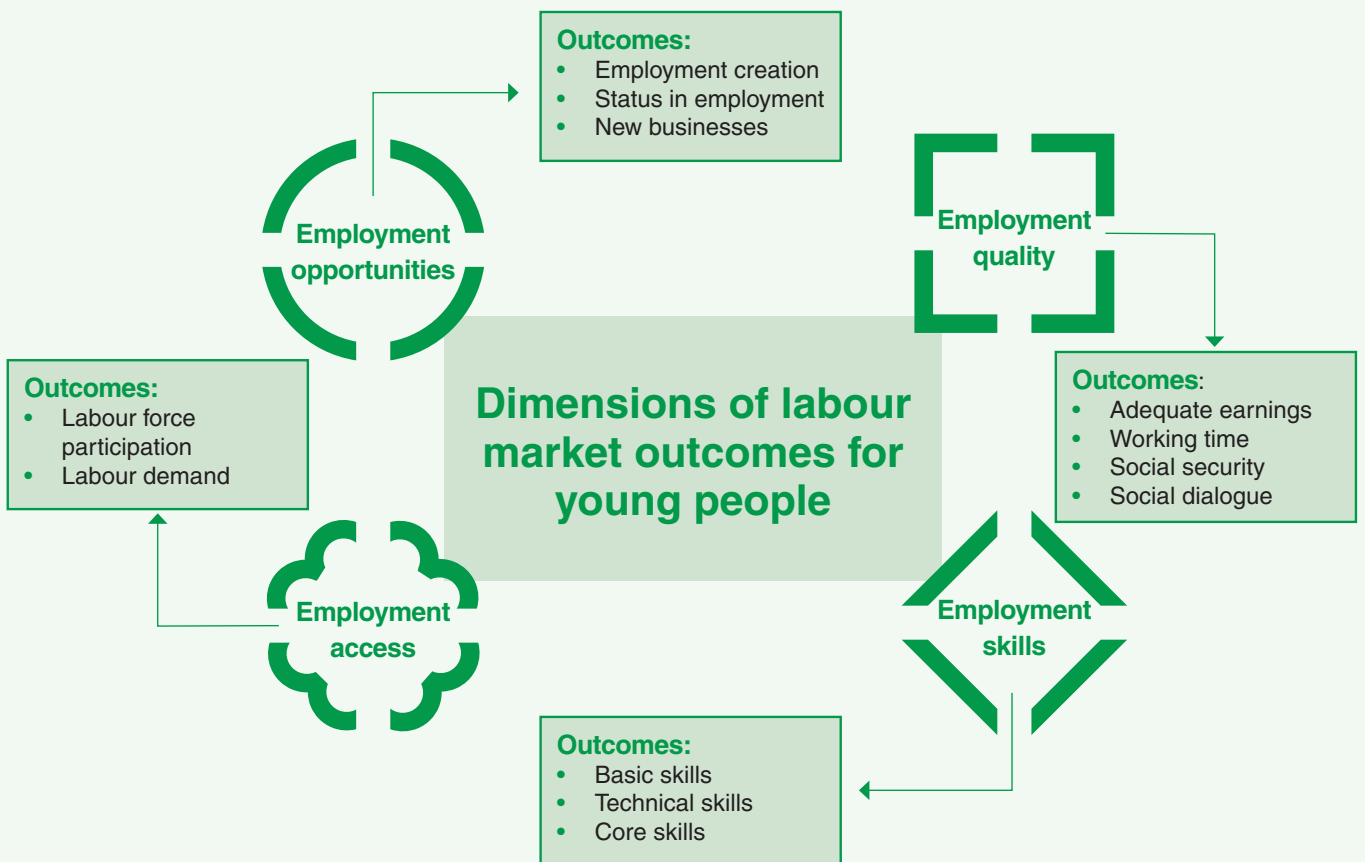
We distinguish between four key dimensions of decent jobs for youth:

(a) **Employment creation:** Outcomes relate to the creation of more jobs for project target

beneficiaries at an individual level. The jobs can either be created for employees or for the self-employed, either as employers or as own account workers. Another key outcome refers to business development, as self-employment and entrepreneurship are the main focus of many youth employment interventions.

(b) **Employment quality:** Outcomes focus on the ability of beneficiaries to achieve better labour market results, through social security provision, social dialogue, increased earnings and decent working time. Many young workers hold jobs of poor-quality in low-

FIGURE 2.1: OVERVIEW OF EMPLOYMENT INDICATORS RELEVANT FOR YOUTH EMPLOYMENT INTERVENTIONS



productivity, low-income activities. These types of results are therefore especially relevant for projects targeting the informal sector and for livelihoods projects.

- (c) **Employment access:** Outcomes relate to activating young people to enter the labour market, improving performance of the labour market and measuring the demand for labour and skills by employers. Outcomes allow the targeting of specific vulnerable or traditionally disadvantaged groups in

the labour market, especially women and youth

- (d) **Employment skills:** Outcomes relate to measuring the supply side of the labour market, young people's skills and competencies. Key employability skills to be measured are inter-related and include basic skills including literacy and numeracy, technical skills or the ability of individuals to perform various tasks and core work skills which are also known as soft skills.

Table 2.1: A menu of labour market indicators relevant to young people

Outcome	Indicator
Employment opportunities	
Employment creation	Number (or rate) of employed project beneficiaries
	Number (or rate) of unemployed project beneficiaries
	Number of full-time equivalent (FTE) jobs
Status in employment	Number of employees
	Number of self-employed: employers
	Number of self-employed: own-account workers
New businesses	Number of newly created firms with more than one paid employee
	Annual sales revenue for beneficiary firms
	Share of newly established beneficiary firms still operational after X months
	Percentage of firms formally registered
Employment quality	
Adequate earnings	Mean nominal monthly/hourly earnings of employees (in local currency)
	The relative difference between the average hourly pay for men and the average hourly pay for women (gender wage gap)
Working time	Average annual working time per employed project beneficiary
Social security	Percentage of youth/employed youth that are active contributors to a pension, sickness or unemployment insurance scheme
Social dialogue	Percentage of young workers whose pay and conditions of employment are covered by a collective bargaining agreement (an agreement between workers' organizations and employers' organizations)
Employment access	
Labour force participation	Labour force participation rates
Labour demand	Numbers of vacancies and hard-to-fill vacancies by occupation
	Average length of time for beneficiary firms to fill a vacancy
	Lists of missing skills
Employment skills	
Basic skills	Youth literacy rate
	Youth numeracy rate
	Youth oral communication skills rate (speaking and listening)

Outcome	Indicator
Technical skills	Job-/task-related measures of skill use at work
	Participation in apprenticeships
	Employer-reported technical skills
	Share of tertiary graduates (enrolment) in STEM subjects
Core skills	Share of beneficiaries with a high level of self-esteem
	Share of beneficiaries with a high level of self-confidence
	Share of beneficiaries with a high level of ability to convey information effectively
	Share of beneficiaries with a high level of problem-solving skills

Three main **disaggregations** are recommended to be applied to youth labour market indicators:

1. **Age:** The ILO uses the United Nations' definition of a young person as between the ages of 15 and 24 years old; however, different youth interventions will have different age requirements, depending on the national context. ILO's school-to-work transition surveys, for example, use the expanded definition of youth – 15–29 years old. It may also be useful to disaggregate by more than one youth age group; for example, ages 15–17 (adolescents of working age) and 18–24 (young adults).
2. **Gender:** Sex-disaggregated data are collected and analysed separately for males and females. The purpose of collecting sex-disaggregated data is to provide a more complete understanding of youth employment data in order to develop better policies and programmes. Collecting and analysing sex- and age-disaggregated data

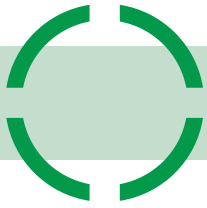
is also critical to integrating gender-aspects effectively into programme design. Without this stratified data, teams cannot accurately analyse differences between the ways in which women and men access and benefit from services and respond to those differences to improve project aims and health outcomes.

3. **Rural/urban:** Official definitions of rural and urban areas are usually country specific and there is no harmonized definition of this widely used concept. We recommend that you apply the national definition when describing the types of work performed in both farm and non-farm jobs and their rural labour-related characteristics. The ILO has published preliminary overviews of national definitions of urban/rural and best practices of international organizations.¹

¹ See: <http://www.ilo.org/global/statistics-and-databases/statistics-overview-and-topics/rural-labour/lang--en/index.htm>

DEFINITION

Disaggregation: Disaggregation is the breakdown of observations into more detailed levels. Data should be disaggregated according to what is to be measured. For example, for individuals the basic disaggregation is by sex, age group, level of education and other personal characteristics useful to understanding how the programme functions and how it affects different (target) groups.



Employment opportunities

The first dimension of employment measurement relates to opportunities. This dimension refers to the quantity of jobs created by a youth employment intervention, with the primary indicator being the employment rate of project beneficiaries. An equally important indicator of employment is the quality of the job – a topic addressed in the next section.

For the selection of indicators on employment we follow definitions as provided by the ILO's ILOSTAT database.² In order to assess decent work for those in employment, we rely on guidance in the ILO's Decent Work Indicators manual (ILO, 2013a) and the school-to-work transition surveys (ILO, 2009). In table 2.2 we briefly describe each indicator and give information on the data type, the history of its use and whether it can be collected at the individual level.

The ILO defines employment from a people-centric perspective (i.e. whether an individual is employed, underemployed or unemployed), rather than focusing on a job as the unit of analysis. The “employed” comprise all persons who worked for pay, profit or family gain for at least one hour in the reference week plus the number of persons who are temporarily absent from their jobs (ILO, 2013a, para. 27 ff.). Three different employment statuses are captured by the terminology: “for pay” captures any person who performed some work for wage or salary, in cash or in kind (wage and salaried workers); “for profit” aims to include workers who are self-employed; and “for family gain” allows for the inclusion of contributing family workers who worked in a family establishment or landholding. A young person is considered to be

employed if they fit this definition and are of “youth age” (the age definition of a young person changes being country-dependent). The employment rate is straightforward to calculate: divide the number of employed persons by the total number of persons of working age (or in the case of youth employment by the number of young people in a given age range, for example 15 to 24 years of age).

Young persons that are not employed are either classified as being “unemployed” or “outside of the labour force”. In order to be considered unemployed a young person must be both (i) carrying out activities to seek employment (during a specified recent period, for example the last week) and (ii) currently available to take up employment given a job opportunity. This definition of unemployment might in some cases lead to an underestimation of labour-utilization, in particular if many young people are discouraged and stop actively looking for work. Therefore, young people not working and not seeking work because they feel that undertaking a job search would be a futile effort, are considered as “discouraged workers” (ILO, 2009).

Defining youth employment in this way runs the risk of veiling the extent of underemployment amongst young people. According to the definition of employment, a young person working for one hour a week would be counted as employed. Therefore, information on employment needs to be complemented with data on the number of hours worked. In a results measurement framework, it would be best to indicate both the number of jobs created and their equivalent in full-time posts in

² See: www.ilo.org/ilostat

order to avoid exaggerating the beneficial effects of the intervention.

Full-time equivalent (FTE) is an approach to measuring jobs that is used and advocated by most international financial institutions as well as by the Donor Committee for Enterprise Development. FTE helps to reveal the total quantity of work created. However, it does not clarify the number of people who benefited from additional jobs and therefore we recommended that FTE be used in combination with youth employment rates. Both measures require the number of hours worked by all project beneficiaries to be monitored.

The below guidance also proposes a number of enterprise-related indicators, given the importance of promoting youth entrepreneurship as a pathway to increasing the number of decent jobs for young people. The indicators includes the number of newly created firms, additional sales revenue, share of newly established beneficiary firms and percentage of firms formally registered. Gathering this information may require increased efforts in data collection, although for the purposes of project-level monitoring and evaluation the information can be collected from the business owners themselves.

Table 2.2: Indicators for employment opportunities

Outcome	Name of indicator	Description	Source
Employment creation	Number of employed young persons / youth employment rate	<p>Persons in employment are defined as all those of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit. They comprise:</p> <ul style="list-style-type: none"> (a) employed persons “at work”, i.e. who worked in a job for at least one hour; (b) employed persons “not at work” due to temporary absence from a job, or to working-time arrangements (such as shift work, flexitime and compensatory leave for overtime). <p>The definition thus includes both, self- and wage employed beneficiaries.</p> <p>The youth employment rate is calculated by dividing the number of employed young person by the total number of young persons (in a given age range).</p>	ILO (2013a, para. 27 ff.)
	Number of unemployed young persons	<p>Persons in unemployment are defined as all those of working age who fulfil the following three criteria:</p> <ul style="list-style-type: none"> (i) not in employment; (ii) carried out activities to seek employment during a specified recent period and (iii) were currently available to take up employment given a job opportunity. <p>Persons who fulfil criteria (i) and (iii) but not (ii) – that is those that are not actively seeking employment – are sometimes referred to as “discouraged individuals”.</p>	ILO (2013a, para. 47ff.)

Outcome	Name of indicator	Description	Source
	Number of full-time equivalent (FTE) jobs	Net additional, FTE jobs created in target enterprises as a result of the programme, per year and cumulatively. Part-time jobs are converted to FTE jobs on a pro-rata basis, based on local definition (e.g., if working week equals 40 hours, a 24-hour per /week job would be equal to a 0.6 FTE job).	ADB (2013)
	Status in employment	<p>Provides information on how jobs held by persons are classified, based on the associated type of economic activity. Employed persons are classified according to the following categories:</p> <p>(a) employees: All those workers who hold "paid employment jobs" who have an explicit (written or oral) or implicit contract of employment with an employer</p> <p>(b) self-employed: all those who are working on their own account where the remuneration is directly dependent upon the profits derived from the goods or services produced</p> <p>(b.I.) self-employed (employers): those who, on a continuous basis, have engaged one or more persons to work for them in their business as "employee(s)"</p> <p>(b.II.) self-employed (own-account workers): those self-employed workers who have not engaged on a continuous basis any "employees" to work for them during the reference period.</p> <p>Although included as categories of employment status, the following three are left out of our guidance: (a) members of producers' cooperatives; (b) contributing family workers; (c) workers not classifiable by status.</p>	ILO (1993, para. 8, ff.)
Status in employment	Number of newly created firms with more than one paid employee	Newly created firm: formal or informal firms that were created as a result of the intervention. Paid employee: Worker holding a job in which the basic remuneration is not directly dependent on the revenue of the employer (can include family members).	ILO (2015)
New businesses	Additional (annual) sales and revenues for beneficiary firms	By keeping track of sales and revenues (on an annual or monthly based), we can measure the incremental sales and revenues (in US\$) for a given period of project-supported businesses.	ILO (2015)
	Share of newly established beneficiary firms still operational after X months	The indicator tracks the ratio of newly established beneficiary firms that are still operational compared to the total number of firms participating in the intervention.	ILO (2015)
	Percentage of firms formally registered	Percentage of firms formally registered when they started operations in the country.	Fiala and Pilgrim (2013)



Employment quality

Job quality is a multidimensional concept and is one of the most important dimensions of decent work. The general or overall quality of a job is the sum of multiple aspects that affect both the employment relation and the work itself. This multidimensional nature of job quality makes the development of a single indicator or system of indicators nearly impossible.

In this section we refer to job quality as those aspects of the employment relationship that have a potential impact on the well-being of workers: these are all the aspects related to the employment contract, remuneration, working hours, social protection and social dialogue.

One of the most important concepts to be measured is earnings or wages. The concept of earnings relates to remuneration both in cash and in kind paid to employees for time worked or work done, together with remuneration for time not worked, such as annual vacation, other types of paid leave or holidays. Earnings should include direct wages and salaries, remuneration for time not worked (excluding severance and termination pay), bonuses and gratuities and housing and family allowances paid by the employer directly to the employee.

Some concepts that might be addressed in a youth employment intervention, such as

Box 2.3: Key components of decent work: Measuring informality

When measuring informality, it is necessary to distinguish between the informal sector and informal jobs. Even workers in the formal economy can have jobs that are classified as informal (for example, due to a lack of social protection coverage or even a contract) and this has important implications for the classification of working conditions in a region.

Certain data must be gathered to determine whether a person is in informal employment. First, core information about the employment characteristics has to be collected, such as occupation and employment status. Second, contextual information, such as workplace characteristics and the type of employment contract, is needed to classify whether the employment is formal or informal. Third, information that identifies whether the employment is taking place in the informal sector, such as whether the enterprise is registered, its legal status and whether both state and employer contributions are being made to social protection provision. Finally, descriptive information on the work characteristics and plans of the employed person under consideration can be collected.

For information on survey design to determine informality, see [ILO \(2013b\)](#).

It is important to acknowledge that different subgroups in the population have a different likelihood of being affected by informality. The report *Women and men in the informal economy: A statistical picture* (ILO, 2018) gives a detailed account of the heterogeneity in informality and its prevalence across many different group. The groups particularly at risk, apart from certain occupations with an overrepresented share of informal employment, are young people and women. This highlights the fact that informality is a concept that should be considered and addressed in youth employment interventions. Gearing interventions towards the informally employed can be a useful in identifying and targeting particularly vulnerable participants.

informal employment, are not directly captured in the indicators suggested. However, when using the indicators jointly, some of these more abstract concepts can still be measured. Given the variety of dimensions to be covered, we decided to provide guidance on only

a few basic measures for each aspect of decent work. Although this does not give exhaustive coverage on these dimensions, it ensures that an entry point is provided for each of the possible fields of youth employment interventions. For each measure, table 2.3 includes a

Box 2.4: Example: How to use selected indicators in project measurement and evaluation

The decision on which indicator can best capture and represent the outcome of interest is just the first step in the measurement process. Data must be collected to compute the indicator, which has to be configured correctly to achieve the desired result. This note cannot give instructions on every proposed indicator. However, the steps involved in computing the NEET indicator are detailed below, which highlight some issues that should be taken into account while constructing indicators. For many of the indicators suggested, detailed information on computation and interpretation can be found in the ILO manual (ILO, 2013c).

Youth not in employment, education or training (NEET)

The NEET indicator is one of the decent work indicators suggested by the ILO and plays a prominent role in the Sustainable Development Goals (SDGs). For SDG 8, decent work and economic growth, it is a key indicator for the labour market situation of young people. The indicator can be calculated as follows (“#” means “number of”):

$$\text{NEET (\%)} = \frac{\text{Total \#youth} - \text{\#youth in employment} - \text{\#youth in education or training who are not in employment}}{\text{Total \#youth}} \times 100$$

The share of NEET among young people is an important indicator because it is a gauge of the potential labour supply among young people. As it includes both discouraged and economically inactive youth, it provides a measure of the degree of social and economic exclusion of young people. Because only young people who are not in employment, education or training are included, it is robust to the educational choices of young people. Rather than counting economically inactive youth, the NEET measure allows us to differentiate between those who are economically inactive but in education and therefore likely to join the labour market successfully and those who have (temporarily) given up on employment.

In order to obtain NEET results that are comparable across countries, it is important to use a common definition of education and training. The ILO suggests that only formal education and training should be considered, excluding non-formal education, such as workshops provided by an institutionalized provider in addition to the formal education system, and informal learning.

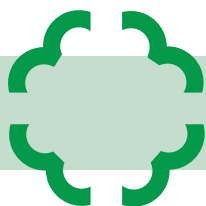
High female NEET rates in comparison those of males in a country can point to gender roles that assign a disproportionate burden of household chores to girls. As this can prevent young girls both from finishing their education and entering the labour market, high female NEET rates, especially among adolescents (15–17 years old), can be understood as an early warning sign of long-term gender equality issues and the presence of barriers to young girls successfully entering the labour market and forging a career.

short description, explains the data type, data sources and indicates whether data can be collected at intervention level.

Efforts have been made to construct aggregate measures for decent work, as discussed

in [Anker et al. \(2003\)](#) and [Ghai \(2003\)](#). However, these measures have been criticized because they do not fully map the concept of decent work and use macro-level indicators that provide very little information about the decent work situation of individuals.

Table 2.3: Indicators for employment quality			
Outcome	Name of indicator	Description	Source
Adequate earnings	Average wage	Mean nominal monthly/hourly earnings of employees (local currency).	ILO (1982, para. 9)
	Gender wage gap	Measures the relative difference between the average hourly pay for men and the average hourly pay for women. The gender wage gap is the difference between the gross average hourly earnings of male and female employees expressed as percentage of gross average hourly earnings of male employees. When the gender pay gap equals “0”, it denotes equality of earnings. Positive values reflect the extent to which women’s earnings fall short of those received by men, where a value closer to “100” denotes more inequality than a value closer to “0”. Negative values reflect the extent to which women’s earnings are higher than men’s.	ILO (1982, para. 9)
Working time	Average annual working time per employed project beneficiary	Hours actually worked is the time spent in a job for the performance of activities that contribute to the production of goods and/or services during a specified short or long reference period.	ILO (2008a)
Social security	Social insurance coverage	Percentage of youth/employed youth that are active contributors to a pension, sickness or unemployment insurance scheme.	ILO (1999, 2013c) .
Social dialogue	Collective bargaining coverage	Percentage of young workers whose pay and conditions of employment are covered by a collective bargaining agreement between workers’ organizations and employers’ organizations.	Ghai (2003) , ILO (2013c) .



Employment access

In order to promote decent employment, it is not enough simply to measure the employment outcomes of a youth employment intervention. Rather, the match between individual skills and employers' requirements is crucial to determine job prospects for young people. The ability to accurately assess the functioning of the labour market in balancing suppliers of labour services (workers) against the demands for labour services (employers), often through labour market policies, is crucial to addressing market distortions. Many youth employment interventions will not directly tackle the demand for skills, as they focus on developing young people's skills rather than improving recruitment in enterprises. However, careful attention to estimating the skills demands of employers when designing and monitoring an intervention can contribute to improved youth job outcomes. Table 2.4 provides an overview of indicators to measure employment access.

A key measure of access to employment is the labour force participation rate, which assesses the extent to which the population is economically active, either employed or actively looking for work. This is an important measure for monitoring those projects which are not only concerned with employment but also with the activation of young people to search and apply for jobs. This measure relates to potentially enhancing the labour supply.

To assess employment demand, some interventions will find it important to monitor the number of vacancies as well as the average length of time that it takes employers take to

fill vacancies. Equally important would be to ask employers to identify the skills gaps and shortages in order to develop training curricula which can improve capacities to fill these reported gaps.

However, it is not advisable for all projects to collect primary quantitative data from employers regarding their vacancy and skills gaps. In many cases, data on vacancies can be collected from public employment services, which often times keep records about their clients (jobseekers) as well as about the vacancies they try to match them with. Vacancy and jobseeker statistics can serve as useful information on current skills supply and demand, and can be complemented by information about the reasons for these vacancies, as it must be established whether vacancies are related to a lack of skills and competencies among job applicants or just the enterprise having difficulty attracting applicants.

In addition to the collection of quantitative vacancy data from secondary sources, qualitative data can be sourced through consultation with a smaller number of companies in selected sectors to provide more detail and capture recent trends in emerging occupations and skills and possible future developments. Content analyses of job requirements, as described in vacancy advertisements, could be used in addition to consultation.

It is also necessary to explore the specific skills that employers find lacking and the impact of skill-shortage vacancies. Projects can consider asking employers questions on recruitment and the skills levels of those leaving

education to take up their first jobs. Employers can be asked whether they have recruited anybody into their first job on leaving education in the past two to three years. It is then

possible to explore employers' perceptions of these recruits in terms of their readiness for work and their skills.

Box 2.5: Employer/establishment skills surveys (ESSs)

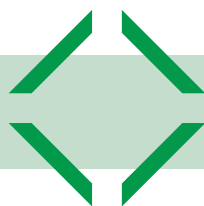
An ESS is an instrument designed to generate data on employer demand for and investment in skills and workers. It helps to define the type, level and composition of skills that individuals need to perform the work demanded by enterprises, as well as determining vacancy rates and whether they relate to a lack of suitable candidates with the required skills. This type of survey not only documents the skill content of current jobs, but, when the correct design is applied, it is also an appropriate tool for investigating future needs, by obtaining information on the type of occupations that will be in higher (or lower) demand or the skills that will be key in future workers' skills portfolios.

This information lays the foundations for generating the knowledge that allows each of the actors in the labour market to adjust their strategies and actions to cope with the challenges they face.

- (1) Policy-makers can shape their education and training policies and also encourage specific human resource strategies, such as:
 - design of initial and continuing education using information on basic areas of expertise in emerging occupations and sectors, as well as upgrading and remedial education for specific segments of the workforce
 - design of active labour market policies: counselling/guidance and retraining for jobseekers.
- (2) Education and training providers (public or private) can receive updated information on skills demand to inform:
 - design of education and training programmes and skills standards, and changes in number of education and training places provided, to serve changing enterprise demand for new entrants
 - design of education and training programmes and skills standards to adjust skills of current workers in response to changing enterprise skills needs.
- (3) Employers can compare themselves to others in terms of training provision or their experience of skills deficiencies, and identify key challenges and opportunities for their sector in terms of:
 - understanding of skills needs drivers to establish the relationship between work organization, product strategies, business positioning and technological changes, and enterprises' skills and training needs.
- (4) Individuals (with the help of careers advisers) can identify skills areas that they might have to upgrade and sectors and occupations with specific skills shortages that offer good job opportunities:
 - identification of skills gaps and labour shortages by level and type of education/ training to contribute to the knowledge generation on future skills needs.

Table 2.4: Indicators for employment access

Outcome	Name of indicator	Description	Source
Labour force participation	Labour force participation rates	Labour force participation rate is a measure of the proportion of a country's working-age population that engages actively in the labour market, either by working or by actively looking for work, relative to the country's total working-age population.	ILO (2016)
Labour demand	Numbers of vacancies	Both the overall number of vacancies and the number of hard-to-fill vacancies, as well as their ratio, can be determined. This data can be obtained through secondary sources, such as public employment services or establishment skills/vacancy surveys.	Řihova (2016)
	Average length of time for employers to fill a vacancy	The average length of time taken for an employer to fill a vacancy can be determined through secondary sources, such as public employment services or establishment skills/vacancy surveys.	Řihova (2016)
	Lists of missing skills	These can be reported either by workers or by employers.	Řihova (2016)



Employment skills

Skills development is essential for increasing the productivity and sustainability of enterprises and improving working conditions and the employability of workers (ILO, 2008b). In order to secure that first job, as well as navigating in the labour market, young women and men need technical skills to perform specific tasks as well as core work skills, such as learning to learn, communication, problem-solving and teamwork. Development of core skills, awareness of workers' rights and an understanding of entrepreneurship are the essential building blocks for lifelong learning and the capacity to adapt to change.

The ILO (2000, para. 9) defines employability skills as:

the skills, knowledge and competencies that enhance a worker's ability to secure and retain a job, progress at work and cope with change, secure another job if he/she so wishes or has been laid off and enter more easily into the labour market at different periods of the life cycle. Individuals are most employable when they have broad-based education and training, basic and portable high-level skills, including teamwork, problem solving, information and communications technology (ICT) and communication and language skills [...]. This combination of skills enables them to adapt to changes in the world of work.

Measuring young people's progress and advances in their work-related skills level is a topic of much debate and there currently is no consensus on the most accurate methods for measuring work skills. A common feature of most skill definitions is to distinguish three types of skills, although these are labelled differently according to organization and

context. The ILO distinguishes basic, technical and core skills (sometimes including a fourth dimension of personal/professional skills, see Brewer, 2013), while in the academic context a grouping of basic, higher-order thinking and affective skills is often used (Cotton, 1996). In this note, we make use of this threefold distinction between basic skills, technical skills and core skills, referring to the ILO definitions.

The success of a skills training programmes depends not only on the skills that young people have acquired through education and training but also on how these relate to the skills required in the labour market. As stated in Note 1, during the diagnostic phase, both labour demand and supply need to be analysed so that the skill gap to be addressed by the intervention is clearly defined. In this subsection, we will first focus on skills outcomes on the labour supply side, discussing measures for basic, technical and core skills. In the next subsection, changes to employer-reported skills to be addressed as a complementary outcome measure for skills will be discussed.

One should be warned that the proposed menu of skills indicators should not be perceived to replace skills accreditation or certification processes in place at national level. Public as well as some private authorities have responsibility for determining the type of skills and how attainment of these skills are measured and awarded. The skills proposed below are intended for individual skills development using academically approved measures which can be applied globally.

MEASURING BASIC SKILLS

Basic skills include literacy, numeracy and the capacity to communicate. These skills are a precondition for almost every type of job. Improving basic skills through youth employment interventions can therefore be an invaluable contribution to enabling young people to find decent employment (Brewer, 2013). These skills are at the heart of each national education system and should be imparted during primary school. Youth employment

interventions can, in some instances, complement or consolidate these skills, but it is beyond both the mandate and the resources of local interventions to provide a full set of these skills. For example, in programmes such as business trainings, basic skills might improve alongside entrepreneurship skills. Standardized tests to measure literacy and numeracy levels are available, which can help to assess basic skills outcomes.

Table 2.5: Indicators for basic skills

Outcome	Name of indicator	Description	Source
Basic skills	Youth literacy rate	<p>The ability to read and write.</p> <p>The OECD offers the Education and Skills Online Assessment, which provides individual-level results linked to the Programme for the International Assessment of Adult Competencies (PIAAC). The assessment contains modules on literacy and numeracy, as well as problem solving. The test can be taken from the age of 15 and therefore fits well with conventional classifications of young people as 15–24 (see http://www.oecd.org/skills/Esonline-assessment/).</p> <p>Enrolment in and/or completion of primary education can be used as a proxy for literacy, although certain individuals may have had some schooling but still be illiterate, while others may have had no schooling but may be literate. Basic literacy is an important prerequisite for many types of further learning.</p>	OECD (2000)
	Youth numeracy rate	<p>The ability to understand and work with numbers.</p> <p>See above for comments and sources.</p>	OECD (2000)
	Youth oral communication rate (speaking and listening)	<p>Oral communication is the dynamic process by which people exchange thoughts, ideas and messages. Listening is the act of interpreting sounds and/or visual stimuli and using those interpretations to give them meaning.</p> <p>The Children, Youth and Families Life Skills project (CYFAR) developed the “communication scale” to assess youth’s ability to communicate by examining the frequency of use of the following skills that are necessary for effective communication practices:</p> <ol style="list-style-type: none"> 1. Awareness of one’s own styles of communication 2. Understanding and valuing different styles of communication 3. Practising empathy 4. Adjusting one’s own styles of communication to match others’ styles (communicative adaptability) 5. Communication of essential information 6. Interaction management. 	Barkman et al. (2002)

MEASURING TECHNICAL SKILLS

The terms “competencies” and “skills” refer to the capacities or abilities of individuals to perform various tasks. For the purposes of this guide, we are concerned primarily with the capacity to perform tasks required for certain jobs; that is, occupational skills. Skills are therefore primarily characteristics of individuals.

Technical skills are normally obtained during specialized vocational or tertiary education. They may also be associated with jobs, where they are prerequisites for the competent performance of tasks, and with education and training programmes and related qualifications, where they are taught and assessed. Evidence suggests that methods of measuring basic skills, such as literacy or cognitive skills, are reliable while technical skills are more difficult to measure and are affected by a variety of external factors (Laajaj and Macours, 2017).

Also, technical skills are not one homogenous set but differ widely according to occupation and the specific workplace. Additionally, the technical skills required in an occupation change over time, with some skills becoming

obsolete. From the intervention perspective, technical skills therefore need to be considered with the skills requirements and potential skills gap firmly in mind. Given that they cannot be as widely applied as basic skills, they are less an end in themselves and their usefulness is more closely tied to whether they make it easier to obtain decent work.

After deciding on specific technical skills to address in a training programme, the best way to measure advancements is to conduct a test that is tailored to the training contents. Ideally, this test would be taken both before and after the training so that progress can be measured. In some special cases, skills acquisition can be monitored through participant outputs during the training. For example, in the ILO Start and Improve Your Business (SIYB) programme, participants are required to draft business plans. Rather than having participants sit a test, which consumes time and resources, the quality of business plans can be directly assessed. For intervention types where participants deliver several outputs during the training, the quality of these

Box 2.6: Most commonly used measurements of skills

Skills characteristics are difficult to measure. At the individual level, the measurement of skills includes psychological tests and various forms of assessment (school grades and tests, assessment centres, worker evaluation). In some types of analysis (especially opinion surveys of employers, employees or graduates), direct questioning on skills is used as well. Other types of analysis use different proxy variables to measure skills supply and demand. These may include:

- (a) qualifications: degrees, diplomas, certificates, acquired in education or training or in a system of recognition learning outcomes. In empirical analysis they are most often expressed as level and field of highest education attained
- (b) occupations: sets of jobs similar in terms of tasks and duties. Standard classifications of occupations (ISCO or similar national classification) are most often used in empirical research
- (c) tasks: activities performed in jobs. This may refer to various types, such as manual tasks, reading, writing, communication with colleagues or customers and work with computers, up to job specific tasks.

outputs can be a useful indirect measure of improved technical skills.

Where an intervention combines skills training with practical work through work-based learning, such as internships or apprenticeships, employer-reported improvements in skills can

be used to complement test results. Where an intervention promotes the placement of participants in formal education, such as vocational training or university programmes, the share of participants in those subjects which are most in demand can be used as an approximate measure of technical skills.

Table 2.6: Indicators for technical skills

Outcome	Name of indicator	Description	Source
Technical skills	Job-task measures of skill use at work	Calculated as the proportion of workers performing various job tasks (either in terms of frequency or at different levels of complexity). These job tasks would cover: reading, writing, numeracy, use of IT, communicating, teamwork, learning new things, physical work and manual dexterity. These job-task indicators provide more direct measures of the broad or generic skills required by employers than do educational qualifications of workers.	Keese and Tan (2013) .
	Participation in apprenticeships	Proportion of youth (aged 15–24) that are apprentices (in either modern or traditional apprenticeships). Needs to be obtained from primary sources. Provides a measure of an important source of learning and skills formation that is not captured by the other proposed indicators of participation in education and training.	Keese and Tan (2013) .
	Employer-reported technical skills	Improved technical skills, as reported by the employer. Needs to be obtained from primary sources. This measure can only be used for programmes with a practical component, such as internships or apprenticeships. Regular feedback on the technical skills of the participants can help to show which modules of the training are most readily assimilated by the participants.	
	Share of tertiary graduates (enrolments) in STEM subjects	The ratio of either graduates or enrolled students in STEM (science, technology, engineering and mathematics) subjects to all graduates or all enrolled students. Note that the relevance of this indicator depends on the skills requirements specified by employers. Provides an indicator of the focus of the tertiary education system on a key area of skills demand that drives economic growth as well as on the potential supply of new labour market entrants with science and technology skills.	Keese and Tan (2013) .

MEASURING CORE SKILLS

The third skill group has been claimed as a major obstacle to obtaining good jobs and has therefore received increasing attention over the past few years. Depending on the field and national context, core skills are also referred to as soft skills, life skills, soft outcomes, practical skills (Blades et al., 2012) or core work skills, key qualifications, essential qualifications and transferable skills (Brewer, 2013). Their role in employment can be summarized as follows:

“They make the difference between being good at a subject and being good at doing a job” (UKCES, 2009).

Core skills comprise the capacity to acquire more skills, or “learning to learn”, social skills such as team work and communication, and problem-oriented thinking. In an environment where technical skills requirements may change quickly, core skills become increasingly important. According to Brewer (2013, p. 5), country evidence suggests “that building the capacity to learn, rather than training to meet detailed forecasts of technical skill needs” may be just as important, “because these may change before curricula can adjust”.

Even where gaps exist in formal education in many countries, leading to a lack of basic skills and technical knowledge that must be addressed, youth employment interventions often have a strong soft skills component. This is

not only because they are seen as particularly important for obtaining employment, but also because they are shown to make progress on traditional academic measures of skills more likely and tend to have a positive influence on reduction of risk behaviours as well (Wilson-Ahlstrom et al., 2014).

In the context of youth employability skills, the use of composite indices is more widespread because different dimensions of skills tend to be closely correlated. Aggregating single indicators also helps to make assessments more reliable and reduces measurement error. We therefore suggest a range of aggregate measures, based on their breadth and the extent to which they are replicable and have been validated.

However, we also offer uni-dimensional measures that are adequate for projects which target just one type of core skill.³ The single indicators we offer can also be combined in a survey to mirror the composition of a specific youth employment intervention. Mixing items from the different scales can help to decrease survey fatigue, as items mapping on the same dimension can trigger recognition effects.

³ For the single-skill measures, we focus on youth of high school age. If a younger age bracket is under consideration, covering ages from 10 to 12 years old, many useful measures are available on the CYFAR platform (see https://cyfar.org/ilm_common_measures).

Box 2.7: The Big Five of measuring non-cognitive skills

Psychologists primarily measure non-cognitive skills by using self-reported surveys or observer reports. They have arrived at a relatively well-accepted taxonomy of non-cognitive skills, called the Big Five, with the acronym OCEAN, which stands for: Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism.

While the Big Five measures are now widely used in psychology, there are several other taxonomies, including the Big Three, the MPQ and the Big Nine, all of which are conceptually and empirically related to the Big Five.

Table 2.7: Indicators for core skills

Outcome	Name of indicator	Description	Source
Core skills	Self-esteem/ self-image	A positive or negative orientation towards oneself, an overall evaluation of one's worth or value. The Rosenberg self-esteem scale is composed of ten items and assesses an individual's feelings of self-worth when the individual compares himself or herself to other people. It is a self-reported measure, aimed at the high-school age cohort.	Schwarzer et al. (1995)
	Self-efficacy/ confidence	Belief in one's ability to succeed in a particular situation. The "general self-efficacy scale" assesses a general sense of perceived self-efficacy with the aim of predicting how an individual will cope with daily stresses and their ability to adapt to different life situations. Self-efficacy is believed to positively affect goal-setting, assertiveness, persistence and effort.	Schwarzer et al. (1995)
	Communication skills	Ability to convey information effectively so that it is received and understood; appropriate verbal/non-verbal communication with colleagues, managers and customers/ others. The "communication scale" is designed for youth aged 12–18 and assesses youth's ability to communicate, focusing on the following skills: awareness of one's own styles of communication, understanding and valuing different styles of communication, practising empathy, adjusting one's own styles of communication to match others' styles, communication of essential information, interaction management.	Barkman et al. (2002)
	Problem solving	Ability to identify problems and devise solutions. The "solving problems scale" is a 24-item scale which assesses youth's problem-solving abilities by examining the frequency of use of the following skills that are needed to engage in problem solving: (1) identify/define the problem; (2) analyse possible causes or assumptions; (3) identify possible solutions; (4) select best solutions; (5) implement the solution; (6) evaluate progress and revise as needed.	Barkman et al. (2002)

Box 2.8: Assessing the skills gap

When talking about a skills gap, this indicates that there is mismatch between the skill level of young people looking for jobs and the type of skills that employers require. Recording unfilled vacancies alone does not give an indication of the size of the skills gap, because many other factors, such as seasonal fluctuations or the sheer number of applicants, can influence that result. Only vacancies that are hard to fill due to the quality of applicants can be considered as indicative of a skills gap. These shortages can either be attributable to a mismatch in the educational level of jobseekers, a lack of experience or a lack of specific skills. Even if it is not possible to obtain reliable information from employers about the specific reasons for unfilled vacancies, the distribution of vacancies across occupations and educational fields can be indicative (see [Řihova, 2016, p. 73](#)). A large number of vacancies in one specific sector of the economy can indicate horizontal skills mismatch. Another channel for detecting this situation is to identify changes in the median wage by occupation. If the median income is rising in one particular sector compared to the dynamics in other sectors or the overall median, this can point to a skills gap.

Box 2.9: Measuring empowerment in rural settings: Women's Empowerment in Agriculture Index (WEAI)

Background

The WEAI is an example of a broad empowerment measure that covers five different dimensions of empowerment and is particularly recommended for projects in a rural context, especially if they have a link to agriculture. It has been collaboratively developed by the International Food Policy Research Institute (IFPRI), the US Agency for International Development (USAID) and the Oxford Poverty and Human Development Initiative. It was developed in 2011 and 2012 and has been extensively piloted and validated. For more information on the index see [Alkire et al. \(2013\)](#).

Dimensions of the WEAI

The WEAI covers five dimensions of empowerment and allows for a holistic and integrated approach to empowerment. Three of the dimensions, “production”, “access to resources” and “income”, are linked to economic empowerment, but the remaining two dimensions, “leadership” and “time use”, contribute to a broader understanding of empowerment. The economic dimensions of the index are strongly oriented towards agriculture: autonomy in production and input in decisions concerning production are the subcategories of production, referring mainly to trade-offs between food and cash-crop farming or livestock. This limits the applicability of the index in other settings, as decisions about entering the labour market or the chosen field of specialization that are important in other economic sectors cannot be captured. Limiting the WEAI to the domain of agriculture does, however, allow for a particularly precise measurement of empowerment.

Data collection in the WEAI

The WEAI is computed using a tailor-made household survey that has to be administered to one female and one male respondent in each participating household. The length of the survey (14 pages, taking about two hours to complete) as well as the need for trained interviewers makes the WEAI a resource-intensive index. Based on feedback from piloting the WEAI, the need for a shortened and easy-to-apply index has been identified. Malapit et al. (2015a) developed the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI). In the A-WEAI, the original ten sub-indicators are reduced to six and the survey takes about 30 per cent less time to administer.

Calculating the WEAI

In the WEAI, empowerment is measured using the five dimensions discussed above and is computed directly in the “five domains of empowerment” (5DE) measure. Additionally, the gender parity index (GPI) is computed, which reflects whether women and men are equally empowered and, if there are differences, how large and in which dimensions these differences are found. This is an important factor because it might be the case that both women and men lack autonomy in their decision-making due to limited resources or other constraints.

Both the 5DE and the GPI range from zero to one, where higher values correspond to higher levels of empowerment. The measures allow for a finely grained distinction between different levels of empowerment, because the index will take a different value dependent on the number of dimensions in which women are disempowered. The weighted sum of both the 5DE and the GPI forms the WEAI score.⁴

Methodological innovations

Indices of empowerment tend to be administered only to women. This has the drawback that the relative autonomy of men in the same household cannot be assessed. In cases where all members of the community lack autonomy, assessing only women's empowerment can therefore be misleading. By computing both the 5DE, which can be considered to be an individual empowerment scale, and the GPI, which assesses relative empowerment, this problem is solved in the WEAI.

How to apply and administer the WEAI: Pilot study

Women's empowerment can be an aim in itself, but it can also be considered as an approach that fosters children's nutrition and well-being, as well as a source of economic growth. This is examined in the case study by [Malapit et al. \(2015b\)](#) on the impacts of gender equity on agricultural production and dietary diversity for women and children. The study looked at the development of child nutrition in Nepal over a ten-year period. The main outcomes are maternal and children's dietary diversity as well as the maternal body mass index (BMI) and child anthropometry. One notable aspect of the study was the decision to test each of the five dimensions of the WEAI both separately and jointly to determine their effects. In the empirical model, women's empowerment is instrumented through the ratio of female and male newborns, as well as the distance to the local market to ensure that the effect of empowerment on nutrition is not driven by a common third variable. The study found that empowerment in general is a predictor for the nutritional outcomes, but that predictive power varies widely by empowerment dimension. Income effects are insignificant, while autonomy in agricultural production and hours worked have a positive and significant impact on both maternal and children's dietary diversity. The disparate findings across dimensions show that it is important to take different dimensions of empowerment into account. Measuring only one aspect of empowerment might lead to under- or overestimation of the outcomes of the intervention.

Source: [Malapit et al.\(2013\)](#)

⁴ For detailed information on the calculation of WEAI scores, see <https://www.ifpri.org/weai-training-materials>



KEY POINTS

1. **Youth labour market indicators can be disaggregated by age, gender and rural/urban location.** The ILO uses the United Nations definition of a young person as between the ages of 15 and 24 years old; however, different youth interventions will have different age requirements, depending on the national context.
2. **Employment opportunity outcomes refer to the quantity of jobs created through a youth employment intervention,** with the primary indicator being the employment rate of project beneficiaries. The ILO defines employment from a people-centric perspective (i.e. is an individual employed, underemployed or unemployed), while other institutions focus on a job as the unit of analysis, using measures such as the full-time equivalent (FTE).
3. **Job quality is a multidimensional concept, covering both the employment relationship and the work content itself.** Outcome indicators can be selected across aspects related to the employment contract, remuneration, working hours, social protection and social dialogue. One of the most common concepts to be measured is earnings – cash and in-kind payments to employees for time worked or work done, together with remuneration for time not worked, such as annual vacation, other types of paid leave or holidays.
4. **The match between individual skills and employers' job requirements is crucial to determining job prospects for young people.** A key measure of access to employment is labour force participation, which assesses the extent to which the population is economically active – either employed or actively looking for work. This is an important measure to monitor for those projects which are not solely focused on employment but are also concerned with the activation of young people to search and apply for jobs.
5. **Measuring young people's progress and advancements in their work-related skills level is a topic of ongoing debate** and there is as yet no consensus on the most accurate methods for measuring work skills. A common feature of most skill definitions is to distinguish three types of skills, although these are labelled differently according to organization and context. The ILO distinguishes basic, technical and core skills (sometimes including a fourth dimension of personal/professional skills).



KEY RESOURCES



- Brewer, L. 2013. *Enhancing youth employability: What? Why? And How? Guide to core work skills* (ILO, Geneva).



- ILO. 2013c. *Decent work indicators: Guidelines for producers and users of statistical and legal framework indicators* (Geneva).



- Keese, M.; Tan, J.P. 2013. *Indicators of skills for employment and productivity: A conceptual framework and approach for low-income countries*, a report for the human resource development pillar of the G20 multi-year action plan on development (OECD, World Bank, ETF, ILO and UNESCO).



- ▶ Malapit, H. J. L.; Kovarik, C.; Sproule, K.; Meinzen-Dick, R.; Quisumbing, A. 2015a. *Instructional guide on the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI)*, (Washington, DC, IFPRI).



- ▶ Řihova, H. 2016. *Using labour market information: Guide to anticipating and matching skills and jobs* (ILO, ETF and CEDEFOP).

REFERENCES

- ▶ Alkire, S.; Meinzen-Dick, R.; Peterman, A.; Quisumbing, A.; Seymour, G.; Vaz, A. 2013. "The Women's Empowerment in Agriculture Index", *World Development*, Vol. 52, Dec., pp. 71–91.
- ▶ Anker, R.; Chernyshev, I.; Egger, P.; Mehran, F.; Ritter, J. 2003. "Measuring decent work with statistical indicators", *International Labour Review*, Vol. 142, No. 2, pp. 147–178.
- ▶ Asian Development Bank (ADB) et al. 2013. *Memorandum regarding IFIs harmonized development results indicators for private sector investment operations*.
- ▶ Blades, R.; Fauth, B.; Gibb, J. 2012. *Measuring employability skills: A rapid review to inform development of tools for project evaluation* (London, National Children's Bureau).
- ▶ Barkman, S.; Machtmes, K., 2002. A research-based model for the design and evaluation of youth development programs. *J. P. Terry (Ed.), CYD anthology*, pp.64-69. The survey instrument can be found at: <https://cyfar.org/content/solving-problems> and <https://cyfar.org/content/communication-scale-0>
- ▶ Brewer, L. 2013. *Enhancing youth employability: What? Why? And How? Guide to core work skills* (ILO, Geneva).
- ▶ Cotton, K. 1996. "Affective and social benefits of small-scale schooling", *ERIC Digest* (Charleston, WV).
- ▶ Fiala, N.; Pilgrim, M. 2013. *Intervention models and performance indicators of the ILO Small Enterprises Unit* (Berlin and Geneva, DIW and ILO).
- ▶ Ghai, D. 2003. "Decent work: Concept and indicators" *International Labour Review*, 142, pp. 113–145. ILO. 1982. *Resolution concerning statistics of the economically active population, employment, unemployment and underemployment*, adopted by the 13th International Conference of Labour Statisticians.
- ▶ International Labour Organization (ILO). 1993. *Resolution concerning the International Classification of Status in Employment (ICSE)*, adopted by the 15th International Conference of Labour Statisticians.
- ▶ ILO. 1999. *Decent work*, Report of the Director-General. International Labour Conference, 87th Session. Geneva.
- ▶ ILO. 2000. *Resolution concerning human resources training and development*, adopted by the International Labour Conference, 88th session, Geneva.
- ▶ ILO. 2008a. *Resolution concerning the measurement of working time*, adopted by the 18th International Conference of Labour Statisticians.
- ▶ ILO. 2008b. *Skills for improved productivity, employment growth and development*, Report V. International Labour Conference, 97th Session, Geneva.

- ▶ ILO. 2009. *ILO school-to-work transition survey: A methodological guide, Module 2, SWTS Questionnaires* (Geneva).
- ▶ ILO. 2013a. *Resolution concerning statistics of work, employment and labour underutilization*, adopted by the 19th International Conference of Labour Statisticians.
- ▶ ILO. 2013b. *Measuring informality: A statistical manual on the informal sector and informal employment* (Geneva).
- ▶ ILO. 2013c. *Decent work indicators: Guidelines for producers and users of statistical and legal framework indicators* (Geneva).
- ▶ ILO. 2015. *Global Wage Report 2014/15: Wages and income inequality* (Geneva).
- ▶ ILO. 2016. *Key Indicators of the Labour Market database* (Geneva).
- ▶ ILO. 2018. *Women and men in the informal economy: A statistical picture*, third edition (Geneva).
- ▶ Keese, M.; Tan, J.P. 2013. *Indicators of skills for employment and productivity: A conceptual framework and approach for low-income countries*, a report for the human resource development pillar of the G20 multi-year action plan on development (OECD, World Bank, ETF, ILO and UNESCO).
- ▶ Laajaj, R.; Macours, K. 2017. *Measuring skills in developing countries*, Policy Research Working Paper, No. 8000 (Washington, DC, World Bank).
- ▶ Malapit, H.J.L.; Kadiyala, S.; Quisumbing, A.R.; Cunningham, K.; Tyagi, P. 2013. *Women's empowerment in agriculture, production diversity, and nutrition: Evidence from Nepal*, IFPRI Discussion Paper 01313 (Washington, DC, IFPRI).
- ▶ Malapit, H. J. L.; Kovarik, C.; Sproule, K.; Meinen-Dick, R.; Quisumbing, A. 2015a. *Instructional guide on the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI)*, (Washington, DC, IFPRI).
- ▶ Malapit, H. J. L.; Kadiyala, S.; Quisumbing, A. R.; Cunningham, K.; Tyagi, P. 2015b. "Women's empowerment mitigates the negative effects of low production diversity on maternal and child nutrition in Nepal." *The journal of development studies*, 51(8), 1097-1123.
- ▶ Olenik, C.; Fawcett, C.; Boyson, J. 2013. *State of the field report: Examining the evidence in youth workforce development*, Final report, USAID youth research, evaluation and learning project (Washington, DC, USAID).
- ▶ Programme for International Student Assessment, 2001. *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Organisation for Economic Co-operation and Development.
- ▶ Řihova, H. 2016. *Using labour market information: Guide to anticipating and matching skills and jobs* (ILO, ETF and CEDEFOP).
- ▶ Rosenberg, M. 1965. *Society and the adolescent self-image* (Princeton, NJ, Princeton University Press).
- ▶ Schwarzer, R.; Jerusalem, M. 1995. "Generalized Self-Efficacy scale", in J. Weinman, S. Wright and M. Johnston (eds): *Measures in health psychology: A user's portfolio. Causal and control beliefs Windsor*, UK: NFER-NELSON), pp. 35-37.
- ▶ UK Commission for Employment and Skills (UKCES). 2009. *The employability challenge* (London).
- ▶ United Nations (UN). 2017. *Report of the inter-agency and expert group on sustainable development goal indicators (E/CN.3/2017/2)*, Annex 3. Revised list of global sustainable development goal indicators.
- ▶ Wilson-Ahlstrom, A.; Yohalem, N.; DuBois, D.; Ji, P.; Hillaker, B.; Weikart, D.P. 2014. *From soft skills to hard data: Measuring youth program outcomes* (Washington, DC, The Forum for Youth Investment).



Case study:

SELECTING INDICATORS FOR THE NORTHERN UGANDA YOUTH ENTREPRENEURSHIP PROGRAMME

This case study is based on "External Evaluation of YDP and NUYP Programmes: Final Evaluation Report, 23rd September 2016" and "Northern Uganda Youth Entrepreneurship Project Review 2013-2016"

Learning objectives

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

- ▶ identify the relevant dimensions of youth employment, selecting from dimensions of employment opportunities, employment quality, employment access and employment skills
- ▶ be able to navigate and select from a “menu” of indicators based on the objectives and design of the youth employment intervention.

Introduction and case study context

Uganda has one of the fastest-growing populations in the world. Over three-quarters of Ugandans are under the age of 30, implying a demographic with great economic and social potential. Yet, many young Ugandans frequently find themselves without opportunities for sustainable work. Unemployment, underemployment and working poverty are widespread phenomena. Young women, in particular, face social and economic barriers due to cultural norms and sexual harassment and discrimination.

In Northern Uganda, a 20-year long civil war has left the local population with additional economic, social and psychological challenges. Young people suffered disrupted schooling and communities were displaced. The legacy of war continues to limit opportunities in the labour market. Youth unemployment rates are high. Employment in the informal sector is common.

The Government of Uganda has put in place a policy to promote the development of micro,

small and medium-sized enterprises (MSMEs) to become the main vehicle for expanding production, providing sustainable jobs and enhancing economic growth. A new MSME Directorate in the Ministry of Trade, Industry and Cooperatives saw entrepreneurship education and training as one means to remedy the problems faced by young people in Northern Uganda. It recognized the lack of wage and salaried employment opportunities in the private sector and sought to build participants' capacity to start up income-generating activities and become self-employed. To this end, the UK Department for International Development (DFID) supported the Government's objectives by funding the Northern Uganda Youth Entrepreneurship Programme (NUYEP).⁵

This case study focuses on choosing appropriate labour market indicators relevant for young people to use in the monitoring and evaluation frameworks of NUYEP.

⁵ NUYEP was implemented by Enterprise Uganda (EUG), in partnership with Youth Business International (YBI), and ran between 2013 and 2016.

Selecting indicators

NUYEP aims to support 10,500 beneficiaries, 80 per cent of whom are youth (for the purpose of this project defined to be person between the ages of 18 to 35). The programme targets five conflict-affected sub-regions in the North of Uganda with the aim of improving the livelihoods of young people and their households through entrepreneurship.

Analysis conducted by NUYEP during the programme design phase revealed that previous interventions in the sub-regions focused on children, women and people with disabilities, while youth needs were largely ignored. Furthermore, a survey of female youth by UNICEF showed that incomes and employment among young women are especially low, with the majority of young women having earnings below the absolute poverty line. Young women with children, orphans, single mothers, former abductees and victims of gender-based violence were identified as the groups most in need of livelihood interventions.

The NUYEP approach emphasizes starting and running an enterprise to transform unemployed and underemployed youth and their household members into business owners with an improved flow of income. The programme is based on a six-stage cycle:

- ▶ Stage 1: Entrepreneurship awareness and mobilization
- ▶ Stage 2: Business and Enterprise Start-up Tool (BEST), mass training workshops
- ▶ Stage 3: BEST follow-up workshops, offering follow-on support
- ▶ Stage 4: Specialized business skills clinics
- ▶ Stage 5: One-on-one volunteer mentoring and business counselling services
- ▶ Stage 6: Linkage to finance, with a focus on savings and investment clubs and savings and credit co-operatives (SACCOs).

Participants comprise individuals who are willing to invest an initial amount of their own money to receive the Business and Enterprise Start-up Tool, which is delivered through mass-training events of up to 800 participants in a purpose-built marquee over five days. The “high flyers” then self-select to receive more intensive levels of support during Stages 3 to 6.

These outputs are expected to lead to the main outcome of the project, which is the creation and expansion of 6,000 youth- or family-owned businesses in Northern Uganda, of which 1,000 will create additional jobs. This is expected to contribute to creating a peaceful and productive youth population with improved livelihoods in Northern Uganda.



Discussion Topics

1. NUYEP wants your help to select appropriate outcome and high-level goal indicators to measure the success of their project. The first step is to decide which key dimension of decent jobs to measure. Which would you recommend – and why?
2. Which indicators would you recommend that NUYEP measures?
3. NUYEP needs particular help with “employment opportunities”. They want your guidance on how to define and measure each key indicator. What would you advise?

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Note

3

Establishing a monitoring system



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

NOTE 3
Establishing a monitoring system

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. 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 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 3: Establishing a monitoring system/ International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131660-2 (print)
978-92-2-131661-9 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلم في برامج سوق العمل النشطة - الملاحظة 3: إنشاء نظام مراقبة، ISBN 978-92-2-630791-9 (print), 978-92-2-630792-6 (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.

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

Printed in Switzerland

Contents

Establishing a monitoring system	1
Step 1: Define the intervention logic	3
Theory of change: The link between programme design and higher-level results	3
The results chain: From theory to practice	5
Step 2. Choose key performance indicators	6
Identifying indicators	6
Good practice when specifying indicators	8
Step 3: Select data collection tools	12
Which data collection tools?	12
When to collect monitoring data?	12
Who collects data?	16
Step 4: Deploy the system to aggregate, analyse and report on results	17
Aggregating results	17
Data analysis	19
Reporting on results	20
Key points	22
Key resources	22
References	23
Case study: Establishing a monitoring system for the Jordan rural economic growth and employment project	24
Figures	
3.1: Stylized theory of change for a youth employment project	4
3.2: Components of a results chain and examples	5
3.3: Example of a projection	11
3.4: Dashboard for data monitoring during half-day classroom-based training	21
Tables	
3.1: Typical components of a monitoring budget	3
3.2: Examples of activity indicators for youth employment projects	7
3.3: Examples of output indicators for youth employment projects	7
3.4: Examples of indicators	9
3.5: Overview of data collection tools	13
3.6: Logical framework for an entrepreneurship promotion programme (example taken from the ILO's Start and Improve Your Business programme)	18
3.7: Tailoring reports to our target audience	21





Establishing a monitoring system



Prerequisites:

This note requires no prior knowledge. However, it would be advisable to first read Note 1 on diagnosing, planning and designing youth employment interventions. This note guides readers through the key steps required to set up a quality monitoring system that can both steer performance improvements and provide the foundation to prove programme impact.



Learning objectives:

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

- ▶ develop a visual theory of change diagram and translate this into a results chain which maps out the intervention logic from inputs through to higher-level goals
- ▶ choose SMART (specific, measurable, attributable, relevant and time-bound) indicators that describe and explain change
- ▶ select appropriate measurement tools and specify the timing and responsibilities for data collection
- ▶ deploy a monitoring system to aggregate, analyse and report on results.



Keywords:

Theory of change, assumptions, results chain, logic model, inputs, activities, outputs, outcomes, higher-level goals, indicators, proxy measures, baseline, targets, quantitative data, qualitative data, logical framework, management information system

Monitoring provides information on a continuous basis to inform programme managers about planned and actual developments. Monitoring involves collecting and analysing data to verify that resources are used as intended, that activities are implemented according to plan, that the expected products and services are delivered and that intended beneficiaries are reached. Effective monitoring should be central to all projects. It helps to detect problems, take corrective actions and lay the groundwork to produce evidence about what works in creating decent jobs for youth. That being said, monitoring systems come with a cost (see box 3.1).

Monitoring also provides the foundation to evaluate an intervention. In fact, a good evaluation is hard to conduct without good monitoring information from actual implementation. If no reliable information about the progress and quality of implementation is available, then any evaluation undertaken will run the risk of misinterpreting the reasons for the success or failure of the project.

This note summarizes the key steps for building a monitoring system that should be followed in any project, regardless of whether an evaluation will also take place:

- ▶ Step 1. Define the intervention logic: Draw a theory of change and results chain
- ▶ Step 2. Choose key performance indicators
- ▶ Step 3. Select data collection tools
- ▶ Step 4. Deploy the system to aggregate, analyse and report on results.

Step 1: Define the intervention logic

THEORY OF CHANGE: THE LINK BETWEEN PROGRAMME DESIGN AND HIGHER-LEVEL RESULTS

Underlying any programme design is a theory of change (ToC). The ToC can be expressed as a diagram showing how a programme plans to impact its beneficiaries, and the set of assumptions we make about why these particular project activities will foster positive change. Figure 3.1 maps a simplified theory of change, showing how on-the-job training can lead, through a series of intermediate steps and external

assumptions, to higher levels of technical and core skills.

Practitioners should draw a visual theory of change for every intervention. Ideally, this is developed during the objective-setting phase of the project design (see Note 1), when all relevant stakeholders can be brought together to agree on a common vision for the project. A

Box 3.1: Resourcing

Monitoring systems can be expensive. In addition to fixed costs (computing hardware and software, staff) there are also variable costs that include training local enumerators, contracting outside consultants and publicizing findings (see table 3.1). It is important that a project's monitoring system is properly budgeted. It is often the case that, when the costs are realized, programme managers hesitate to spend significant resources on a monitoring system, as this expenditure appears to be at the expense of intervention activities. Yet, without suitable monitoring systems, a programme runs the risk of underperformance or even failure. At the end of the day, monitoring systems are critical to project management and a crucial component of any intervention.

Table 3.1: Typical components of a monitoring budget

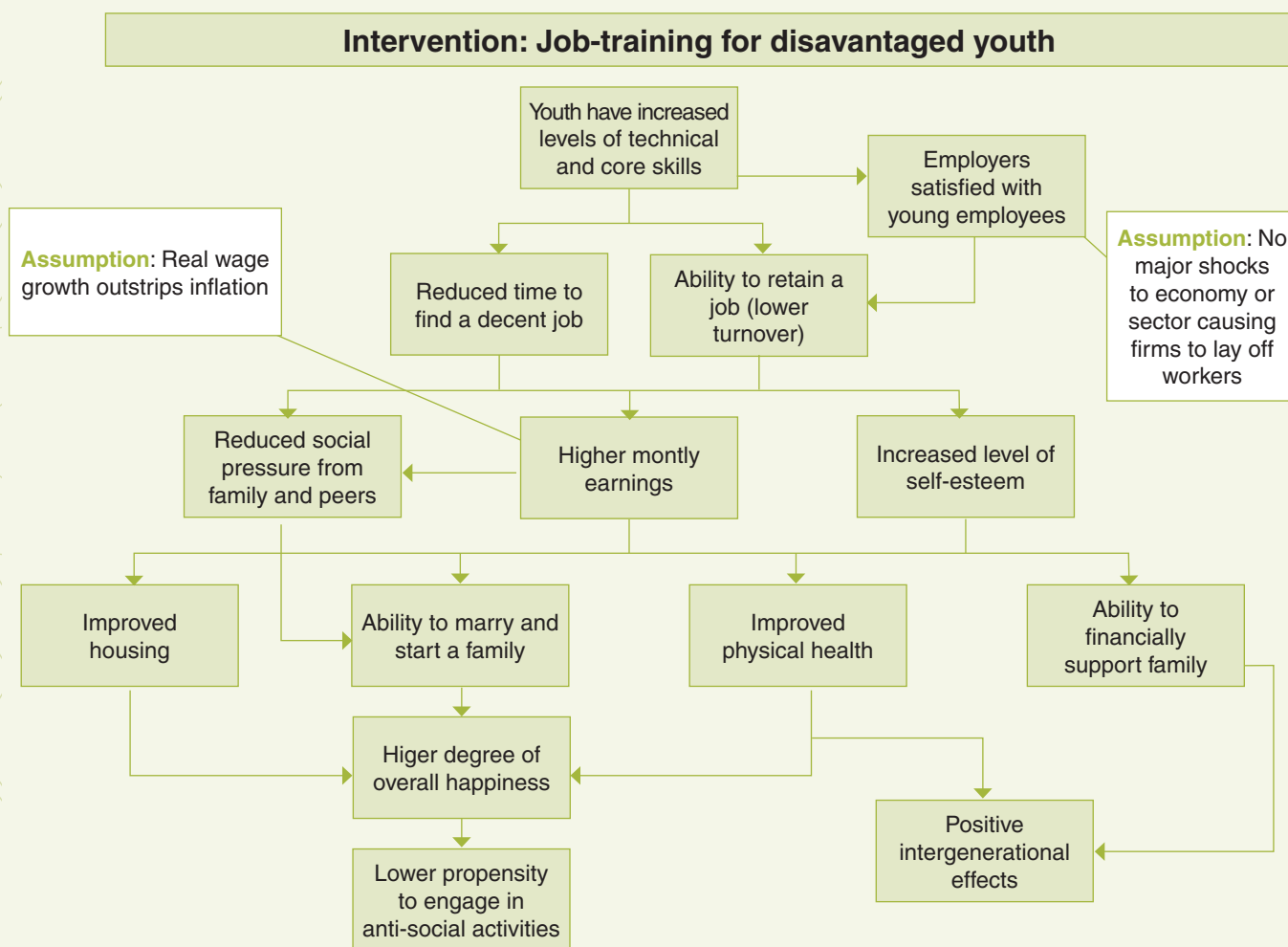
Fixed costs	
Staff costs	<ul style="list-style-type: none"> Headquarters: Percentage of a monitoring and evaluation (M&E) coordinator's time to manage M&E system. Can range from 10 per cent to 100 per cent, depending on project size Locally: Typically, 50–100 per cent of a local M&E officer's time to manage implementation of M&E activities, plus junior support staff
Equipment	Computers, voice recorders, cameras, etc.
Software	Licences for quantitative and qualitative analysis tools
Variable costs	
Training	Capacity building for staff, enumerators, community members, etc.
Travel	Travel from HQ to the field for periodic check-ins and technical assistance. Local travel to field sites to ensure standardized implementation of M&E activities
Data collection and analysis	Contracting of third-party vendors, such as survey firms
Consultants	Contracting of external experts for specific tasks
Printing	Instruments, reports, etc.

theory of change helps both the programme manager and the evaluator to:

- ▶ reach a common understanding about the programme
- ▶ refine and enhance the programme logic using a visual tool

- ▶ differentiate between “what the intervention does” and “what it wants to achieve”
- ▶ communicate externally about what the intervention does, and how it makes a difference
- ▶ identify the important outcomes to be measured.

FIGURE 3.1: STYLIZED THEORY OF CHANGE FOR A YOUTH EMPLOYMENT PROJECT



DEFINITION

A **theory of change** is an articulation of the way in which programme/project planners aim to produce results. Often, theories of change are represented in a visual format that can range in complexity and which can include a logframe approach or a results chain to depict programme/project theory.

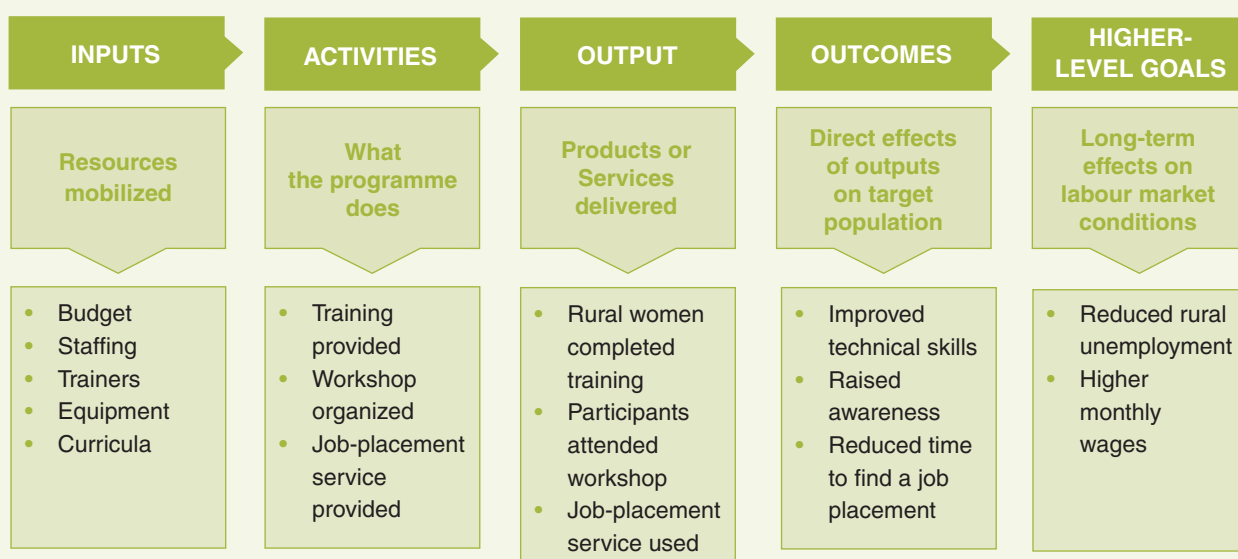
Assumptions describe the conditions that must exist if cause–effect relationships in the theory of change are to occur as expected. They are external factors, beyond the intervention’s control, but nonetheless critical for the success of the intervention.

THE RESULTS CHAIN: FROM THEORY TO PRACTICE

A theory of change can be “operationalized” in the form of a results chain. Results chains (also known as logic models) provide stakeholders with “a logical, plausible sequence” of how the planned intervention will lead to the

desired results. As shown in figure 3.2, this sets out a sequence of resources, activities and services provided are expected to influence the direct and long-term effects on our target population.

FIGURE 3.2: STYLIZED EXAMPLE OF A RESULTS CHAIN FOR A RURAL EMPLOYMENT INTERVENTION



DEFINITION

Inputs: The resources used by the project, including budget, staff, partners, and equipment.

Activities: The actions, processes, techniques, tools, events and technologies of the programme. Describe these activities with an action verb (*provide, facilitate, deliver, organize, etc.*).

Outputs: The products and services provided by the implementing organization. They indicate if a programme was delivered as intended. Outputs are typically expressed as completed actions (trained, participated, used, funded, etc.).

Outcomes (also known as immediate objectives): The short- to medium-term effects (usually within several months of and up to two years after the implementation) on the beneficiary population resulting from the project outputs. These may include changes in attitudes, knowledge and skills, which can often be relatively immediate effects, as well as changes in aspects such as behaviours, labour market status, job quality, etc., which may take more time to manifest themselves. The key outcomes targeted should be those defined in the project development objective. Outcomes are typically expressed at an individual level and indicate an observable change (*increased, improved, reduced, etc.*).

Higher-level goals: The long-term project goals, usually relating to overall living standards in the area where the intervention takes place. They can be influenced by a variety of factors. This level of the results chain is also often labelled “development objective” or “impact”. We prefer the phrase “higher-level goals” to avoid confusion with the specific meaning of “impact” in the context of impact evaluation (see Note 5).

Step 2. Choose key performance indicators

Once we have a results chain, how do we know whether what has been planned is actually happening? One of the biggest challenges in monitoring is choosing what kind

of information to collect in order to know whether we are achieving our objectives. We now need to identify appropriate (key performance) indicators.

IDENTIFYING INDICATORS

Indicators answer the question “How will I know?”. They are:

- ▶ key aspects of (or **proxies** for) the element that we want to measure, even though they may not necessarily be fully representative
- ▶ tangible signs that something has been done or that something has been achieved; they are the means we select as markers of our success (Shapiro, 2003).

Indicators are a crucial element of a monitoring system because they drive all subsequent data collection, analysis and reporting. Without a clear set of indicators, monitoring and evaluation (M&E) activities lose their capacity to compare actual progress with what was projected and agreed upon (Gosparini et al., 2003). They are required at each level of the results chain.¹ Indicators on the level of outputs, outcomes and higher-level goals are referred to as “key performance indicators”.

Input indicators measure resource utilization. They track whether you have the inputs required to implement the intervention. Common input indicators include:

- ▶ budget allocation and expenditure
- ▶ amount and share of matching funds raised
- ▶ number of programme staff, by level
- ▶ number of local facilitators under contract
- ▶ number of local organizations who provide in-kind contributions.

Activity indicators measure what the intervention does. They seek to understand the extent to which a project was delivered as planned, and to highlight obstacles to implementation. Table 3.2 presents examples of activity indicators for different types of youth employment interventions.

DEFINITION

A **proxy** is an indirect measure of the desired change, which is strongly correlated to that change. It is commonly used when direct measures are unobservable and/or unavailable. For example, when looking for ways to measure young people’s engagement and participation it may be appropriate to collect information about the numbers of young people involved in volunteering locally and how much time they spend doing these activities.

¹ The following section is based on guidance issued by the UK Government. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/304626/Indicators.pdf.

Table 3.2: Examples of activity indicators for youth employment projects

Type of project	Activities
Training and skills development	<ul style="list-style-type: none"> • Number of workshops offered • Number of training hours • Number of youth screened/enrolled • Number of employers offering internships • Number of internships available
Subsidized employment (e.g. public works and public services programmes)	<ul style="list-style-type: none"> • Number of workfare projects by type and location • Number of municipalities providing public works/services
Employment services (e.g. job placement support)	<ul style="list-style-type: none"> • Number of career counselling services created (in labour offices, in schools, etc.) • Number of job counselling sessions offered • Number of career and job fairs organized
Youth enterprise and entrepreneurship promotion	<ul style="list-style-type: none"> • Number of business plan competitions organized • Number of hours of support services provided • Average number of hours of mentoring provided per week/month
Youth-inclusive financial services	<ul style="list-style-type: none"> • Number of workshops organized for participating financial institutions • Micro-loan scheme for young entrepreneurs launched • Youth-targeted savings account create

Output indicators measure what your intervention produces. They describe the delivery of tangible products and services, such as

training and technical assistance. Table 3.3 presents examples of output indicators for different types of youth employment interventions.

Table 3.3: Examples of output indicators for youth employment projects

Type of project	Outputs
Training and skills development	<ul style="list-style-type: none"> • Number and percentage of youth who attend at least 80 per cent of the training • Number of certificates awarded • Number of youth placed in internships • Average length of internships completed (in weeks)
Subsidized employment (e.g. public works and public services programmes)	<ul style="list-style-type: none"> • Number of beneficiaries employed in each activity • Number of temporary jobs created (by type and sector)
Employment services (e.g. job placement support)	<ul style="list-style-type: none"> • Number of youth participating in job placement services • Number and percentage of youth matched with employers • Number of companies and youth participating in local careers/jobs fairs
Youth enterprise and entrepreneurship promotion	<ul style="list-style-type: none"> • Number of youth submitting completed business plan • Number of youth enterprises supported annually • Number and percentage of youth talking to their mentor at least once every two weeks
Youth-inclusive financial services	<ul style="list-style-type: none"> • Number of staff trained in partner financial institutions • Number of business loans issued to young people (by type of enterprise) • Average loan size • Number of youth saving accounts opened

Outcome and higher-level goal indicators require particular attention. The results of youth employment interventions can be highly diverse and are not limited to labour market outcomes. We therefore need to choose indicators across different domains of

employment, including employment opportunities, job quality, equal access to employment and skills development. Note 2 provides a “menu” of indicators that youth employment interventions can choose from when determining the outcomes and higher-level goals.

GOOD PRACTICE WHEN SPECIFYING INDICATORS

Bring in other stakeholders: Choosing indicators without the proper involvement of internal and external stakeholders can result in a lack of ownership (Kusek and Rist, 2004). Collaborate with local partners and stakeholders in the community to arrive at a mutually agreed set of goals, objectives and performance indicators for the programme.

Choose the right number of indicators: It is common to define several indicators for each element in the results chain, especially regarding outcomes or higher-level outcomes. However, choosing too many indicators will complicate the monitoring system and increase the burden of data collection, analysis and reporting. It is important to identify two or three key indicators that best capture each change in the results chain (see table 3.4 for examples).

Meet quality standards: Even though there are no absolute principles governing what makes a good indicator, the commonly cited SMART characteristics can be useful. SMART indicators are:

- ▶ **Specific** – to clearly set out what information is required

- ▶ **Measurable** – to ensure that the information can actually be obtained at a reasonable cost
- ▶ **Attributable** – to the efforts of the intervention
- ▶ **Relevant** – to the result we want to measure
- ▶ **Time-bound** – to ensure that the data can be obtained in a timely fashion, with reasonable frequency.

Don't just be SMART ... BUILD:

For indicators to drive learning as well as fulfil upwards accountability and reporting requirements, indicators have not only to be technically robust (fulfilling the SMART criteria) but they also need to help those implementing the programme to measure, analyse and improve the impact that they are having on **target groups**. The social investor, the **Acumen Fund**, has proposed that indicators also need to conform to BUILD characteristics. Information collected through BUILD indicators is:

- ▶ **Bottom-up** – nurturing the habit of listening to programme partners and target groups in order to provide actionable insight on their needs and interests

DEFINITION

A **target group** comprises the specific individuals or organizations for whose benefit a development intervention is undertaken.

- ▶ **Useful** – yields data that is of sufficient quality to support decision-making
- ▶ **Iterative** – allows for learning, adaptation and replication
- ▶ **Light-touch** – uses low-cost tools and technologies that require a minimal investment in terms of time and money
- ▶ **Dynamic** – enables rapid data collection within a fast-changing environment.

Table 3.4: Examples of indicators

Category	Example result	unSMART indicators	The problem?	SMART indicators
Input	Two trainers and facility within budget of US\$10,000.	Two trainers complete at least two training courses each under budget.	Not relevant (describes the activity level of the results chain).	<ul style="list-style-type: none"> • Two trainers skilled, equipped and deployed. • Cost of programme in US dollars within desired budget.
Activity	Provide life skills training for youth (20 hours).	Training delivered.	Neither specific (not clear what information should be collected) nor measurable (no way of objectively verifying training delivery).	<ul style="list-style-type: none"> • Number of training hours delivered. • Number of youth participating by age, gender, level of education. • Date by which training was provided.
Outputs	100 youth participated in training.	Number of youth who finished the training (by age, gender, level of education).	Not time-bound (unclear when the information should be collected and assessed).	Number of youth who finished the training (by age, gender, level of education) at the end of each calendar month.
Outcomes	Increased knowledge of effective communication.	By the end of the programme: <ul style="list-style-type: none"> • number and percentage of youth increasing their self-confidence due to improved communication skills. 	Not attributable (self-confidence may be an indirect effect of skills training, but is not directly linked to intervention efforts).	By the end of the programme: <ul style="list-style-type: none"> • number and percentage of youth able to communicate effectively, measured against a predetermined communication scale. • number and percentage of youth with improved problem solving skills, measured against a predetermined problem solving ability scale.
Higher-level goal	Improved employment of youth aged 18–24.	Youth will find jobs more easily than they could before the intervention.	Neither specific (vague as to what information is required) nor time-bound.	Number and percentage of youth aged 18–24 who are in employment and receive above minimum wage in their field of training within three months of completing the programme.

Establish a baseline: The baseline tells us the value of an indicator at the beginning of, or, ideally, just prior to, the implementation period. Knowing the baseline value of indicators allows us to define realistic targets and track future progress against the initial situation. For example, if we want to monitor participants' incomes over time, data from our programme registration forms may tell us that the average monthly income of participants on entering the programme is US\$100. This is our baseline value that allows us to compare how incomes change during and after our intervention.

Define targets: If indicators are not assigned a time frame or magnitude of change, we will not know whether we are on track or have reached our objective. For example, if the desired outcome is increased household income, our indicator may be monthly earnings in US dollars. Then, the target may be set at a 30 per cent increase (magnitude) within three years (time frame). Each indicator should have no more than one target per specified period. If setting firm numerical targets is too arbitrary, targets can also be expressed as a range (see projecting results in box 3.2).

TIP

It is usually a good idea to pilot indicators during the early phases of an intervention before integrating them into the monitoring system. This will highlight how well they work in practice, and whether they actually capture the information that the project manager and other stakeholders are interested in collecting.

Ensure consistency: Although it is not always possible, in order to ensure consistent monitoring we should seek to keep the same indicators over the lifetime of an intervention. Having said that, it is not uncommon to add new indicators and drop old ones as we modify the programme design or streamline the monitoring system. However, it is important to retain a level of fidelity to our original objectives: if we find that our project will not achieve its original goal but will instead achieve some other goal (which may be of even greater value), we must acknowledge that factor in our reporting. Indicators accepted at the beginning of the intervention should not be changed unless objective criteria exist to justify the change.

Box 3.2: Projecting results

Projections are the results that are expected from an intervention, within a given time limit. They serve a number of important purposes.

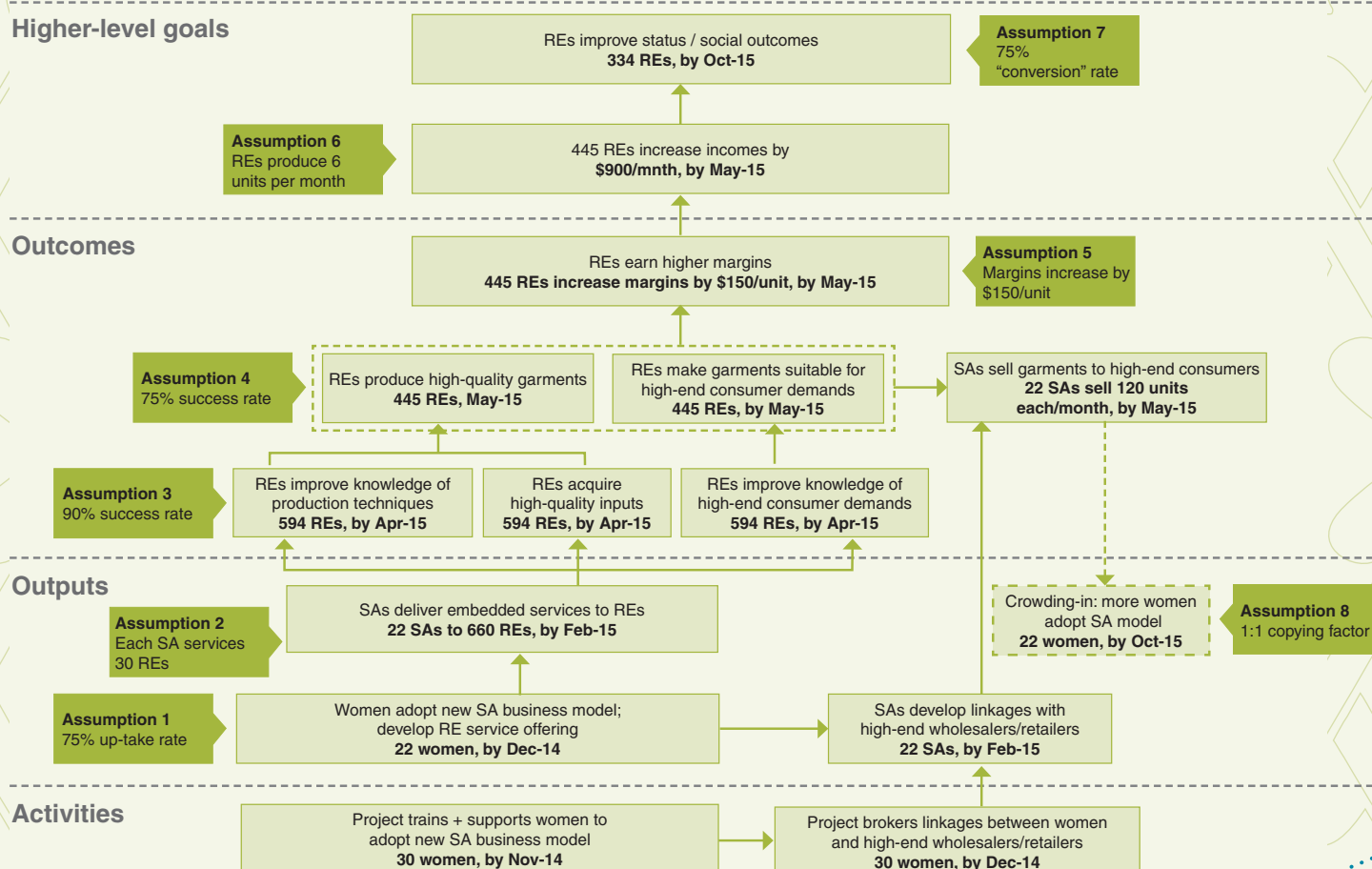
At the intervention-level, projections help implementing teams to think through when change is likely to happen, which is an important factor in effective measurement planning for monitoring activities. Projections and the assumptions on which they are based, also help to identify gaps in knowledge and flag areas for data collection. Projections can also feed into portfolio management decisions about where, when and how to intervene to achieve greatest sustainability and scale, and to make decisions about different interventions' relative value for money.

In order to be effective, projections should be made at the outset of an intervention. They are reasoned estimates, not wild guesses, of the changes we can expect to be brought about by intervention activities. Projections are made for every box in each intervention results chain against key indicators. As projections predict the change that will result from the intervention, they can be made for the two years following the end of the intervention. Each projection must be based on carefully thought-out assumptions and findings from market analysis and research, field observations or other credible sources, such as government data, relevant studies by development projects, etc.

It is important to note that projections are not targets. Targets tend to be fixed and denote funders' performance expectations. Projections should be regularly reviewed (at a minimum, twice a year) and updated where necessary to reflect new data collected and to clarify any assumptions.

An example of a projection against a results chain and assumptions, which anticipates the impact of an intervention on rural entrepreneurs (REs) and sales agents (SAs), is given in figure 3.3.

FIGURE 3.3: EXAMPLE OF A PROJECTION



Step 3: Select data collection tools

For each indicator, we need to choose how to collect information. In general, we can use two types of data: quantitative and qualitative.

Quantitative data come in numerical form and can be put into categories, ranked or ordered. Quantitative methods aim to provide objectively measurable data on demographic or socio-economic profiles, often using statistical techniques. They are usually based on standardized structured instruments that facilitate aggregation and comparative analysis. Common examples include tests, surveys and censuses.

Qualitative data come in non-numerical form and aim to provide an understanding of how

and why people think and behave the way they do. Qualitative methods seek to understand events from stakeholder perspectives, to analyse how different groups of people interpret their experiences and construct reality. Common examples of qualitative methods include unstructured or semi-structured interviews, focus groups and direct observation of participants. Qualitative methods tend to be quicker to implement than quantitative methods and are often less expensive.

Combining qualitative and quantitative data (applying a “mixed-methods” approach) is often recommended to gain a comprehensive view of the programme’s implementation and effectiveness.

WHICH DATA COLLECTION TOOLS?

A variety of different data collection tools can be used at each level of the results chain. Table 3.5 describes a range of common tools, along with their respective advantages and disadvantages. Direct observation and field visits can provide data for output indicators; for instance, the number of small businesses created. Measuring outcomes often requires a combination of formal surveys that provide reliable quantitative information as well as qualitative methods, such as key informant

interviews or focus groups, in order to understand the underlying mechanisms of whether and how certain effects were achieved. Finally, since higher-level outcomes usually relate to broader changes beyond the control of the project, official statistics can be useful when they are available for small geographic areas (such as municipalities) and can be disaggregated by socio-demographic characteristics.

WHEN TO COLLECT MONITORING DATA?

Many development programmes are deploying an “adaptive management” approach – shortening the feedback loop between activities and their effects by monitoring changes

in as near to real time as possible and flexibly adjusting plans based on experience.

Access to good-quality data has long been the factor constraining such rapid, iterative learning.

Table 3.5: Overview of data collection tools

Tool	Description	Use	Advantages	Limitations
Administrative and management records	Documents that provide information on project management processes	To examine the effectiveness of project management or strategy implementation	<ul style="list-style-type: none"> Provide information on process that is difficult to obtain through other means 	<ul style="list-style-type: none"> Programme specific, not generalizable Dependent on reliable management records systems
Field visits (combination of observation and interviews)	In-depth examination of a specific site or location	To monitor and understand context	<ul style="list-style-type: none"> High level of detail Access to observational data 	<ul style="list-style-type: none"> Programme specific, not generalizable Highly dependent on access to appropriate field sites
Key informant interviews	In-depth data collection method with highly informed individuals	To obtain specific and highly detailed information on a particular issue or set of issues	<ul style="list-style-type: none"> High level of detail Can address unanticipated topics Has flexibility to explore issues in depth Can capture a range of stakeholder perspectives 	<ul style="list-style-type: none"> Programme specific, not generalizable Quality is highly variable depending on interviewer skills and interviewee comfort
Focus groups	In-depth data collection method with informed members of a specific sub-population (e.g. women, youth, elderly workers)	To obtain specific and highly detailed information on stakeholder perspectives on a specific issue or set of issues	<ul style="list-style-type: none"> Same as for key informant interviews Allows for interaction with and among participants 	<ul style="list-style-type: none"> Programme specific, not generalizable Quality highly dependent on group dynamic (e.g. participants can be influenced by moderator or dominant group members) Interpretation challenges Time-consuming analysis
Direct observation	Method of collecting data through direct observation (e.g. classroom observation), information is recorded in a log or diary	To obtain naturalistic data	<ul style="list-style-type: none"> High level of detail from a neutral observer Provides information on actual behaviour rather than self-reported behaviour 	<ul style="list-style-type: none"> Not generalizable High potential for observer bias Interpretation and coding challenges
Review of official records	Official documents that provide background information or historical data on certain phenomena	To examine underlying processes or historical trends/data for certain phenomena	<ul style="list-style-type: none"> Provides information that may be difficult to obtain through other means Inexpensive 	<ul style="list-style-type: none"> Possible access restrictions Must verify validity and reliability of data Data may not be exactly what is needed
Mini surveys (often mobile-based)	Brief questionnaire/survey that collects limited data set on certain phenomena	To obtain quantitative data on a limited number of people or issues	<ul style="list-style-type: none"> Faster and less expensive than household surveys 	<ul style="list-style-type: none"> Limited scope and therefore usually not representative
Household surveys	An extensive set of survey questions whose answers can be coded consistently	To obtain information on a large number of respondents regarding their socio-economic status, demographic data, consumption patterns, etc.	<ul style="list-style-type: none"> Provides in-depth information on population of interest More generalizable than mini surveys May be designed to collect data of specific interest 	<ul style="list-style-type: none"> Expensive Requires special expertise to ensure validity Difficult to persuade people to respond to long questionnaire
Panel surveys	A longitudinal study in which variables are measured on the same units over time	Same as for household surveys, with particular interest in measuring changes over time	<ul style="list-style-type: none"> Same as for household surveys Can capture dynamics over a period of time 	<ul style="list-style-type: none"> Same as for household surveys May have problems with participant retention over time

However, a number of recent tech-based innovations have brought down both the costs and the time required for data collection (see box 3.3 and 3.4). The surge in mobile phone ownership in many parts of the developing world has made both SMS and voice calling popular means of surveying. These tech-enabled methods not only collect data in near-real time, they also significantly reduce the costs of measurement compared to face-to-face techniques. In Kenya, for example, the price of administering an SMS survey can be as little as \$0.17 per respondent, and \$7.17 for a voice call survey.²

There is value in using quick feedback mechanisms, especially tech-enabled ones, even

² Based on a standard 10-question SMS survey and a 30-minute call centre survey, excluding labour costs. Prices from EngageSpark, correct as of January 2018.

TIP



Use quantitative methods when:

- numerical or generalizable data are required to convince decision-makers or make decisions regarding scaling-up of a programme
- you need statistically representative information about the target population, their situation, behaviours and attitudes.

Use qualitative methods when:

- “how and why” questions need to be understood; that is, when quantitative data need to be explained by motivation and attitudes affecting behaviours
- participatory approaches are favoured.

Box 3.3: Harnessing technology for real-time data

The social investor Acumen leverages mobile surveys as part of their Lean Data Service. Lean data (a type of mini survey, as described in table 3.5) aims to collect meaningful monitoring data while making limited demands on the respondent's time and attention. In 2015, Acumen worked with Edubridge, a vocational training company that seeks to improve labour market outcomes for workers in India who are migrating from rural to urban areas. The company wanted to know the answer to a question critical to their theory of change: How do “successful” trainees – those who are offered and accept job placements immediately after they undergo Edubridge training – differ from less “successful” trainees?

Acumen Lean Data conducted a phone-call-based survey of several discrete populations: people who had expressed an interest in Edubridge courses but had never signed up for one; people who had completed an Edubridge course but had not accepted a job offer that they had received afterwards; and people who had both completed a course and accepted a job offer. The project took just four months. Existing Edubridge call centre operators acted as enumerators, setting aside one hour of their time per day for survey calls. They completed a total of 650 calls, and each call lasted seven to eight minutes.

The results provided rich insight into Edubridge outcomes. The theory of change had hypothesized that trainees with close friends in urban areas would be more likely to accept jobs than other trainees. This turned out to be true: trainees who had friends in a city where a job was located were 21 per cent more likely to take that job than trainees who had no friends there. Another hypothesis was that trainees from higher-income families would be more likely to accept jobs than trainees from lower-income families. That turned out not to be true. Those who had accepted jobs were 8 per cent poorer than those who had not. The company is now using data from the survey to shape its strategy as it prepares to expand its operations to 100 training centres over the coming years.

Source: Adapted from [Dichter et al. \(2016\)](#).

when outcomes are more predictable, as they can bring down the costs of data collection. This approach is particularly suited to fragile, fast-moving and complex operational environments – or innovative intervention models – where the evidence underpinning the theory of change is more uncertain. However, certain studies have shown that over-reliance on mobile surveys in particular can lead to bias against more marginalized/vulnerable groups, who may not have access to mobile phones or good reception. Projects should therefore start with an assessment of likely mobile phone penetration in target communities.

Key decision points:

- ▶ Is the programme operating in a data-constrained environment?
- ▶ Is the theory of change less “proven” and eventual outcomes more uncertain?
- ▶ Are mobile phones, and particularly smartphones, widely used in the country/community?

TIP



The timing of data collection should be planned according to local realities so that collection does not impose a burden on an individual or a family. Data collection should not coincide with youth taking school exams, for example, or when young people’s labour is needed during particular agricultural seasons.

Recommendations:

- ▶ Consider tech-enabled monitoring wherever possible, but especially where physical access to target groups is constrained and would come at prohibitive costs, for example when there are large distances between beneficiaries.
- ▶ But carefully consider whether exclusively mobile-based monitoring will lead to vulnerable groups being excluded. Consider mixing different data-collection tools, such as more “traditional” face-to-face interviews and field observation with remote digital data collection.

Box 3.4: Mobile data collection is cheaper ... but is it reliable?

In Kenya, the World Food Programme wanted to see whether mobile phones could be used to collect reliable information on nutrition indicators. They conducted an experimental trial to see whether the mode of data collection affected survey responses. Comparing computer-assisted telephone interviewing (CATI – a telephone surveying technique in which the interviewer follows a script provided by a software application) with face-to-face (F2F) interviews, they found a number of advantages and disadvantages.

Advantages:

- higher response rates with CATI
- one-third the cost of F2F per survey (\$5 vs. \$16)
- more feasible to collect data from insecure areas.

Disadvantages:

- higher risk of sampling bias: women without mobile phones had fewer assets overall
- patchy network coverage
- more difficult to collect a range of social outcomes, especially regarding more sensitive topics.

Source: https://ec.europa.eu/jrc/sites/jrcsh/files/hachhethu_kusum_day3_session9_room1.pdf

WHO COLLECTS DATA?

It is important to clearly define data collection responsibilities. In practice, different types of monitoring will fall under the responsibility of different actors, both in the field and at headquarters. The following people could be recruited as data collectors in various circumstances:

- ▶ programme managers
- ▶ local project team members or M&E officers
- ▶ local implementing partners (e.g. teachers, training providers, loan officers)
- ▶ beneficiaries
- ▶ other local stakeholders (including parents and community members)
- ▶ volunteer enumerators (e.g. university students)
- ▶ external consultants
- ▶ survey firms.

While defining the responsibilities for gathering data, clarify what happens to the information, once collected. Integrate data collection plans with procedures for storing, aggregating and analysing the data to guarantee that those who need the information have timely access to it.

A number of “full package” providers are available, who offer tech-enabled solutions and often use off-the-shelf surveys and question banks to automate data collection, aggregation and presentation. Service providers relevant to data collection on employment outcomes are ULULA, a multi-language supply chain management, stakeholder involvement and M&E software for responsible supply chains, Laborvoices for supply chain management, which includes feedback from workers on suppliers, and Laborlink, an on-line worker-centric solution that uses direct

TIP



Be mindful of conflicts of interest when assigning responsibilities for collecting and reporting information. For example, teachers or training providers may have an incentive to cheat when recording outputs (such as the number of hours of training conducted) or outcomes (such as the number of youth who improved their test scores or found a job). To ensure data reliability, we recommend (1) using neutral observers to ensure independent monitoring, and (2) verifying the accuracy of information provided, at least sporadically, through unannounced site visits or other means.

worker feedback to identify and evaluate factory improvements. Other service providers offering technology-enabled data collection include:

- ▶ <https://www.engagespark.com/> (voice, SMS and social media)
- ▶ <https://telerivet.com/> (SMS and voice)
- ▶ <https://www.ushahidi.com/> (crowd-sourced mobile surveys)
- ▶ <http://www.frontlinesms.com/> (SMS texts)
- ▶ <https://www.echomobile.org/public/main> (SMS, voice, sensors, web)
- ▶ <https://viamo.io/> (mobile phone surveys).

To learn more about participatory monitoring and evaluation, consult Sabo Flores (2008), Powers and Tiffany (2006) and Chambers (2007).

For an example of how photo monitoring improved teacher attendance and reduced the need for monitoring visits in India, see Duflo et al. (2012).

Step 4: Deploy the system to aggregate, analyse and report on results

The key parts of the monitoring system can be brought together in the form of a **logical framework** matrix. This popular project management

tool summarizes the results chain, indicators, data collection tools and assumptions. An example is included in table 3.6.

AGGREGATING RESULTS

Depending on the complexity of the programme, data may need to be aggregated: that is, compiled into a single location to aid data processing and analysis. Data can be aggregated at the level of an intervention (for example, different data on the same indicator, but from different training cohorts or geographic locations) or across a programme (for example, the number of jobs created for youth across different interventions).

The project team must decide on the best ways of organizing the data to conduct efficient analysis. For most projects, a simple Excel workbook will suffice. To facilitate analysis and reporting in bigger programmes, it may be advisable to set up a **management information system** that connects all the databases used by different programme units.

If the data collected is particularly complex, it may be beneficial to employ an experienced

database manager to develop codes and procedures that allow multiple users to query the data and derive results with a limited amount of training. A variety of database systems are appropriate for this purpose, and the project should select a software program that offers a balance of analytical sophistication and user-friendliness.

► For **qualitative data**, computer-based qualitative analysis software can be used. There are many brands to choose from (such as Atlas.ti, NVivo or MaxQDA), and each works in a similar way. Software for qualitative analysis allows the user to import all relevant documents (such as transcripts from interviews and focus groups, project documents and photographs) and then apply a set of predetermined codes. Depending on the user's level of training, the codes can function as an organizing tool (grouping all similar topics from various sources) or allow sophisticated analysis that identifies relationships within these topics.

DEFINITION

A **logical framework** (or "logframe") is a management tool used to improve the design of interventions, usually at the project level. It involves identifying strategic elements (inputs, outputs, outcomes, higher-level goals) and their causal relationships, indicators and the assumptions or risks that may influence success and failure.

DEFINITION

A **management information system** is the combination of computer technology, people and procedures put in place to collect, organize and analyse information in order to support decision-making. It allows large amounts of data to be managed centrally and comparison of indicators both by beneficiary characteristics and over time.

Table 3.6: Logical framework for an entrepreneurship promotion programme (example taken from the ILO's Start and Improve Your Business programme)

Key components of the results chain					
	Indicators	Sources	Frequency	Assumptions	
Longer term goals	<ul style="list-style-type: none"> Additional employment created Quality of employment improved Labour demand improved 	<ul style="list-style-type: none"> Administrative data Employee surveys 	e.g. yearly	<ul style="list-style-type: none"> Businesses meet consumer demand Adequate regulatory, economic and institutional environment Start-ups benefit from investment/credit 	
Outcomes	<ul style="list-style-type: none"> Improved core work skills More and better business plans Better business practice New start-ups established Increased business performance and competitiveness 	<ul style="list-style-type: none"> Participant surveys Employee surveys Administrative data 	e.g. bi-annually	<ul style="list-style-type: none"> Participants learn from training Training prompts expected behavioural change Credit/grant is used for enterprise Credit agency/franchisor does not exploit entrepreneur 	
Outputs	<ul style="list-style-type: none"> New curriculum available Trainers trained Demand-driven training services delivered 	<ul style="list-style-type: none"> Programme monitoring data Participant registry 	e.g. monthly	<ul style="list-style-type: none"> Content, intensity and delivery of the training is tailored to the needs of the target group Correct group is targeted and participates in complete training 	
Activities	<ul style="list-style-type: none"> Carry out market assessment Develop training of trainers Train and advise training providers on how to deliver training services Link businesses to microfinance organizations Entrepreneurship training, advice and mentoring 				
Inputs	<ul style="list-style-type: none"> Budget, staff, trainers, partnership, facilities, supplies, technical expertise, training tools 				

- For **quantitative data**, several different systems can be combined. One should use a relational database, such as Microsoft Access. Relational databases facilitate the investigation and display of data along several different variables. Typically, however, the analyses performed in relational databases are descriptive in nature, providing measures of central tendency (e.g. means, modes, medians, standard deviations). If the project demands more sophisticated

analysis, and the instruments are designed and administered to allow it, M&E staff might use a statistical software package, such as SPSS or Stata. In addition to commonly available statistical software packages that are based on the hard drive of a single computer, there is also an increasing use of “cloud”-based data management and analysis systems, which allow a large team to collaborate on monitoring and analytical tasks (see previous sub-section).

DATA ANALYSIS

Monitoring has little value if we do not learn from and act on the data that we gather. A constant cycle of reflection and actions helps to keep interventions responsive to often-dynamic implementing contexts and the shifting needs of project beneficiaries. Organizations and projects stagnate when they don't learn and rigorous monitoring forces us to keep learning (Shapiro, 2003).

Monitoring processes often produce an array of data but little actionable knowledge. The step of interpreting and analysing the data is therefore vital to help ensure that monitoring generates useful information which supports evidence-based decision-making.

Consider data collected about the knowledge and attitudes of youth who participated in a skills development training course. Survey responses have been collected and processed, and the data entered and tabulated in an Excel file by project team staff. To analyse this data and turn it into useful information (see an example in box 3.5), some “tips” for the project team include:

- **Analyse information throughout the research period, including during data gathering:** For example, if survey results came only from one particular district: does the limited amount of data (only 25 per cent of the respondents were able to accurately recall key learning outcomes) already signal a problem with the

training product? Does this trigger a need for more in-depth data collection?

- **Structure analysis around the results chain:** The project team should take a specific box in the theory of change as the “entry point” for analysis and check to see if their logic still holds true (e.g. “training participants demonstrate improved technical skills”).
- **Analysis should include not only problems but also potential solutions:** Avoid simply producing a list of problems without identifying any avenues for follow-up/potential solutions. Analysis should explore all the problems identified and shorten the list to highlight a few key issues that must be addressed/discussed.
- **Make sure that there are sufficient data to underpin the analysis:** If the information is weak, then say so, and be careful not to draw conclusions which are not justifiable. For example, even if the project team collects an initial round of surveys that show low levels of knowledge among training participants, the sample size (of five respondents) or locality (restricted to one district) might be too small to allow accurate conclusions to be drawn.
- **Triangulate the data using multiple data sources to verify findings and develop a more robust basis for decision-making:** The findings from one data source may prompt questions that require further

investigation, perhaps using other tools. For example, if survey results show only modest interest in seeking employment among youth, the project could conduct focus group discussions/in-depth interviews with a small sample of respondents to

determine the reason (if this is not clear from the survey results).

A detailed discussion on evidence uptake and policy formulation related to youth employment is available in Note 7.

REPORTING ON RESULTS

Different stakeholders will have different data reporting demands. Typically, the higher our audience is in an organization’s hierarchy, the less detail we need to provide and more we need to summarize findings. Presenting clear messages, substantiated by aggregated data, and concise information tends to be more appropriate for high-level audiences, who are mainly interested in the big picture.

We can tailor the format of our reports to suit each audience (see table 3.7).

Monitoring data should always be reported in comparison with their baseline and target values and presented in a straightforward and easy to understand manner. Visual tools, such as graphs, charts and maps, can be very useful in highlighting key data and messages.

Table 3.7: Tailoring reports to our target audience

Target audience	Format	Timing/frequency
Project staff	Oral presentation and written summary statistics at team meetings	Weekly
Management team	Written reports and oral presentation	Monthly
Partners	Oral presentation and written summary statistics	Monthly
Donors	Depends on donor requirements. Usually short written reports highlighting project progress, issues experienced, outcomes and impact, efficacy of intervention/strategy, etc.	Quarterly/biannually

Box 3.5: Bringing monitoring data to life

Kuza is a UK Department for International Development (DFID) project that takes a market systems approach to expanding employment opportunities for poor and marginalized youth in targeted sectors in Mombasa County, Kenya. It aims to increase the employability of 8,000 young people over the life of the project, while creating at least 5,000 sustainable jobs.

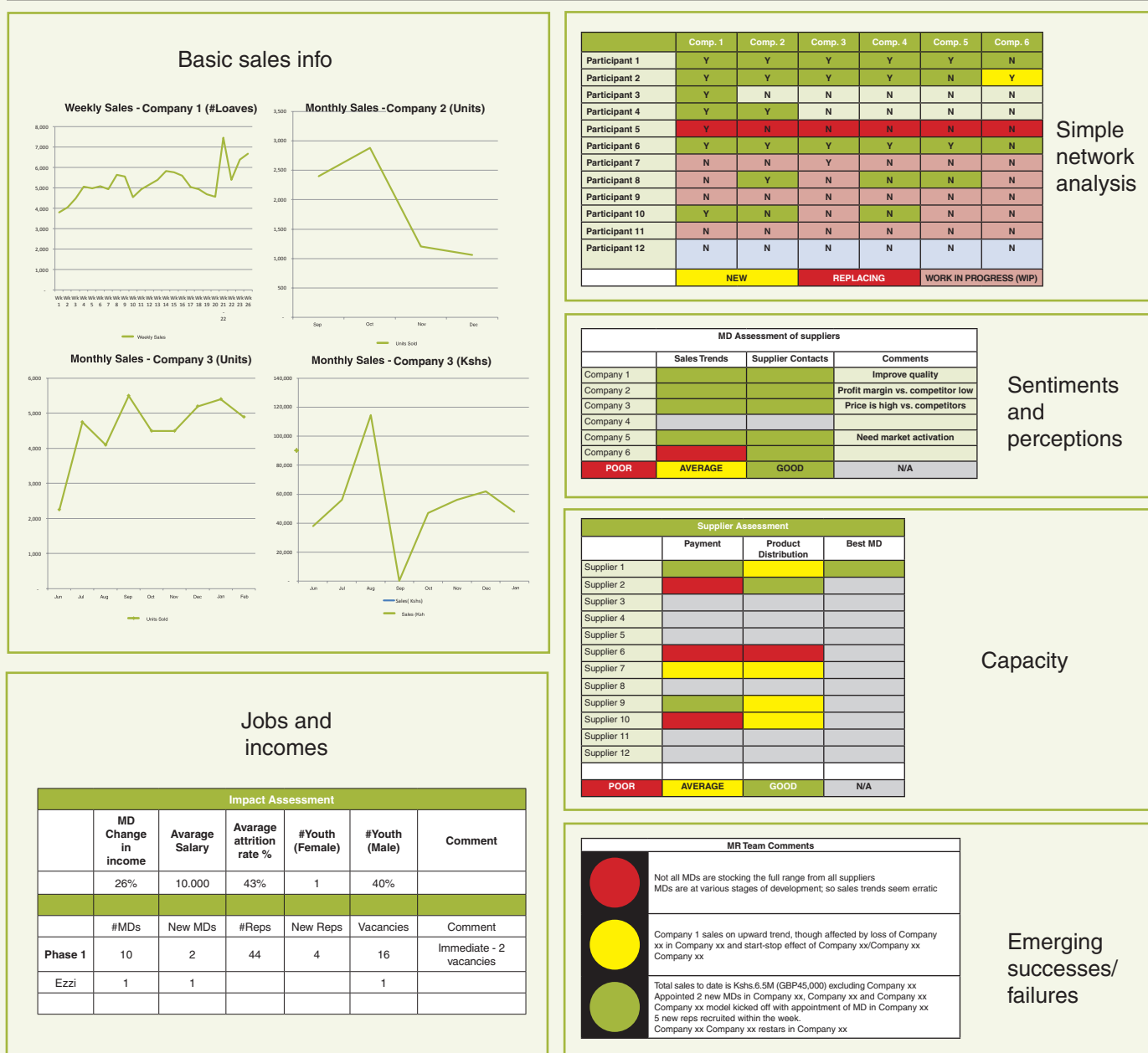
The project’s early analysis laid out the youth employment challenge in Mombasa. The 44 per cent unemployment rate among youth was higher than the national average, and there was also a significant mismatch in the types of skills that youth were being taught in existing educational institutions and those required by the job market.

Kuza adopted a sectoral focus on manufacturing, waste management, investment promotion and micro-retail. Operating in a dynamic context where access to good data proved difficult, Kuza collected monitoring data in as near real time as possible to allow it to adapt and evolve its intervention design during the implementation phase. This involved:

- developing hypotheses and running short-term experiments (called “micro pilots”) to prove or disprove hypotheses
- quickly gathering useful information about the hypotheses and any unexpected developments
- drawing meaningful conclusions and adapting behaviour
- reacting quickly to identify and build on emerging good practice and limit/learn from failures.

To aid in implementing this approach, Kuza developed a simple dashboard for monitoring data, accessible to all staff (see figure 3.4). The idea was to integrate basic market information (sales data) with target group data (on jobs and incomes) to inform rapid decision-making.

FIGURE 3.4: DASHBOARD FOR DATA MONITORING DURING HALF-DAY CLASSROOM-BASED TRAINING



Source: ILO (2016) and DCED (2016).

KEY POINTS

1. **Every intervention needs to have a solid monitoring system to allow continuous tracking of implementation and results.** Effective monitoring helps to detect problems in a timely manner and allows corrective actions to be taken. Monitoring also provides the foundation to evaluate an intervention.
2. **Underlying any programme design is a theory of change, which can be expressed as a diagram showing how a programme plans to impact its beneficiaries.** It also shows the set of assumptions we make about why these particular project activities will foster positive change. Programme managers and key stakeholders need to collaborate to develop the theory of change and translate it into a results chain, in order to provide stakeholders with “a logical, plausible sequence” of how the planned intervention will lead to the desired results.
3. **Indicators are required at each level of the results chain.** They help us to understand whether what has been planned is actually happening. Indicators are a crucial element of a monitoring system because they drive all subsequent data collection, analysis and reporting. For each indicator, we need to choose the appropriate method of collecting information. In general, we can use a combination of quantitative (numerical) and qualitative (non-numerical) data. A variety of different data-gathering tools can be used at each level of the results chain, from simple low-cost observation to complex high-cost surveys.
4. **The monitoring system provides continuous information on the direction, pace and magnitude of change.** It also allows us to identify unanticipated developments in the project or its environment. This provides the foundation for determining whether an intervention is moving in the intended direction and makes good monitoring critical to effective project management. To achieve this aim, data must be systematically aggregated, analysed and reported.

KEY RESOURCES



- ▶ Donor Committee for Enterprise Development. 2016. *The DCED Standard for Measuring Results in Private Sector Development. Control Points and Compliance Criteria. Version V.*



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



- ▶ ILO. 2016. *Development Cooperation Manual* (Geneva).



- ▶ Roberts, D.; Khattri, N. 2012. *Designing a results framework for achieving results: A how-to guide* (Washington, DC, Independent Evaluation Group/World Bank).

REFERENCES

- ▶ Baker, J.L. 2000. *Evaluating the impact of development projects on poverty: A handbook for practitioners* (Washington DC, World Bank)
- ▶ Chambers, R. 2007. *Who counts? The quiet revolution of participation and numbers*, IDS Research Summary of IDS Working Paper 296 (Brighton, Institute of Development Studies).
- ▶ Creswell, J. 2008. *Research design: Qualitative, quantitative and mixed methods approaches* (Thousand Oaks, CA, Sage).
- ▶ Donor Committee for Enterprise Development. 2016. "Measuring better and more inclusive jobs", a presentation at the 2016 DCED Global Seminar, Bangkok, March 14–16.
- ▶ Dichter, S.; Adams, T.; Ebrahim, A. 2016. "The power of lean data", *Stanford Social Innovation Review*, Winter.
- ▶ Duflo, E.; Hanna, R.; Ryan, S.P. 2012. "Incentives work: Getting teachers to come to school", *American Economic Review*, Vol. 102, No. 4, pp. 1241–1278
- ▶ 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).
- ▶ Gosparini, P.; Russo, L.; Sirtori, M.; Valmarana, C. 2003. *The monitoring and evaluation manual of the NGOs of the Forum Solint* (Rome, Development Researchers' Network).
- ▶ International Labour Organization (ILO). 2016. *Crafting Kuza: Towards a systemic approach to job creation for youth in Mombasa* (Geneva).
- ▶ Kusek, J.Z.; Rist, R.C. 2004. *Ten steps to a results-based monitoring and evaluation system: A handbook for development practitioners* (Washington, DC, World Bank).
- ▶ Majurin, E. 2014. *Start and improve your business (SIYB): Implementation guide* (Geneva, ILO).
- ▶ Powers, J.L.; Tiffany, J.S. 2006. "Engaging youth in participatory research and evaluation", *Journal of Public Health Management Practice*, November (Suppl.), pp. S79–S87.
- ▶ Sabo Flores, K. 2008. *Youth participatory evaluation: Strategies for engaging young people* (San Francisco, CA: Jossey-Bass).
- ▶ Samarth. 2015. *Results measurement system: User manual*, version 2.0 (SAMARTH - Nepal Market Development Programme).
- ▶ Shapiro, J. 2003. *Monitoring and evaluation* (Washington, DC, CIVICUS).



Case study:

**ESTABLISHING A MONITORING SYSTEM
FOR THE JORDAN RURAL ECONOMIC
GROWTH AND EMPLOYMENT PROJECT**

This case study is based on the Rural Economic Growth and Employment Project, final project design report (IFAD).

Learning objectives

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

- ▶ identify key programme results and translate them into an intervention logic in the form of a visual theory of change, complete with underlying assumptions
- ▶ select appropriate measurement tools to capture qualitative and quantitative change
- ▶ identify constraints to effective data analysis and suggest solutions to overcome them.

Introduction and case study context

Economic growth over the past decade has not led to a significant increase in the number of jobs available to Jordanians. The unemployment rate is estimated at around 15 per cent, and among young people, aged 15 to 24 years old, unemployment stands at 36 per cent as of 2017.³ Almost 90 per cent of all the unemployed are under the age of 40.

Jordan's agricultural sector has been experiencing a shift from traditional labour-intensive methods to more modern, mechanized capital- and water-intensive systems. However, agriculture's contribution to GDP has declined, from 8 per cent of GDP in 1990 to 3.4 per cent in 2012. It is estimated that 20 per cent of the population depends on agriculture for some part of their income and that the sector is the main source of income for 7–8 per cent of Jordanians.

Studies have identified a range of high water value crops, which Jordan can produce at prices which are competitive on the export market, particularly during winter (which is the

off-season for European producers). These crops can create employment along the value chain and much of the work involved is particularly suited for rural women. There is an urgent need to support job-creating agricultural growth.

To this end, and to support the Government's Poverty Reduction Strategy 2013–2020, the International Fund for Agricultural Development (IFAD) funded an agricultural development initiative called the Rural Economic Growth and Employment Project (REGEP) to be implemented by the Jordan Enterprise Development Corporation (JEDCO). Launched in 2015, REGEP runs over 6 years with a budget of US\$11.34 million. The main outcomes are to enhance the technical capacity and competitiveness of smallholder farmers and rural micro, small and medium-sized enterprises (MSMEs) and to increase access to sustainable and readily accessible rural financial services.

³ ILO modelled estimates for 2017, available at: www.ilo.org/ilostat.

Part I: Define the intervention logic

In line with the Government of Jordan's poverty reduction goal, REGEP aims to create a wide range of financial and non-financial benefits for the target group, including increased incomes, empowerment and institutional strengthening, training and capacity building, access to finance, value chain linkages and job opportunities.

To achieve this goal, the programme aims to target smallholders, rural entrepreneurs and the unemployed and underemployed in rural areas directly, through existing and new associations, Saving and Credit Groups (SCGs), and through MSMEs which create employment in rural areas, in order to:

- ▶ build their technical and business capacity
- ▶ increase access to rural financial services
- ▶ build win-win value chain linkages and enhance access to high value domestic

and export markets through marketing initiatives

- ▶ improve post-harvest handling, quality standards and certification.

As a direct consequence of the above, REGEP expects to:

- ▶ improve access to finance in rural areas through enhancing the technical capacity and competitiveness of small farmers and small and medium-sized agricultural companies
- ▶ integrate smallholder farmers in value chains
- ▶ create employment opportunities in rural areas for women and youth
- ▶ contribute to economic growth and increase income
- ▶ increase agriculture exports of fruit and vegetables.



Discussion topics

1. Based on what you now know about REGEP, how would you formulate the main programme's outcomes and outputs – as well as its higher-level goal?
2. A results chain is a diagram that shows how a programme plans to impact its beneficiaries. Can you draw a results chain for the REGEP programme?
3. A good results chain and underlying theory of change also includes the assumptions we make about why this particular logical sequence of programme results will foster positive change. Can you identify some key assumptions underlying the REGEP logic?

Part II: Selecting data collection tools

Based on their theory of change, REGEP came up with a set of indicators to measure programme progress. At the outcome level, these included:

- ▶ the percentage increase in farm gate value of sales for smallholders
- ▶ the percentage increase in value of sales for supported MSMEs
- ▶ improvement in access of poor to financial services
- ▶ level of effectiveness of pro-poor policies.

And at the output level:

- ▶ number of smallholders participating in “farmer field schools”
- ▶ number of MSMEs benefiting from business mentoring
- ▶ the amount of disbursements for value chain upgrading investments
- ▶ number of “Global GAP” certificates issued
- ▶ value chain working group established and operating
- ▶ number of policy briefs prepared and disseminated.



Discussion topics

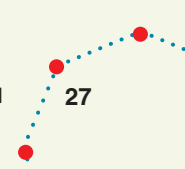
1. REGEP has asked for support in choosing suitable data collection tools to gather information on each of the indicators. Pick two

outcome indicators and two output indicators and complete the table below.

Indicator	Data collection tools	Frequency (when and how often to collect data)	Responsibility (implementing partner, M&E officer, etc.)

2. No measurement tool is “perfect” – each has its respective advantages and disadvantages. Refer to table 3.5 in Note 3. What are some of the possible disadvantages of the data collection tools you have selected – and how would you

overcome them? Add two more columns to the table you used in question 1 to list the disadvantages of each data collection tool and possible risk mitigation strategies to overcome them.



Indicator	Data collection tools	Responsibility (implementing partner, M&E officer, etc.)

Part III: Data analysis and interpretation

Before beginning implementation, REGEP set up an M&E system with several components, including:

- ▶ *Output monitoring:* Focusing on physical and financial inputs, activities and outputs. Data will flow directly from records at different levels and will be monitored quarterly. Annual Work Plans and Budgets provide yearly targets for first level monitoring.
- ▶ *Outcome monitoring:* Assesses the use of outputs and measures their benefits at beneficiary, association, enterprise and value chain levels. Performance indicators will be compared with historical values and expected targets.

The REGEP was implemented through an array of partners. To ensure a consistent approach to data collection and analysis, a mandatory reporting system was established. This includes three sets of reports:

1. a quarterly progress report by each implementing partner, consolidated at the level of the Project Management Unit (PMU)
2. a biannual progress report by the PMU, and
3. an annual progress report by the PMU.

Each report is then submitted to the Programme Steering Committee for approval and then to IFAD and the Government.



Discussion topics

1. Monitoring systems often produce an array of data but little actionable knowledge. The process of interpreting and analysing the data is therefore vital to help ensure that monitoring generates useful information which supports evidence-based decision-making. What do you think are the two main strengths and two main challenges facing the REGEP's M&E system in terms of effective data analysis?
2. REGEP wants to decide on the best way of organizing data to conduct efficient analysis. What technology, systems and processes do you recommend that they use to manage the large amounts of data coming into the PMU?

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Note

4

Enhancing youth employment learning through evaluation



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

NOTE 4
**Enhancing youth employment learning
through evaluation**

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. 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 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 4: Enhancing youth programme learning through evaluation/*International Labour Office. Geneva, 2018.

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

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

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

Printed in Switzerland

Contents

Diagnosing, planning and designing youth employment interventions	1
Why conduct an evaluation?	2
Assessing the evaluability of our intervention	6
Evaluation criteria and evaluation questions	7
Linking evaluation questions to evaluation types	10
Monitoring (no evaluation)	11
Performance evaluation	11
Impact evaluation	14
Cost-effectiveness and cost-benefit analyses	15
Understanding evaluation needs based on the operational context	17
Timing	17
Resources	19
The political context	19
Key points	22
Key resources	22
References	23
Case study: Developing terms of reference for a mid-term evaluation of a youth employment project in Egypt	24
Figures	
4.1 Benefits of conducting evaluations	3
4.2: Different evaluation types	10
4.3: Key steps in planning and managing an evaluation	13
4.4: Overview of typical programme phases	18
Tables	
4.1: Evaluation criteria	7
4.2: Examples of evaluation questions formulated along a results chain	9
4.3: Overview of the main evaluation types	21





Enhancing youth programme learning through evaluation



Prerequisites:

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



Learning objectives:

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

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



Keywords:

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

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

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

Why conduct an evaluation?

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

Evaluations can help project management:

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

DEFINITION

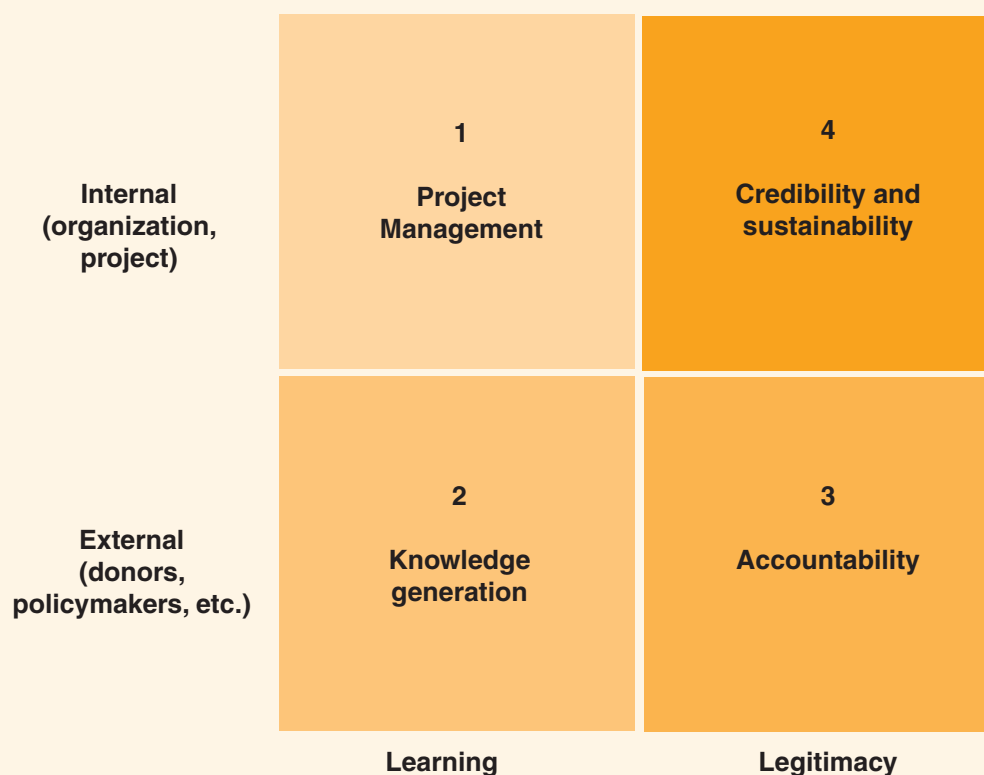
Evaluation is an evidence-based assessment of strategy, policy or programme and project outcomes, by determining their *relevance, impact, effectiveness, efficiency and sustainability*. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned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

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

evaluations allow us to show the true value of our work and inform the design and planning of other interventions.

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

FIGURE 4.1: BENEFITS OF CONDUCTING EVALUATIONS



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

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

Evaluations fulfil accountability requirements: Evaluations expose programmes to external scrutiny and hold programmes to account over the achievement of results and use

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

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

Box 4.1: Youth Employment Evidence Gap Map

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

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

All evaluation results convey learning:

While, positive evaluation results can be used in advocacy efforts to obtain greater support from donors, governments and the general public, negative findings are arguably even more useful in helping to iterate and refine intervention modules. As in any other field, building successful products and services

requires testing, prototyping and adapting to local circumstances. Failures are a necessary step toward state-of-the-art programming. And, especially when evaluation is not simply asking what works or doesn't work but also how and why certain approaches work, negative results will improve both programme design and ongoing operations. If, early in the process, we are able to understand the problems that may reduce the effectiveness of our intervention, then we are in a good position to build successful projects in the long run.

Box 4.2: The link between programme design and evaluation

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

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

Assessing the evaluability of our intervention

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

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

Furthermore, such a scoping exercise can assess the presence of a well-developed results measurement framework and the availability and quality of baseline data for assessing

TIP



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

changes (Note 3). For impact evaluations, these assessments also consider how likely it is that an impact study will lead to real improvements in programme performance and success and, thus, whether the costs in terms of effort and money will realistically be outweighed by the anticipated benefits. Following this assessment, the scope, methods and timing of the evaluation can be adjusted accordingly (ILO, 2017).

DEFINITION

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

Evaluation criteria and evaluation questions

As already mentioned, evaluations are periodic assessments of the **relevance, effectiveness, efficiency, impact and sustainability** of our intervention. These are the common

evaluation criteria as originally defined by the Organisation for Economic Co-operation and Development (OECD) (see table 4.1 for more detail).

Table 4.1: Evaluation criteria

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

Source: Based on [OECD/DAC, 2002](#) and [ILO, 2017](#).

Formulate evaluation questions: In practice, we may have many evaluation questions across all criteria that we would like to assess. While evaluation criteria are formulated in a general manner, evaluation questions need to be adapted and tailored to the

specific intervention. This also helps to define the menu of appropriate monitoring and evaluation (M&E) tools that will allow those questions to be answered ([GAO, 2012](#)). There are several important considerations when formulating evaluation questions:

DEFINITION

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

DEFINITION

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

1. **Involve stakeholders:** Before getting started, it is important to identify the audience for the evaluation and what that audience wants to know. The set of stakeholders depends on the intended goals of the evaluation (see the section “Why conduct an evaluation?” above) and typically includes project managers and staff members in the originating organization as well as key national and local partners (for example, line ministries and social partners involved) as well as donors. Involvement of our target audience to jointly identify evaluation questions is important for the assessment of all evaluation criteria and particularly helpful for evaluating the relevance and sustainability of an intervention.
 - ▶ **normative questions** compare what is taking place to what should be taking place. They compare the current situation with the specific objectives and targets that have been defined (*Has our project been implemented as intended? Is it performing as expected?*)
 - ▶ **cause-and-effect questions** examine outcomes and try to measure the difference that an intervention makes. They ask whether objectives have been achieved *as a result of* our project (*To what extent can we attribute observable change to our intervention?*).

2. **Choose complementary questions for each evaluation criterion:** All types of evaluation questions examine different aspects of the project and provide different kinds of information. Rather than being substitutes, they often complement or build on each other. For most assessed evaluation criteria, it makes sense to ask a mixture of *descriptive, normative and cause-and-effect* questions (Imas and Rist, 2009):
 - ▶ **descriptive questions** seek to define processes, conditions, organizational relationships and stakeholder views (*What is going on in our project?*)

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

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

	Inputs	Activities	Outputs	Outcomes	Higher level goals
Descriptive	<ul style="list-style-type: none"> How does the cost of the programme compare to similar interventions? What qualifications do the service providers have? What other ongoing interventions are there? 	<ul style="list-style-type: none"> Do youth know about the programme and how they qualify to join? What delivery mechanisms are being used? To what extent does the programme implementation differ by site? 	<ul style="list-style-type: none"> How many youth participate (by age, sex, etc.)? Who drops out? Which services are used the most? 	<ul style="list-style-type: none"> Are participants satisfied with the programme? Are there any observable changes in participant's basic, technical or core skills? How many programme participants find employment within 3 months? 	<ul style="list-style-type: none"> What are the current local youth unemployment rates? How high is the average household income?
Normative	<ul style="list-style-type: none"> Do we spend as much as we have budgeted? Are the staff and financial resources adequate? Is the programme duplicating other efforts? 	<ul style="list-style-type: none"> Is the process for selecting participants fair and equitable? Is the programme implementation delayed? Are operational manuals being followed? 	<ul style="list-style-type: none"> Do we achieve the desired gender balance in participants? Will we reach the goal of training 5,000 youth per year? 	<ul style="list-style-type: none"> Does participant income increase by 20%, as planned? Do 80% of beneficiaries find a job within 3 months of graduation, as required? 	<ul style="list-style-type: none"> Is local youth unemployment falling, compared to the programme start? Are household incomes evolving? Are more households becoming self-sufficient?
Cause-and -effect	<ul style="list-style-type: none"> n/a 	<ul style="list-style-type: none"> n/a 	<ul style="list-style-type: none"> n/a 	<ul style="list-style-type: none"> As a result of the job training, do participants have higher paying jobs than they otherwise would have? Do 80% of beneficiaries find a job within 3 months of graduation due to their participation in the programme? Does including internships increase the effectiveness of technical training offered? Does the programme affect boys and girls differently? What, if any, unintended positive or negative direct effects are there? 	<ul style="list-style-type: none"> Does the project contribute to reducing poverty in the area? What other (positive or negative) impacts does this intervention have on the living conditions of the wider community?

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

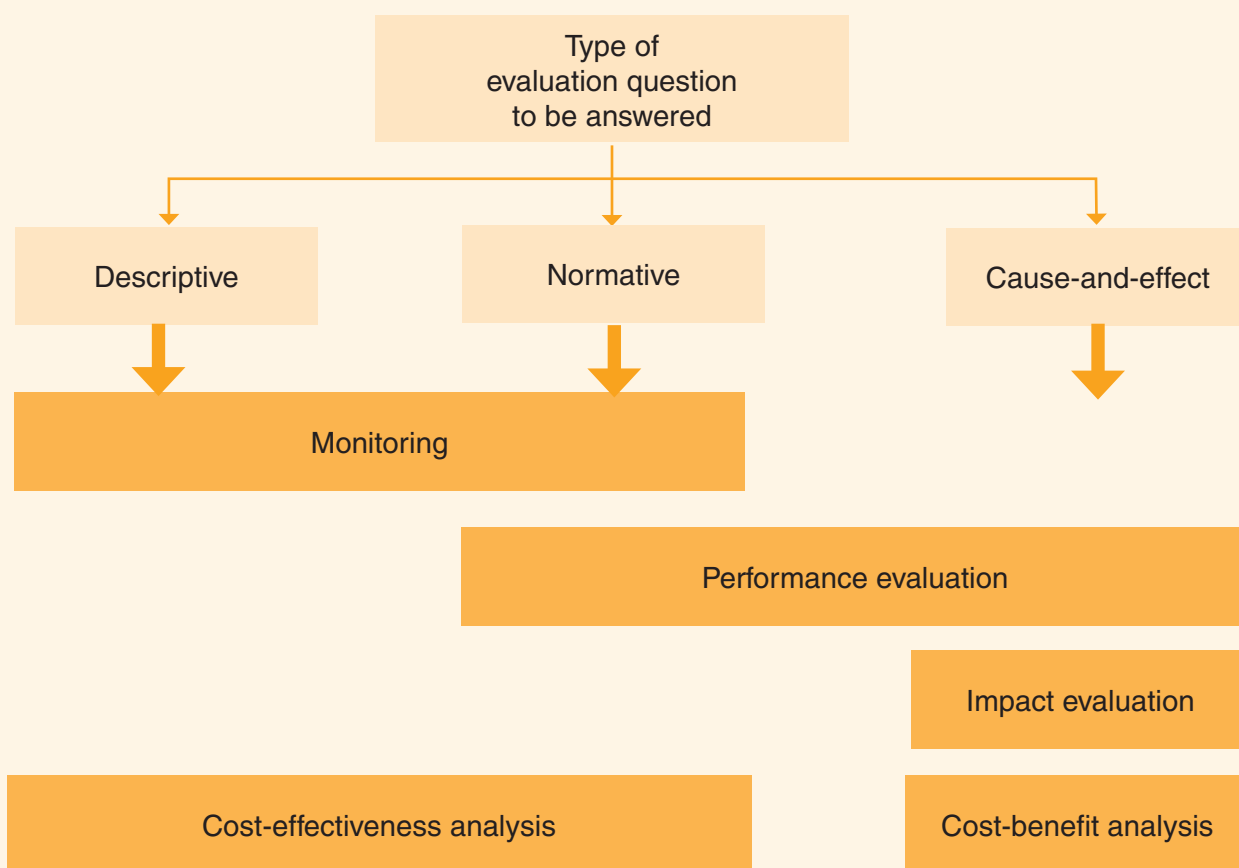
Linking evaluation questions to evaluation types

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

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

Figure 4.2 provides an overview of available evaluation options, depending on the type of

FIGURE 4.2: DIFFERENT EVALUATION TYPES



MONITORING (NO EVALUATION)

If a programme manager requires only descriptive information about the intervention, for example, because the project is in a very early stage and the objective is to obtain some general information about how the programme is being implemented, then a full-fledged evaluation may not be necessary. In that case, the knowledge obtained from monitoring may well be sufficient. Obviously, this requires the existence of a well-functioning

results measurement system, with a clearly defined results chain, indicators, data collection tools and so on (see Notes 2 and 3). A solid monitoring system is the backbone of all processes of results measurement and internal learning and provides a firm foundation for conducting thorough and informative evaluations. If such a system is in place, descriptive information about the programme should be relatively easy to obtain.

PERFORMANCE EVALUATION

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

This implies that they also assess whether the results framework is appropriate; that is, whether there are inconsistencies between

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

DEFINITION

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

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

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

type of evaluation to be conducted later on, in order to ensure a coherent process and access to good quality data.

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

Box 4.3: Categorizing evaluations by timing

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

Source: ILO, 2017.

Typically, performance evaluations cannot assess if a certain outcome happened explicitly due to the programme activities. Claiming the causal attribution of an observed change for an intervention requires robust impact evaluation methodologies to be applied (see the following section). However, the transition between the two evaluation types is fluid. In particular, qualitative tools can be used in performance evaluations in order to establish a causal link between the project activities and the measured outcome. This is usually done by identifying and revising alternative hypotheses on the causal process underlying the outcome in question. In contexts where a counterfactual-based impact evaluation is not feasible, reasonable or desired,

it is recommended that the range of qualitative methods for establishing causal claims in performance evaluations be explored and assumptions about contributions of the intervention to broader development impacts and goals be assessed (ILO, 2017). A useful overview of qualitative methods for addressing cause-and-effect questions that can also form part of a performance evaluation can be found in White and Phillips (2012).

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

FIGURE 4.3: KEY STEPS IN PLANNING AND MANAGING AN EVALUATION



Source: ILO, 2017.

IMPACT EVALUATION

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

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

Based on the information they provide, impact evaluations are particularly useful for informing strategic approaches, from scaling up effective interventions to curtailing unpromising programmes (Rubio, 2011). Decisions linked to replication and expansion often require evidence on whether smaller pilot projects worked for a comparable population. As governments, donors and non-governmental organizations are accountable for spending scarce resources efficiently, they increasingly demand the reliable evidence that well-conducted impact evaluation can deliver

(ILO, 2013). Importantly, the current global evidence base on what works in youth employment is still limited, so more impact evaluations are needed to support external learning processes. Impact evaluations also help us to understand which programme design options (dosage, delivery channel, etc.) are most important within a specific programme category, such as skills training, employment services or entrepreneurship promotion.

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

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

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

DEFINITION

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

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

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

youth employment programmes, for which the evidence base is still slim.

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

COST-EFFECTIVENESS AND COST-BENEFIT ANALYSES

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

Cost-effectiveness analysis (CEA) measures the cost per output or outcome (e.g. US\$300 per youth trained, US\$500 per job created) and compares this cost to similar

interventions of our own and other organizations. It thus answers the question of how much output or outcome we get per dollar spent (descriptive) and whether there is a gap between those findings and our expectations (normative).

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

DEFINITION

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

DEFINITION

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

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

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

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

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

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

Understanding evaluation needs based on the operational context

After formulating evaluation questions and identifying potential types of evaluation, we need to explore which evaluation can be carried out under which conditions. This implies analysing the operational context of our

intervention to gain a clearer understanding of how time and resource constraints, as well as political considerations, might affect any potential evaluation. This also involves taking factors outside our intervention into account.

TIMING

Early identification of evaluation demands:

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

Varying information needs over the programme life cycle:

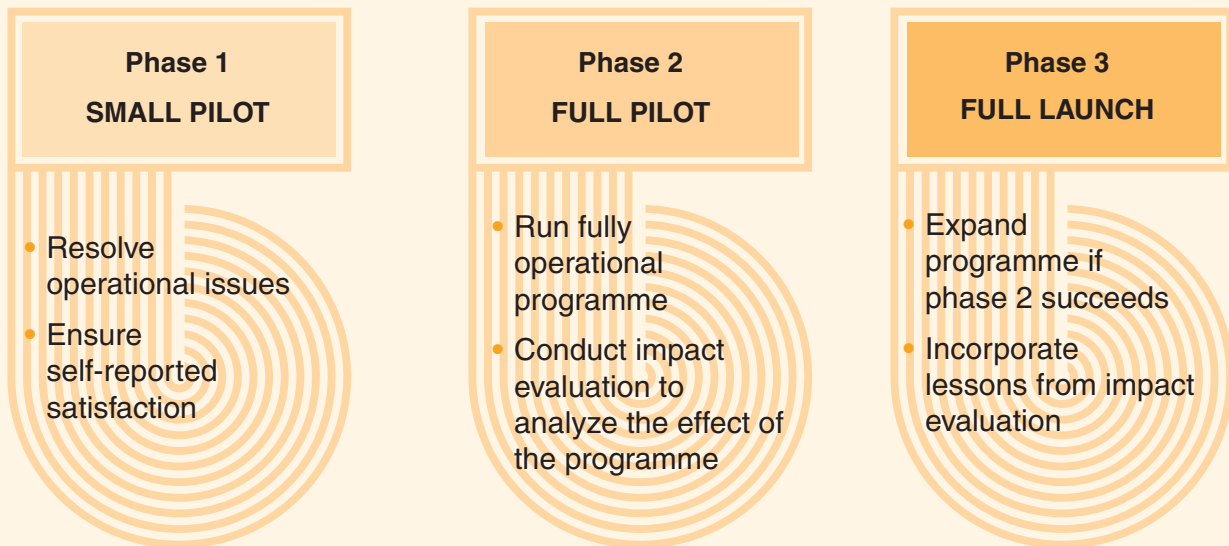
A programme that has just been planned may require a cost-effectiveness analysis to help determine whether or not it should be implemented. Alternatively, for a recently launched intervention, we may need to know how closely programme procedures are being followed and whether any adjustments are necessary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

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

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

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

FIGURE 4.4: OVERVIEW OF TYPICAL PROGRAMME PHASES



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

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

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

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

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

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

RESOURCES

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

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

The differences in scope and varying forms of data collection and analysis create a wide range of evaluation costs. Relying on desk research and key informant interviews is

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

THE POLITICAL CONTEXT

Different stakeholders within and outside our organization may have potentially competing interests in terms of whether or not an evaluation should take place, the issues to be studied, the type of evaluation and its

methodology, the data collection strategy, and who, if anyone, should be hired for the evaluation. All of these factors can exert pressure on the choice of an evaluation and influence the relevance and quality of the planned

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

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

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

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

Table 4.3: Overview of the main evaluation types

	Performance evaluation	Impact evaluation	Cost-effectiveness and cost-benefit analyses
What are the main questions answered by this type of evaluation?	<ul style="list-style-type: none"> • Do programmes have clear objectives? • Is the programme design appropriate to achieve the objectives? • Are adequate resources and systems (management, information, etc.) in place? • To what extent have programme objectives been achieved? • Do priorities need to be changed? • Is the programme being implemented according to design? 	<ul style="list-style-type: none"> • How has the well-being of participants changed as a result of the intervention? • Are there any unintended consequences, positive or negative, on programme participants? 	<ul style="list-style-type: none"> • Are programme costs justified, compared with similar interventions? • Are aggregate programme costs justified in terms of the benefits achieved?
When can this evaluation be conducted?	It may be conducted at early stages of implementation, for mid-term review or at programme completion	It should be designed during the planning of a programme but the final results will typically not be available until after the programme (phase) has been completed	It is commonly conducted during an ex ante analysis to determine whether the programme is worth implementing or continuing, or after the programme is completed to determine the final costs and their relation to the achieved benefits
How long does it take?	1–6 months (more if additional data collection is involved)	<ul style="list-style-type: none"> • At least 6 months (retrospective evaluation) • Approximately 12–24 months (prospective evaluation) 	1–3 months
What data collection and analyses are required?	Desk review of existing documents and selected field visits and interviews with programme staff and clients. Possibly complemented by monitoring data analysis, beneficiary and stakeholder interviews, mini-surveys, focus groups, etc.	Statistical and econometric analysis of survey and administrative data, ideally combined with qualitative data analysis	Desk review of existing programme documents and relevant literature as well as key informant interviews
Who carries out the evaluation?	Usually independent evaluator (but can also be internal)	Independent evaluation team, including, for example, lead evaluator, field coordinator, survey firm	Independent evaluator (can be the same as for performance or impact evaluation)
What skills are needed?	Programme analysis, possibly qualitative and simple quantitative methods	Statistical and econometric analysis, possibly qualitative methods	Valuation and econometric analysis of programme costs and benefits
What are the costs?	\$10,000–\$60,000	Can range from \$15,000 to \$1 million or more, depending on the size and complexity of the programme	\$10,000–\$30,000
What programmes are best suited for this evaluation?	Any programme	Programmes that are <ul style="list-style-type: none"> • innovative and untested • strategically relevant and influential • replicable 	<ul style="list-style-type: none"> • For cost-effectiveness analysis: any programme • for cost-benefit analysis: same as impact evaluation

Source: Adapted from [Rubio, 2011](#).

KEY POINTS

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

KEY RESOURCES



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



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



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

REFERENCES

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



CASE STUDY

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

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

Learning objectives

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

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

Introduction and case study context

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

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

Through an integrated multidimensional approach, the eight-year project (2011–2019), with an overall budget of US\$12.5 million, funded by Global Affairs Canada, contributes to

the development and implementation of youth employment initiatives in collaboration with national and local partners. The project engages extensively with policy-makers to strengthen youth employment policies and programmes. The project's activities are implemented at the national level and in the four governorates of El-Minya, Luxor, Port Said and Red Sea.

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

Part I: Formulating evaluation questions

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

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

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

Key outputs and activities: (a) Conduct training programmes and workshops for at least 200 government officials on how to design youth employment projects (including diagnostic analysis, project design and monitoring); (b) conduct quarterly knowledge-sharing

TIP



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

workshops through a “Youth Employment Forum” to be attended by ministries, civil society organizations and development partners.

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

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



Discussion Topics

1. Formulate at least one evaluation question for each of the following evaluation criteria:

relevance, effectiveness, efficiency, impact, sustainability (see also table 4.1 in Note 4).

Part II: Defining the evaluation's methodology

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

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

- ▶ key informant interviews
- ▶ focus group discussion.



Discussion Topics

Select one of the two data collection methods and:

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

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Note

5 Impact evaluation methods for youth employment interventions

A large, intricate, light gray decorative pattern covers the entire page. It features a central vertical axis with symmetrical, ornate scrollwork and floral motifs, resembling a stylized crown or a traditional Islamic geometric design. The pattern is composed of multiple overlapping layers of lines, creating a complex, lace-like appearance.

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

NOTE 5
**Impact evaluation methods for
youth employment interventions**

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. 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 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 5: Identifying an appropriate impact evaluation method/ International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131664-0 (print)
978-92-2-131665-7 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلم في برامج سوق العمل النشطة - الملاحظة 5: طرق تقييم أثر التدخلات التي تستهدف عمالة الشباب، ISBN 978-92-2-630795-7 (print), 978-92-2-630796-4 (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.

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

Printed in Switzerland

Contents

The attribution challenge	2	Key points	41
		Key resources	41
Challenges specific to evaluating youth-focused active labour market programmes	7	References	42
Quantitative methods of impact evaluation	10	Case study: Assessing rural micro-enterprise growth through different evaluation methods	43
Randomization – lottery design	12		
How it works	12	Tables	
When can I use a lottery design?	15	5.1 Sample size required to detect significant impacts	9
Advantages	15	5.2 Overview of different impact evaluation methods	11
Disadvantages	15	5.3 Categories of impact evaluation questions	38
Adapting random designs to different contexts	18	Figures	
Randomized phase-in design	19	5.1 A visual illustration of intervention impact	3
Randomized promotion design	20	5.2 Consider all possible evaluation methods during a planning stage	10
Difference-in-differences (DID)	22	5.3 Steps in a lottery design	12
How it works	22	5.4 Choosing samples for small and large programmes	14
When can I use a DID design?	24	5.5 Evaluation design	16
Advantages	24	5.6 Take-up of the EI Mashrou3 show	21
Disadvantages	24	5.7 Example of difference-in-differences analysis	23
Propensity score matching (PSM)	26	5.8 Comparing participants and non-participants	27
How it works	26	5.9 Impact evaluation design (simplified)	29
When can I use PSM?	28	5.10 Impacts on labour market outcomes	30
Advantages	28	5.11 Sample discontinuity chart	31
Disadvantages	28	5.12 Discontinuity in the probability of districts participating in the programme	33
Regression discontinuity design (RDD)	31	5.13 Comparing before-and-after outcomes	35
How it works	31		
When can I use RDD?	32		
Advantages	32		
Disadvantages	32		
Simple comparisons: Before and after	34		
Improving the relevance of quantitative impact evaluations	36		
Measuring a variety of impacts	36		
Combining quantitative and qualitative approaches	37		





Impact evaluation methods for youth employment interventions



Prerequisites:

A basic understanding of quantitative research methods would be helpful. This note describes a number of commonly used impact evaluation methods and explains the advantages and disadvantages of each, including theoretical and practical considerations.



Learning objectives:

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

- ▶ appreciate the major considerations and challenges to be taken into account when seeking to establish impact by asking: “What would have happened to the same people/household/community if the intervention had not taken place?”
- ▶ construct a counterfactual to estimate the change in outcomes that can be attributed to an intervention, and identify the key characteristics that treatment and comparison groups must share to ensure internal validity
- ▶ weight the pros and cons of different evaluation techniques and how they aim to eliminate selection bias
- ▶ comprehensively understand different quantitative research methods, from fully randomized designs to quasi-experimental methods, such as difference-in-difference, propensity score matching and regression discontinuity design
- ▶ use qualitative methods, to not only find out “what” happened – determining the average treatment effect of the intervention – but “why”.



Keywords:

Attribution, before-and-after comparison, comparison group, counterfactual, difference-in-differences, external validity, internal validity, lottery design, process tracing, propensity score matching, randomized phased-in design, randomized promotion design, regression discontinuity design, treatment group.

This note¹ provides practitioners with an overview of the different tools available for an impact evaluation and offers guidance on which ones to select under specific circumstances, and how to implement these tools to assess the effects of youth employment interventions. While impact evaluations can be based on both quantitative and qualitative methods, this note focuses primarily on quantitative methods and introduces qualitative methods as a valuable complement in the context of mixed-methods approaches.

The attribution challenge

Before moving on, we need to clarify what we mean by *impact*. In previous notes within this guide, we have used the term as synonymous with higher-level goals or outcomes relating to changes in a young person's employment situation such as reducing unemployment or increasing the well-being of individuals and households. In the context of impact evaluations, however, we understand *impact* more narrowly as the change in outcomes (e.g. employment status, working time, earnings) that can be attributed to our intervention.

As discussed in Note 4, impact evaluations try to answer cause-and-effect questions; that is, whether an intervention (the cause) improves outcomes among beneficiaries (the effect). For example:

- ▶ Can observed changes in trainees' likelihood of securing employment be attributed to our vocational training intervention?

- ▶ Does our job counselling intervention lead to a higher level of satisfaction among employers and a higher job retention rate?
- ▶ Does our start-up mentoring intervention foster business creation and sustainability?

The labour market outcomes that we are interested in are determined by many complex factors, such as the overall social and economic development context, changes in political and/or personal circumstances, etc. Hence, establishing the degree to which changes in such outcomes can be attributed to a particular intervention is challenging. The purpose of impact evaluation is precisely to overcome this **attribution** challenge by measuring the extent to which a particular programme, *and only that programme*, contributed to the change in the outcomes of interest.

DEFINITION

Attribution: The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention.

¹ This note of this guide draws on materials originally developed by Duflo et al. (2006), Khandker et al. (2010), and Gertler et al. (2016), adapting some of the material and illustrations to the youth employment field and providing a more concise presentation of impact evaluation methods.

In other words, impact evaluations try to assess whether, to what extent and why observed changes in outcomes of interest can be attributed to an intervention or project.

The focus of this note is the so-called **counterfactual** framework on which quantitative impact evaluations are typically based.² This approach defines the impact of an intervention as the difference between the observed outcomes under the intervention and the so-called counterfactual scenario: “What would have happened to the same people/

household/community if the intervention had not taken place?”. Figure 5.1 visualizes the concepts of impact and counterfactual.

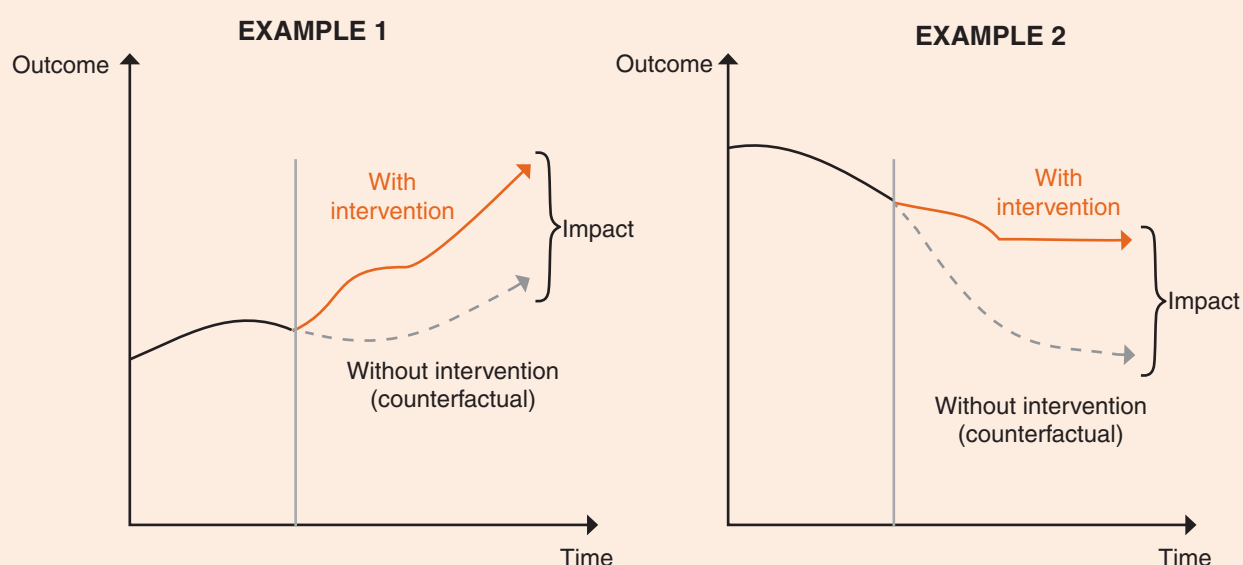
In practice, the real counterfactual is impossible to measure. Impact evaluation methods try to quantify causal effects through estimating or constructing the counterfactual typically – though not always – with comparison groups, sometimes known as *control groups*. The group of participants is known as the *treatment group* or *participant group*.

² Depending on the intervention being evaluated and its context, an impact evaluation design that mixes qualitative and quantitative methods is usually more appropriate, as explained in greater detail below.

DEFINITION

Counterfactual: The counterfactual describes what a certain outcome would have been for a programme participant in the absence of the programme. By definition, the counterfactual cannot be observed directly. Therefore, it must be estimated, for example using comparison groups.

FIGURE 5.1: A VISUAL ILLUSTRATION OF INTERVENTION IMPACT



Treatment and **comparison groups** should share the same characteristics in at least three ways (Gertler et al., 2016):

1. They should be similar in terms of both observable and unobservable characteristics: **Observable characteristics** may include age, gender, level of education, socio-economic status, family characteristics, employment status, and the like. **Unobservable characteristics** could include motivation, interest, values and ideologies and the level of family support, among other factors. Not every person in the treatment group must be identical to every person in the comparison group, but both groups should share similar average characteristics.
2. Treatment and comparison groups should be expected to react to the intervention in a similar way: For example, outcomes, such as skills or income, should be as likely to increase for members of the treatment group as for those in the comparison group.
3. Treatment and comparison groups should have similar levels of exposure to other interventions: For example, both groups should have the same access to other support services provided by local government, NGOs, etc.
4. When the treatment and comparison groups share the similarities listed above, we can confidently infer that any differences we see in outcomes between the two groups

can be attributed to the intervention. If, on the other hand, the comparison group differs from the treatment group in significant ways, comparisons of outcomes between the treated and comparison groups will reflect not only the impact of the intervention, but also the consequences of these differences. This is called selection bias.

Selection bias usually occurs when intervention participants and non-participants differ in characteristics that are not observed, which affect both the individuals' probability of taking part in (and/or finishing) the intervention and the outcomes of interest.

In most youth employment programmes, it is likely that those who apply to participate differ in significant ways from those who do not apply, and that these differences cannot be easily observed. For example, participants of a job counselling project might be more motivated and have access to better information about how to find a job than non-participants, even before the intervention starts. In that event, participants might be more successful in terms of labour market outcomes but it would be unclear whether this was due to the intervention or because of their initial advantage in starting conditions.

One of the key objectives of the evaluation techniques presented here is to eliminate selection bias. In the absence of selection bias,

DEFINITION

A **comparison group** is a group used to estimate the counterfactual in an impact evaluation. In contrast to members of the treatment group, members of the comparison group have not been exposed to the intervention we want to evaluate. The terms “comparison group” and “control group” are often used interchangeably. For the purposes of this document, we will use the generic term comparison group throughout.

DEFINITION

Treatment group: The group of people that actively take part in an intervention is known as the treatment group or participant group.

observed differences in outcomes between treatment and comparison groups can be attributed to the intervention.

Just as counterfactuals are not observable, nor is selection bias. Unless care is taken in selecting treatment and comparison groups appropriately, a simple comparison of labour market outcomes of treated and comparison groups will include both the impact of the intervention and selection bias. It is, in general, impossible to know the exact extent to which this is due to one or the other.

The evaluation techniques presented below attempt to select treatment and comparison groups to eliminate selection bias, so that comparisons between treated and comparison groups reflect only the impact of the intervention.

A good comparison group is essential for the **internal validity** of the evaluation which

determines the reliability and credibility of the evaluation results. External validity comes into play when we start to think about the transferability of these results: Should the intervention be scaled-up to other communities, or implemented at a regional or nationwide level? Can we expect similar results if we design this programme in other contexts and/or for a different target population? These are usually questions of significant interest to policy-makers.

It is important to bear in mind that the conditions will never be exactly the same when replicating or upscaling an intervention. Hence, to achieve **external validity**, it is crucial to understand the complex aspects surrounding the programme in the specific time, place and context of its implementation, and their potential influence on the evaluation results. For example, employment services for young graduates might be effective in a region because there is corresponding demand from the

DEFINITION

Observable and unobservable characteristics: Observable characteristics can be measured through appropriate data collection methods (such as surveys). They often include age, gender, level of education, socio-economic status, family characteristics, employment status, etc. Unobservable characteristics are those factors that cannot be, or are not, measured in an (impact) evaluation and could include motivation, interest, values and ideologies and the level of family support. For many of those unobservable characteristics (imperfect) proxy measures have been developed.

DEFINITION

Selection bias: Selection bias occurs when the reasons for an individual's participation in a programme are correlated with outcomes. This bias often occurs when the comparison group self-selects out of the programme (for example, drop-outs).

DEFINITION

Internal validity: To have internal validity, an impact evaluation must have a comparison group that provides a valid estimate of the counterfactual. An **internally valid** impact evaluation will be able to clearly attribute changes in outcomes to the intervention by controlling all possible differences between the treatment and comparison group. This can be achieved through appropriately applying experimental or quasi-experimental techniques.

local economy for this target group, the educational institutions have a good reputation, and the programme was carried out in a season when the employer demand for work was high. Implementing the same intervention in other regions and throughout the whole year, the results might show a very different picture.

In order to understand not only *if* something works, but *why and in what context* it can be expected to work, it is necessary to analyse the causal mechanisms underlying the

observed results. Qualitative methods, e.g. those applied in the context of theory-based evaluation, are of fundamental importance in this work.

In the course of this note, different quantitative impact evaluation methods will be introduced, followed by an example of a qualitative method and remarks on how a mixed-methods approach can help to achieve both internal and external validity of evaluation results.

DEFINITION

External validity: In impact evaluation, external validity means that the causal impact observed can be generalized to all eligible individuals. Therefore, for an evaluation to be externally valid, it is necessary that the evaluation sample is a representative sample of all eligible individuals.

Box 5.1: ILO's support for impact evaluation

ILO's Evaluation Department (also known as "EVAL") has developed a variety of resources to support impact evaluation (IE):

- ▶ *An Impact Evaluation Framework:* EVAL developed a position paper about how, when and why IEs should be considered and implemented, based on input from ILO staff. The position paper covers key issues, such as the specific use and purpose of IE; the match between evaluation research questions and appropriate methodology; use of a range of complementary and available methodologies; the feasibility and value of IEs; and the need to not only identify impact (what) but also the how and why.
- ▶ *An Impact Evaluation Review Facility (IERF):* EVAL established a review mechanism, which allows ILO staff to ask questions and request reviews of concept papers, full proposals, plans and reports to assist with planning, designing or implementing IEs (EVAL_impact@ilo.org). A Briefing Note on the operation of this facility is available.
- ▶ *An inventory of impact evaluations conducted at the ILO:* The inventory allows easier access to institutional knowledge in a variety of intervention areas.
- ▶ *A quality appraisal of ILO impact evaluations:* In order to monitor and report on the progress that the ILO is making in its use of IE and the quality of IE, EVAL will periodically commission a quality appraisal of IEs across the organization.
- ▶ *An Informal Impact Evaluation Network as a community of practice:* This informal group of colleagues who are involved with and interested in IEs meets on a regular basis to share experiences and provide peer review of IE, as required.

These resources are intended to support the ILO in further enhancing its capacity in terms of the use of IE, in documenting knowledge of what works and for whom, and in assessing impact.

Challenges specific to evaluating youth-focused active labour market programmes³

The nature of active labour market programmes (ALMPs), specifically those which focus on targeting a youth population, affects many aspects of the design of a valid evaluation. As background for a more detailed discussion of design issues in Note 6, this section presents an overview of some of the most common features of youth focused ALMPs and describes certain evaluation design features which are particularly relevant for these types of programmes. Understanding which of these features are likely to be present in a given setting will help in formulating the appropriate evaluation design.

Mandatory or voluntary programmes

A fundamental characteristic of a youth employment intervention is whether the programme is mandatory or voluntary. Mandatory programmes are built into many public employment services, including those related to unemployment insurances and training programmes. In these settings, young people are required to participate in an ALMP which is linked to an unemployment benefit. That being said, and as will be explained in the next sections, mandatory participation in a youth employment intervention does create challenges for impact evaluation, where valid estimates of impact typically need a treatment and an equivalent comparison group. Some impact evaluation methods can only be applied to voluntary programmes that recruit participants from a wider pool of applicants who can decide whether or not to participate.

Non-compliance: No-shows and dropouts

In many voluntary youth employment interventions a substantial fraction of people who are assigned to the programme will either fail to register for the programme (so-called no-shows) or will drop out prior to completion of the programme (dropouts). This challenge is particularly relevant to young people who are highly mobile, tend to change address and place of work frequently and alternate between working and studying.

Indeed, [Card et al. \(2011\)](#) state that:

it is rare to achieve programme completion rates over 80 percent and rates as low as 50 percent are common and failure to anticipate the problems caused by no-shows and dropouts is one of the leading causes of a broken design in ALMP evaluations (2011, p. 13).

While non-compliance by members of either the programme group or the comparison group does not invalidate an evaluation design per se, it does complicate the interpretation of the results, and means that the evaluation has to collect data on the actual programme participation rates of the treatment group and the comparison group.

The validity of a randomized design relies critically on the equivalence between the observed outcomes of the comparison group and the counterfactual outcomes of the treatment

³ This section is based on [Card et al. \(2011\)](#).

group. In most cases this equivalence is compromised when members from one group or the other are dropped or lost. For this reason, the analysis of a randomized design should be based on a comparison of the treatment and comparison groups as initially assigned, using data on everyone who was initially assigned to these groups. In the experimental evaluation literature this is known as an “intention to treat analysis”.

Recruitment and screening

Because only some of all the young people recruited into an impact evaluation are assigned to actually receive the programme, intake for an evaluation may disrupt the normal flow of clients into an ongoing programme.⁴ This is not a particular concern in a setting where there are many more applicants than available slots: in these cases random selection serves as a convenient and objective rationing device. In settings where the regular flow of recruits is needed to fill the available programme slots, however, programme operators may object to having some of their potential clients allocated to the comparison group and may try to override the assignment process. It is extremely important to know in advance whether this is likely to occur. If so, planning for the evaluation may have to include a budget for extra recruitment efforts to increase the flow of new clients, and extra resources to closely monitor compliance with recruiting protocols. For example, ALMPs for young people may be limited to unemployed men and women between the ages of 16 and 30. Normally, the same eligibility screening procedures and rules should be used to select participants for the evaluation.

Sample sizes

Guidelines for the necessary sample sizes for an ALMP evaluation are based on a standard power calculation. The main ingredient for this calculation is an estimate of the plausible effect size of the programme (e.g., the effect of the programme on the outcome of interest, expressed as a fraction of the standard deviation of this outcome). Given this value, and standard choices for the statistical significance level (e.g., 5 per cent) and the adequacy of the power of the design (e.g., 0.80), it is straightforward to calculate the appropriate sample sizes for the treatment and comparison groups of a randomized design with equal-sized groups. [Card et al. \(2011\)](#) developed guidance (shown in table 5.1) showing detailing the sample size required to measure a range of impacts. Each row shows the employment rate of the comparison group, and each column represents the difference between treatment and control groups. For example, if the employment rate is 50% per cent in the comparison group, to detect a significant impact of 2.5 percentage points in employment, the required sample size is of 6,354 participants and the same number of non-participants.

⁴ For example, if 100 new clients present themselves at the programme sites each month, and there are 80 open programme slots each month, then, at most, 40 people per month can be recruited into the evaluation: 20 will be assigned to the programme (along with the other 60 new clients who are not part of the evaluation) and 20 to the comparison group.

Table 5.1: Sample size required to detect significant impacts

	Impact of programme						
	2.5%	5.0%	7.5%	10.0%	12.5%	15.0%	
Employment rate of control group	30%	5475	1417	650	376	247	176
	35%	5883	1511	688	396	259	183
	40%	6166	1574	713	408	265	186
	45%	6323	1605	723	412	266	186
	50%	6354	1605	719	408	262	183
	55%	6260	1574	702	396	254	176
	60%	6040	1511	671	376	240	165
	65%	5695	1417	625	349	221	151

Source: [Card et al. \(2011\)](#)

* Under the standard assumptions (power = 0.8, significance = 0.5, equal-sized groups), using the `sampsi` command in Stata

In thinking about the effect size of interest for an ALMP, [Card et al. \(2011\)](#) recommend to place that these programmes be put in context. They state that

A very large body of research has shown that in most countries around the world each additional year of formal schooling is associated with a gain in earnings of about 10 percent. Arguably, a typical ALMP involves a smaller investment than a typical year of formal schooling, so an effect size of less than 10 per cent is reasonable, and for less intensive programs, effect sizes of no more than 5 per cent may be plausible (2011, p.19).

Timing of follow-up surveys

The timing for the follow-up survey (or surveys) is an important decision in terms of guaranteeing programme impacts. Many ALMP evaluations use a one-year follow-up survey, in part because the terms of the evaluation contract often require a final report within two or three years. On the other hand,

the existing ALMP literature suggests that the impact of more intensive programmes, such as classroom training and on-the-job training programmes, only tends to manifest itself two or three years after entry into the ALMP, rather than after just one year ([Card et al., 2011](#)). Based on these studies, and consideration of the interruption effects of many ALMPs, a post-programme horizon of at least two years is desirable for ALMPs of longer duration.

There is, however, a trade-off between being able to observe long(er)-term impacts and ensuring a valid impact evaluation design: as young people are highly mobile and might move across the country (or even migrate) after finishing their education or training programmes, it becomes increasingly difficult to track down a sufficient number of programme beneficiaries as time progresses. When a large number of the young people who participated in the baseline survey of an impact evaluation cannot be contacted for follow-up survey(s) (high attrition rate), it becomes increasingly difficult, or impossible, to reliably detect and quantify impacts due to reduced statistical power and possible bias.

Quantitative methods of impact evaluation

Recommended quantitative methods of impact evaluation achieve internal validity and avoid selection bias by comparing groups with and without treatment, which ideally differ only in this respect. This can best be achieved if we have control over who receives the intervention and who does not. In this case, **experimental evaluation** designs are possible, the most common of which is a **randomized controlled trial**. If the assignment to treatment and comparison group is totally random, the two groups will be, on average, very similar before the programme starts and we will have gone a long way towards assuring internal validity.

For a variety of reasons, to be discussed

below, randomization is not always possible or desirable. In that case, other methods can be used that seek to undertake internally valid comparisons by constructing a valid counterfactual. These are called **quasi-experimental impact evaluation methods**. The most commonly used ones are difference-in-differences (DID), propensity score matching (PSM) and regression discontinuity design (RDD), all of which will be briefly introduced in this section. It is generally an excellent idea to consider all possible impact evaluation efforts and to carefully weigh advantages and disadvantages before proceeding with the evaluation. Table 5.2 provides an overview over different quantitative impact evaluation methods.

FIGURE 5.2: CONSIDER ALL POSSIBLE EVALUATION METHODS DURING A PLANNING STAGE



Source: www.freshspectrum.com

DEFINITION

Experimental design: Experimental designs rely on some element of randomization in the allocation of participants into treatment and comparison groups. They can produce highly credible impact estimates but are often costly and, for certain interventions, difficult to implement.

A **randomized controlled trial** is a study in which people are allocated at random (by chance alone) to receive a treatment, such as participating in a specific intervention.

Quasi-experimental design: Quasi-experimental design approaches are used to construct a valid comparison group by using statistical means to control for differences between the individuals treated with the programme being evaluated and those not treated.

Table 5.2: Overview of different impact evaluation methods

Methodology	Description	Who is in the comparison group?	Required assumptions	Required data
Before-and-after	Measures how programme participants improved (or changed) over time	Programme participants themselves –before participating in the programme	The programme was the only factor influencing any changes in the measured outcome over time	Before-and-after data for programme participants
Comparing participants to non-participants	Measures difference between programme participants and non-participants after the programme is completed	Individuals who didn't participate in the programme (for any reason), but for whom data were collected after the programme ended	Non-participants are identical to participants except for programme participation	After-programme data for participants and non-participants
Differences-in-differences	Measures improvement (change) over time of programme participants relative to the improvement (change) of non-participants	Individuals who didn't participate in the programme (for any reason), but for whom data were collected both before and after the programme	If the programme didn't exist, the two groups would have had identical trajectories over this period (would share the same "common" time trend)	Before-and-after data for both participants and non-participants
Propensity score matching	Individuals in the treatment group are matched with non-participants who have similar observable characteristics. The average difference in outcomes between matched individuals is the estimated impact	Non-participants who have a combination of characteristics which predict that they would be as likely to participate as participants	The factors that were excluded (because they are unobservable and/or have been not been measured) do not bias results because they are either uncorrelated with the outcome or do not differ between participants and non-participants	Outcomes as well as "variables for matching" for both participants and non-participants
Regression discontinuity design	Individuals are ranked based on specific, measurable criteria. There is a cut-off point to determine who is eligible to participate. Impact is measured by comparing outcomes of participants and non-participants close to the cut-off line	Individuals who are close to the cut-off line, but who fall on the side of that line where they (just) do not get the programme	After controlling for the criteria (and other measures of choice), the remaining differences between individuals directly below and directly above the cut-off score are not statistically significant and will not bias the results. A necessary requirement for this to hold is that the cut-off criteria are strictly adhered to	Outcomes as well as data of ranking criteria (e.g. age, index, etc.). Socio-economic background variables highly desirable.
Randomized evaluation	A sample of eligible individuals is randomly assigned into two groups: those who receive the intervention and those who do not. Impact is the difference in outcomes between the two groups. There are different ways of carrying out the randomization	Participants are randomly assigned to the treatment and comparison groups	Randomization is successful and complied with; that is, the two groups are statistically identical (in terms of both observed and unobserved factors)	Outcome data for comparison and treatment groups. Baseline data and background variables are desirable

Randomization – lottery design

A lottery is a simple and transparent way to assign youth to groups which will receive our services (the treatment group) and those which will not (the comparison group). This is the method used to design randomized controlled trials. If a large enough sample of people from the same population of interest is randomly assigned to one of two groups, then both groups will, on average, have similar observable characteristics (age, gender, height,

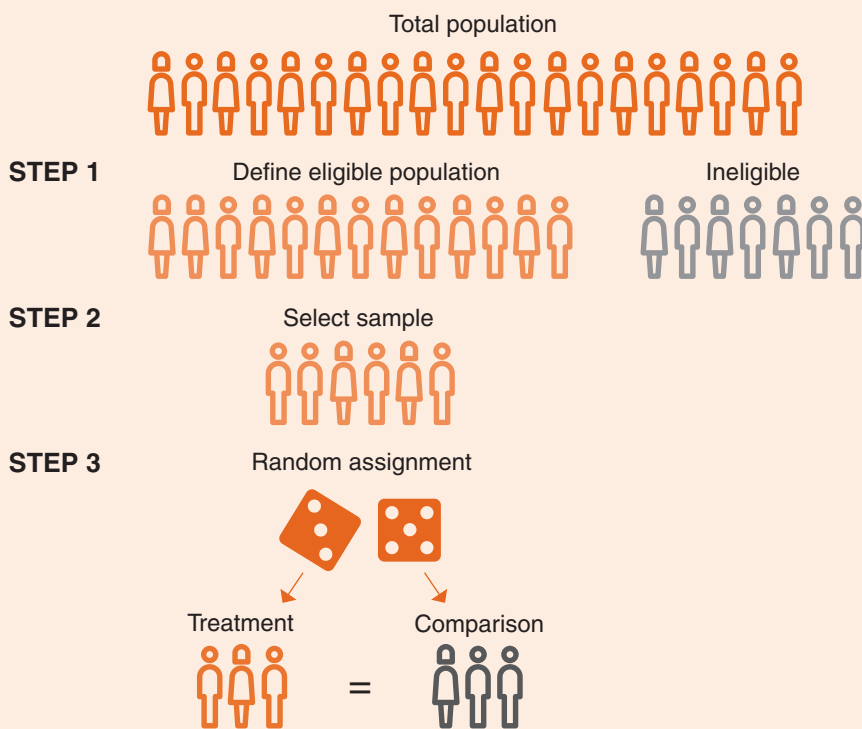
level of education, etc.). Equally importantly, they will also, on average, share the same unobservable characteristics (such as motivation and state of mind).

Through randomization, the difference in outcomes that we observe between the two groups at the end of our intervention can be attributed to the intervention, because all other factors that could influence the outcomes are, in general, equal.

HOW IT WORKS

There are three steps to a lottery design (see figure 5.3).

FIGURE 5.3: STEPS IN A LOTTERY DESIGN



Step 1: Define the eligible population

The first step in a randomized controlled trial is to find a group of eligible young people for an intervention. If a medical scientist is studying the effect of a drug on a childhood disease, she searches for a specific group of children and will not enrol adults or elderly people in the intervention. Likewise, a job training programme may target urban street youth of a specific age range, and so will not include adults or rural youth. What is important here is to have very clear and transparent criteria (age, gender, income level, employment status, etc.) and to be able to communicate who will be eligible to join the intervention and who will not.

Step 2: Select a sample for the evaluation

To evaluate an intervention, we do not need to test everyone who will participate in the intervention. We just need to choose a representative group of people that is numerous enough for the purposes of our evaluation; this is called our **sample**. These will be the young people on whom we will collect data. While Note 6 provides more details about how to determine the sample and its size, the typical sample size for a youth employment intervention evaluated through a lottery design is somewhere between 500 and 2,000 study participants (usually with a roughly equal split between the treatment and comparison groups).

Choosing the sample for the evaluation can be done in two ways, depending on whether the intervention is large or small. A small intervention may find that there are 10,000 eligible beneficiaries, such as urban street youth aged 16–24 years old. The intervention may have sufficient budget to help 500 of them. Ideally, a comparison group will be similar in size to the treatment group, so 1,000 out of the 10,000 street youth will need to be selected for the intervention and evaluation (see figure 5.4, left-hand image).

Large programmes may be bigger than the sample size needed for an evaluation. If the job training is able to serve 4,000 young people, it is not necessary to find an additional 4,000 young people for comparison. Instead, only 1,000 may be needed. The intervention can then identify a sample of 5,000 youth from the total population of 10,000. Of these, 3,000 youth can be guaranteed admission to the intervention. The remaining 2,000 will then be randomly split between the intervention and the comparison group (figure 5.4, right-hand image).

In order to make the selection representative of the total eligible population of 10,000 street youth, the sample (whether 1,000 in the first case or 5,000 in the second case) should be selected at random from the eligible population. By selecting randomly, participants will, on average, have similar characteristics to the total eligible population. Even though we include only a limited number of youth in the study, the potential impact of the intervention can be extrapolated to cover the entire eligible population, in this case, 10,000 young people.

DEFINITION

A **sample** is a subset of a population. Since it is usually impossible or impractical to collect information on the entire population of interest, we can instead collect information on a subset of manageable size. If the subset is well chosen, then it is possible to extrapolate results to the entire population.

Step 3: Randomize assignment

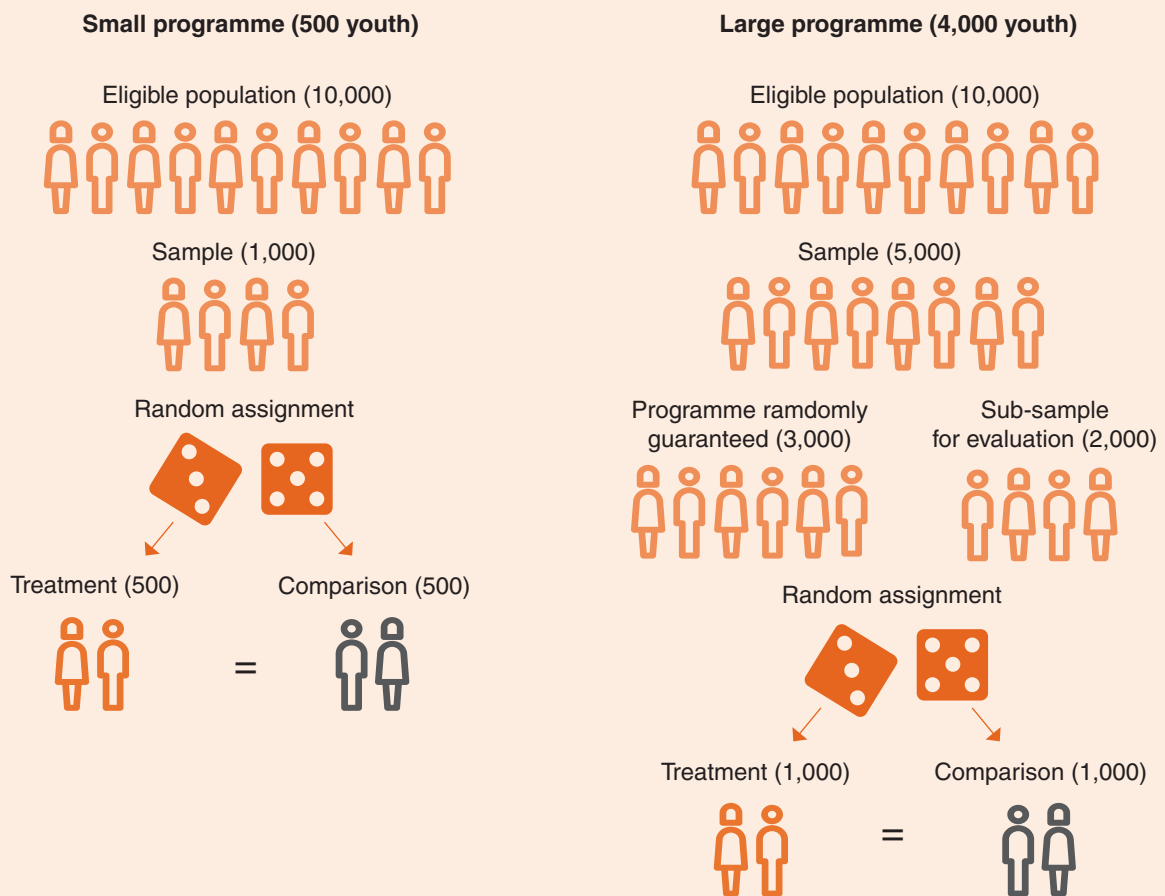
The next step is to assign the selected sample of youth to treatment and comparison groups which are roughly equal in size. In randomized controlled trials, every youth has the same chance of receiving the intervention. Randomization can be done via traditional techniques, such as flipping a coin, rolling dice or drawing names out of a hat. Randomization can be done publicly, if desired, if the sample is relatively small (drawing 2,000 names out of a hat, for example, would not be very practical). Alternatively, and more appropriately if the number of people is large, we can randomize by using computer software, such as Microsoft Excel. Randomization

TIP



One way of obtaining a random sample of youth is to get a list of the total population of street youth from a census, voter registration records or some other database, and randomly select from that list. If that approach is not possible, randomly targeting areas where street youth interact, such as an urban centre, will produce a random sample. If young people are known to spend time at 50 different centres around a city or country, randomly selecting centres and then selecting a portion of youth at these centres to participate in the study is likely to result in a selection of youth with minimal bias. Note 6 will discuss sampling more in detail.

FIGURE 5.4: CHOOSING SAMPLES FOR SMALL AND LARGE PROGRAMMES



can occur at several levels. By assigning our sample to treatment or comparison groups randomly, we select participants fairly, and we also develop a good counterfactual: if the

sample size is big enough, youth in the treatment group have, on average, the same observable and unobservable characteristics as those in the comparison group.

WHEN CAN I USE A LOTTERY DESIGN?

A randomized lottery evaluation can be used when the evaluation is planned in advance of implementation (prospective) and when the intervention can serve only a fraction of the eligible youth. As long as resource constraints prevent the intervention from serving the entire eligible population, there are no ethical concerns in having a comparison group, because a subset of the population will necessarily be left out of the intervention. In such a situation, comparison groups can be maintained to measure short-, medium- and

long-term impacts of the intervention (Gertler et al., 2016). Importantly, the central advantage of randomizations – that treatment and comparison groups, on average, share the same characteristics – will only be maintained if we manage to follow up with (almost) all members of the treatment and comparison group. High attrition rates pose a severe threat to the internal validity of our results for every impact evaluation method, and methods that use randomization techniques are no exception.

ADVANTAGES

- ▶ A lottery design is the most robust method for developing a counterfactual because it leads to a very well-matched comparison group (relying on fewer assumptions than other methods). It is therefore considered the most credible design to measure impact.
- ▶ It is by far the simplest of all evaluation methods in analytical terms. The impact of

the intervention in a random trial is simply the mean difference in outcomes between treatment and comparison groups.

- ▶ It allows for communities to be directly involved in the selection process for a fair and transparent allocation of benefits.
- ▶ It is easy to implement and communicate to programme staff.

DISADVANTAGES

- ▶ Conducting a randomized experiment can be very cost- and time-intensive.
- ▶ No ex-post implementation of this method is possible. Planning the evaluation has to be part of planning the intervention (which is good practice in any case but does not always represent the reality in project work).
- ▶ It requires a comparison group to be excluded from the intervention for the duration of the impact evaluation. Political and/or ethical concerns might emerge in spite of the

transparent allocation criterion of randomization (see more in the section “Adapting random designs to different contexts” below).

- ▶ Organizations must ensure that partners and local stakeholders consent to the method.
- ▶ The internal validity of a lottery design depends on the fact that the randomization works and is maintained throughout the study, which may not be easy to achieve. This condition may be threatened if randomization is implemented incorrectly, if treatment

or comparison groups do not comply with their status (that is, if treatment individuals do not take up the intervention or comparison individuals receive the programme), if participants drop out of the study prior to completion or if there are **spillover effects**: for example, young people who received the job

training might transfer the acquired skills and knowledge to their peers, thereby blurring the clear separation between treatment and comparison groups. These cases are highly problematic, as they can substantially bias the results and thereby threaten the overall validity of the evaluation.

DEFINITION

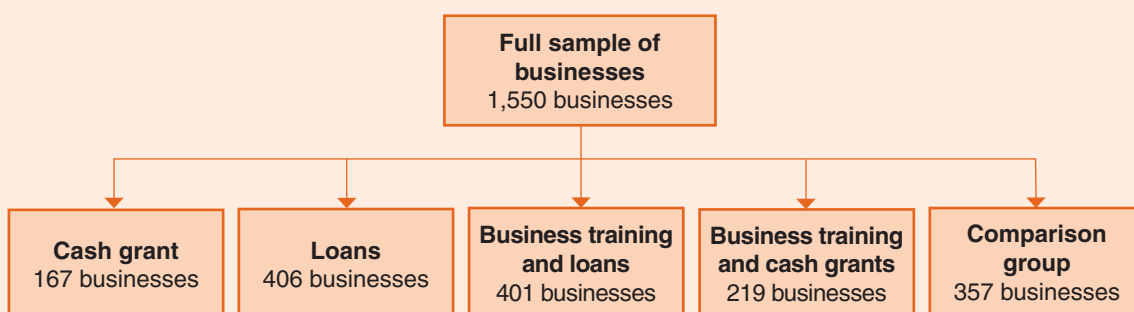
Spillover effects: Spillovers are effects of an intervention on non-participants; for instance, if knowledge from a skills training spreads within a village, even to those who did not attend the course.

Box 5.2: Evaluation of ILO’s intervention Start and Improve Your Business (SIYB)

The evaluation was designed to test whether expanding access to capital via grants or loans would increase the profits of micro-enterprises owned by men or women, and whether the ILO’s SIYB entrepreneurship training could further increase impacts.

The research team surveyed 4,637 micro-enterprises from a census of businesses and selected 1,550 business owners to be included in the evaluation sample – based, among other criteria, on an expression of interest in receiving ILO training and participating in the loan intervention. The sample included small business owners interested in improving their businesses (for example, hair salons, retail shops and tailors). The sample was randomly split into five treatment arms, which received the following interventions: (1) a loan; (2) a cash grant; (3) business training and a loan; (4) business training and a cash grant; and (5) no intervention (the comparison group) (figure 5.5).

FIGURE 5.5: EVALUATION DESIGN



The main interventions which were delivered to the business owners were:

- Business training: The Start Your Business (SYB) training programme targets starting (or nascent) entrepreneurs and consists of a five-day training course, followed by fieldwork and group-based and individual counselling sessions. The trainees prepare their detailed bankable business plan and action plan (see www.ilo.org/siyb).
- Unconditional cash grants, valued at US\$200, were delivered via free bank accounts at a local microfinance institution (MFI). The business owners were given free choice in the use of the loan.
- Semi-conditional loans, valued at US\$180 to US\$220, were offered at a discounted annual interest rate of 20 per cent by the MFI. Loans had to be paid back to the MFI, but there were no consequences in the event of misuse of the money.

The size of the grants and loans is equal to approximately 1.5 times the monthly profits of the average businesses.

To check whether the randomization “worked”, the evaluation compared business owners in the treatment group with those in the comparison group with respect to 26 different variables and found that, for virtually every characteristic, treated and non-treated enterprises, on average, looked alike before the intervention.

The business owners were surveyed before the intervention (baseline survey) and six months, nine months and two years after the intervention (three follow-up surveys).

The main outcome variable of interest was business profit and the evaluation found a significant increase in earnings after the intervention, but only for male business owners. None of the interventions led to sustained increases in profits for female entrepreneurs. Women with high initial profits also saw negative effects through all interventions. While the initial response to the grants was positive, this increase disappeared entirely and even became negative over time. Women who received the grant made 35 per cent less profit than their peers who received no intervention. After nine months, women were either not better off or were even worse off than their counterparts in the comparison group.

The evaluation also found that the proximity of family members represents a positive force on business for men and a negative one for women. Married women with family living in the same district experienced large and significant decreases in their profits.

Source: [Fiala, 2015](#).

Adapting random designs to different contexts

Some programmers are reluctant to randomly assign potential beneficiaries into treatment and comparison groups. The general concern is that the evaluation leads to withholding seemingly obvious benefits (such as training opportunities) from needy individuals, which

would be unethical. Still, for many interventions, demand considerably exceeds what can be supplied and, as further elaborated in box 5.3, randomization may in fact be more ethical than other selection methods.

Box 5.3: Is randomization ethical?

Sometimes randomly assigning potential beneficiaries into treatment and comparison groups is considered unethical. These concerns might be valid in certain cases, for example when a policy or intervention that is likely or proven to work can be extended at little cost to a large population. However, more often than not one of the following situations arise:

- **Uncertainty of project impact.** For most programmes, it is not clear if the intervention has a positive and sizable impact on the individual and the community that justifies the resources being spent. For instance, programmes geared toward girls at the exclusion of boys may increase gender violence. A microfinance intervention for youth may leave participants worse off if they are not able to repay their loans. A poorly designed training programme may actually decrease job prospects. An increase in incomes (e.g., a US\$100 per participant) may come at a very high cost (e.g., US\$1,000 per person). Thus, in the case of interventions whose impact and cost-benefit structure has not yet been sufficiently proven, it is well justified to evaluate the intervention based on randomly assigned treatment and comparison groups.
- **Budget constraints.** In reality, because of limited resources, it is rarely possible to serve everyone in need. That is, most programmes provide benefits and services only to a limited number of beneficiaries, thereby excluding others, whether this is made explicit or not. For example, if a youth training intervention has a limited number of available spots, then some young people will receive the training while others will not. Similarly, if an intervention is carried out in one particular district, eligible youth in other districts are excluded. Randomization allows programme managers to allocate scarce places in their interventions in a way that is fair and that gives the same chance for participation to everyone. If the randomization is done in an open manner (for example as a lottery during a public event), it also enhances transparency in the selection process and may reduce fears in the population that selection was based on personal or political preferences.

Nevertheless, creating a pure comparison group by random lottery assignment in which young people are never given the intervention is sometimes impossible. Both random phase-in and random promotion designs avoid the strict separation into treatment and comparison groups and might be a viable alternative for an experimental impact evaluation design when lottery designs are not feasible or desirable.

TIP

With a phase-in approach, it is critical to have enough time between each of the phases for the intervention to show effects. If, for example, an intervention officer believes that it will take two years for the impact of the intervention to take effect, the time between the first and last phase must be at least two years. Small or short-term programmes may not be suitable for this approach.

RANDOMIZED PHASE-IN DESIGN

Because many programmes are active in a community for years, never giving the intervention to a group of needy individuals can be both politically and programmatically difficult. A variation of the lottery design is the phase-in design. The main difference between a phase-in design and a lottery design is the method of assigning people to treatment and comparison groups. In practice, potential beneficiaries are randomly divided into two or more groups. The intervention is then rolled-out over time, so that individuals of group one participate in the intervention first, followed by group two, group three, and so on. During the time when groups are on the waiting list, they can serve as the comparison group until they receive the intervention.

For example, a non-governmental organization (NGO) may have sufficient budget to train 1,500 youths, but it may not have the capacity to conduct all of the training simultaneously. Instead, it chooses to train 500 people per year over three years. If it can identify all 1,500 participants in the beginning, a phased-in randomization may be the best evaluation method to adopt. The 1,500 youths are randomly split into three groups. In year one, while group 1

receives training, groups 2 and 3 remain on the waiting list and can serve as the comparison group. In year two, only group 3 remains for comparison purposes. By year three, all three groups will have received training.

As individuals are selected at random for the different groups, it is possible to compare those offered treatment first with those offered treatment later. This method often suits the natural roll-out of many programmes.

However, because everyone eventually benefits from the programme, the phase-in design approach is usually not ideal for finding the long-term impact of an intervention because eventually there is no comparison group. Even large, longstanding programmes will have difficulty in asking participants to wait for three or four years before their turn comes, so the time span of results is often limited to one or two years. Moreover, there is a risk that participants may change their behaviour while waiting to join the intervention. This could invalidate their ability to serve as a good comparison group. For example, they may stop looking for jobs in anticipation of joining a skills training intervention.

RANDOMIZED PROMOTION DESIGN

There may be cases where it is not possible or desirable to exclude any potential beneficiaries and where the intervention is not rolled out over time. In such cases, the randomized promotion method (also called encouragement design) may be suitable. When it is not possible to randomly assign young people into a group that receives benefits and a group that does not, it may be possible instead to randomly promote the intervention. That is, rather than randomizing those who *receive* the benefits and services, we randomize who is *encouraged to receive* those benefits.

Random encouragement may take many different forms. In the case of youth savings accounts, we may randomly advertise the initiative in selected schools. For a training programme, we could hire a social worker to randomly visit homes of unemployed young people, describe the programme and offer to enrol young people on the spot. In the case of a financial literacy campaign, we may want to randomly send text messages to one part of the target audience, but not to another. In all cases, there will still be people in the promoted group that will not take up our intervention, as there will be people in the non-promoted group who actually will. But the idea is that, if the encouragement is effective, then the enrolment rate among the promoted group should be higher than the rate among those who did not receive the promotion.

To assess the impact of the intervention, we cannot, unfortunately, simply compare the outcomes of those who participated in the

TIP



Random promotion evaluation may be suitable for:

- ▶ interventions that distribute training vouchers
- ▶ interventions encouraging youth to open saving accounts
- ▶ interventions leveraging mass-media based campaigns.

intervention with the outcomes of those who did not. People who choose to participate in an intervention are almost always different from those who do not, and many of these differences may not be observable or measurable. Even if promotion is random, participation in the intervention will not be random, so comparing participants to non-participants would be like comparing apples to oranges.

We can, however, compare outcomes between everyone who received the encouragement and all young people in the comparison group. Given that the promotion is assigned randomly, the promoted and non-promoted groups have, on average, equal characteristics. Thus, the difference that we observe in average outcomes between the two groups can be attributed to the fact that those people only enrolled in the intervention because they received the promotion.

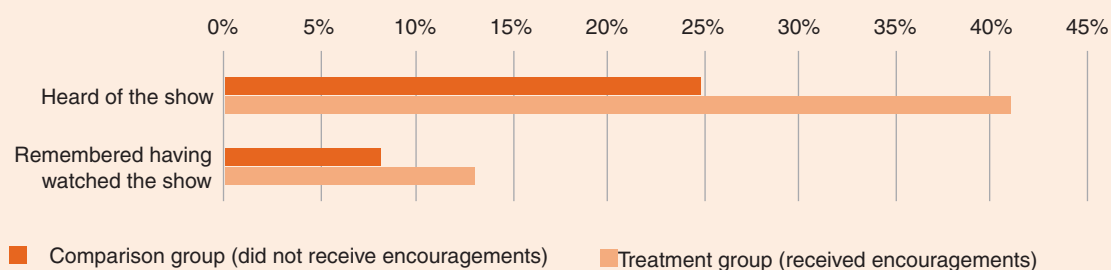
A key advantage of this design is that randomized promotion campaigns never deny anyone the programme, but instead allow people to make their own decisions about whether or not to take up the intervention. However, these studies often need larger sample sizes to provide reliable impact estimates, which increases costs.

Box 5.4: Evaluating an edutainment intervention in Egypt

El Mashroua is a reality TV show designed to promote entrepreneurship among young adult viewers and broadcasted on one of the most popular Egyptian television channels. To evaluate the impact of the show, a research team, supported by the ILO, chose a randomized promotion design. From the study sample of 9,277 individuals, a randomly selected treatment group received SMS reminders about the show that were designed to encourage recipients to tune in.

The follow-up survey that was conducted approximately 1.5 years after the broadcast clearly showed that young people from the treatment group (those who received messages) were more likely to have heard of the show and to have watched at least one episode compared to youth from the comparison group (those who did not receive reminders), see also figure 5.6.

FIGURE 5.6: TAKE-UP OF THE EL MASHROU3 SHOW



These statistically significant differences can be exploited to estimate the actual impacts of the show on the viewers' attitudes and labour market outcomes. This is possible because two assumptions can be made:

1. Because of the randomization, treatment and comparison group do not differ systematically in any observable or unobservable characteristics that could be correlated with the outcome variables.
2. Because having received the messages alone does neither affect attitudes nor labour market outcomes, any difference between treatment and comparison groups can be attributed to the difference in likelihood of having watched the show.

The study finds that having watched the show did not have impacts on young people's propensity to start a business, but that it significantly reduced gender-discriminatory attitudes held by men against women.

Source: Barsoum et al., 2017.

Difference-in-differences (DID)

For the reasons already explained, it is sometimes not possible or desirable to employ experimental evaluation methods. In that case, there is a range of quasi-experimental impact evaluation methods, which can also deliver robust, internally valid results. One of the most

commonly used techniques is the difference-in-differences (DID) approach, which compares the change in outcomes experienced by the treatment group with the change in outcomes experienced by the comparison group.

HOW IT WORKS

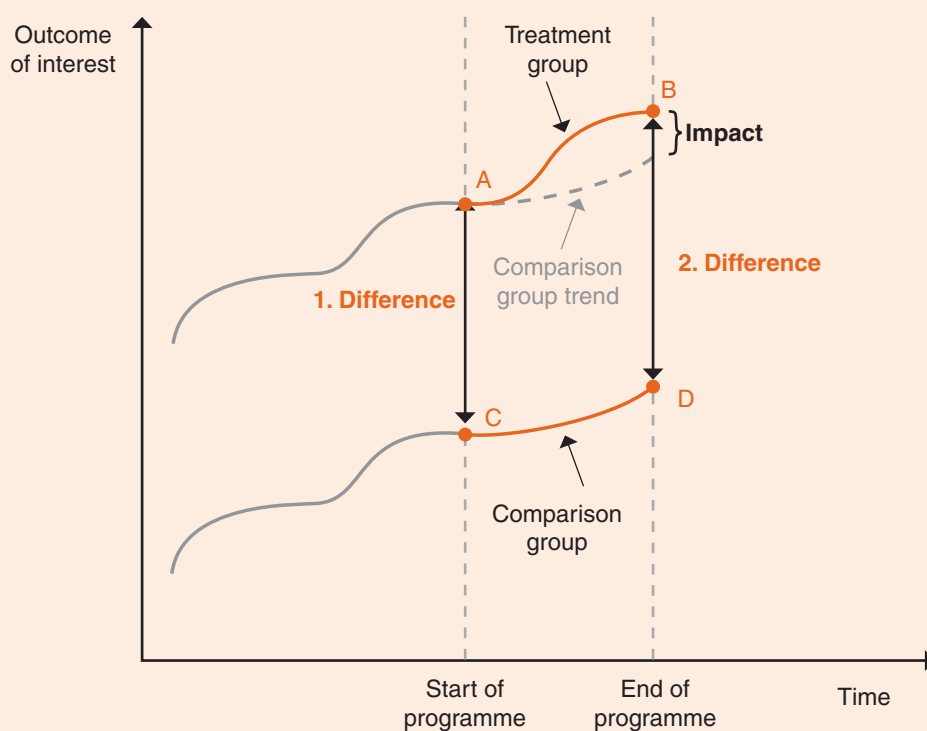
Identifying the comparison group: DID designs rely on having a comparison group whose development in our key outcomes of interest we can reasonably assume would be the same as the development of the treatment group over the time period of the intervention. To this end, it is desirable to choose groups with similar characteristics.

Let us imagine a six-month job training intervention for young people, for which we want to evaluate impacts on labour market outcomes. Randomly distributing training places is not possible. Instead, we take a sample of young people of similar age, education level, socio-economic background and labour market situation from another community as the comparison group.

Estimating the impact: To apply the DID evaluation technique, we need to (a) measure our outcomes of interest (for example, labour market status, see Note 2) for both the treatment and the comparison groups *before* the job-training intervention begins and (b) measure the outcomes of both groups at a

given time *after* the intervention took place. Even though we tried to identify a comparison group of young people that appear similar to the treated youth, it is likely that there are differences between the two groups prior to the job training, and that these differences remain afterwards. Figure 5.7 shows a situation where the comparison group has a considerably lower outcome indicator (say, employment status) at baseline. However, this does not affect the method. The DID technique compares the difference in outcomes between both groups at the end of the intervention (point B minus D) and adjusts it for the difference in outcomes between both groups at the beginning (A minus C). Subtracting these differences from each other (that is, taking a difference from two differences, which gives the method its name) yields an idea of the programme's impact; it shows whether and to what extent the training intervention increased employment status for participants relative to those who did not participate. The scenario in figure 5.7 indicates a moderate positive impact of the job-training intervention.

FIGURE 5.7: EXAMPLE OF DIFFERENCE-IN-DIFFERENCES ANALYSIS



$$\text{Impact} = (\text{2. Difference}) - (\text{1. Difference})$$

Source: Adapted from Gertler et al., 2016.

The “common trend” assumption: The assumption underlying this method is that, although the observable and unobservable characteristics of the treatment and comparison groups may be somewhat different (reflected in different levels of income at the beginning of the intervention), their *differences are constant over time*, or time-invariant. This allows us to use the trend of the comparison group as an estimate for what would have happened to our treatment group in the absence of the intervention. We therefore do not have to assume that without the intervention outcomes would have remained constant but rather that the treatment and the comparison groups share the same *trend* over time. This is what we refer to as the “common trend” assumption.

Coming back to the job training example above, in order to be able to apply DID, we have to be sure that over the next six months there will be no factors that systematically influence the outcomes of youths from the “treated community” differently to the outcomes of those from the “comparison community”, apart from the training assignment. For example, faster economic growth, a new local policy providing incentives to companies for the employment of young people or a major employer closing down in only one of the two communities would violate that assumption and consequently bias our evaluation results.

A good test to establish whether it is realistic to assume equal trends between participants

and non-participants is to compare their changes in outcomes before the intervention is implemented. This approach requires multiple data points prior to the intervention. As several baseline surveys can quickly become very costly, this test can more easily be carried out if administrative data on our key outcome indicators are available at little cost (for example, employment status from public

employment agencies or test scores from previous school years). If the outcomes of the two communities moved in tandem before the intervention started, we can be more confident that their outcomes would continue this trend during the intervention. If, however, pre-intervention trends are different, the equal trend assumption may not be correct.

WHEN CAN I USE A DID DESIGN?

Since it assumes that the differences between participants and non-participants are constant over time, this method is most usefully applied when there are good data available at multiple points before the intervention begins. To improve the credibility of impact estimates it is preferable to have at least three rounds of data collection: two prior to treatment, and at

least one at the end of the intervention (see above). This means that, unless data on participants and non-participants are available through other channels, such as an existing household survey, the costs of such an evaluation can be much higher than those of other impact evaluation techniques.

ADVANTAGES

- ▶ This method provides a way to account for both observable and unobservable differences between participants and non-participants. More precisely, it controls for all individual effects that remain constant over time, or that share the same course of change over time (i.e. treated and comparison groups show similar trends in the outcomes of interest).
- ▶ Even if the method is not experimental, it allows for a (partial) check of the assumption that renders it internally valid. This implies that we can have a sense of whether our estimated impacts are valid or not. If good administrative data is available, the method can be applied fairly easily and even ex-post, based on before and after data from the programme.

DISADVANTAGES

- ▶ This method produces less reliable results than randomized selection methods.
- ▶ In order to test the key assumption of “*common trends*”, at least three data collections are required, so the implementation can be expensive if data are not available initially.

Box 5.5: Evaluation of a labour market activation component for participants of a conditional cash transfer programme in Argentina

The Plan Jefes programme is a conditional cash transfer programme introduced during the Argentinian economic crisis of 2001–2002. Reforms of this programme following the recovery after the crisis included the implementation of the Training and Employment Insurance (Seguro de Capacitacion y Empleo, SCE) in 2006, in order to provide support in skills upgrading, vocational training, jobseeking and job placement to the eligible participants of Plan Jefes.

Participants in the SCE receive a monthly stipend and are provided with the following activation measures:

- ▶ assistance for the completion of primary and secondary education
- ▶ vocational training and apprenticeships
- ▶ labour intermediation services
- ▶ indirect job creation measures (e.g. employment subsidies)
- ▶ promotion of self-employment and micro-enterprise creation.

The ILO studied the effect of implementing these active labour market tools for beneficiaries of the Plan Jefes programme on their labour market status and job quality with DID estimators. In order to isolate the effect of these tools, a comparison group with similar features to those of the SCE participants had to be identified.

As the transfer from Plan Jefes to the new programme was gradual, the researchers could select participants in Plan Jefes who met the requirements to be beneficiaries of the SCE but had not yet been transferred to the new programme. An important key assumption of this identification strategy was that the transition between the programmes was not influenced by any factors which might be driving differences in the outcomes of interest. A total of 1,149 non-participants were selected, based on data from Argentina's Permanent Household Survey – a survey conducted quarterly by the Argentinian National Institute of Statistics (INDEC) which contains questions about individuals' personal characteristics, education and labour market performance. The selected participants and non-participants were similar in gender, age and level of educational attainment.⁵

The evaluators compared a range of outcomes between the two groups at two different moments in time (baseline and follow-up). This approach allowed causal effects of the SCE programme to be identified while controlling for selection bias due to observable and unobservable characteristics of the participants.

The panel structure of the survey allowed the researchers to gather data on both participants and non-participants at several points in time, both before and after programme participation. This allowed the assumption that, in the absence of the programme, the outcomes of participants and non-participants would have changed in the same way (common-trend assumption) to be tested and confirmed.

The study results showed that the programme had a positive effect on the participants' job quality, i.e. the probability of having a formal job and higher hourly wages, and a lower probability of having a low-paid job and working an excessive number of hours. It did not affect their employment status (i.e. the probability of being employed). The evaluation also showed heterogeneous effects, revealing that the programme had a higher impact for young beneficiaries, but no effect for women.

⁵ In order to correct for observed differences, the researchers also applied the propensity score matching (PSM) method. See the following section for a more detailed description of this methodology.

Propensity score matching (PSM)

Propensity score matching (PSM) is a very commonly used approach among the quasi-experimental evaluation methods. Its basic principle is to construct a comparison group by matching participants with similar

non-participants, based on their predicted probability of participating in the intervention. This is called the propensity score, which is calculated based on a range of observed characteristics.

HOW IT WORKS

A range of potentially relevant covariates have to be selected in order to calculate the propensity score for non-participants, based on their probability of being treated. The aim is to include in the propensity score calculation all covariates that affect both programme participation and outcomes. Non-participants are then matched with participants based by their respective scores. There are different ways of matching procedures, the most common approach being nearest neighbour matching, where each participant is matched to the non-participant with the closest propensity score. The closer the score, the better the matching quality. Balancing tests can be conducted to assess how well the matching worked. Consequently, the average difference of the two groups in the relevant outcomes of interest is equivalent to the impact of the intervention.

As an example, consider a skills training programme targeting rural youth which has 1,000 participants. Pre-programme data on key characteristics of the participants are available, e.g. sex, age, education and key aspects of their labour market history. Existing secondary survey data can be used to construct a comparison group based on their propensity score, estimating the probability of treatment for a large number of individuals based on the abovementioned characteristics. A total of 1,000 people with the best matching propensity scores will be selected as a comparison group for the intervention. Post-intervention data from the comparison could be gathered through the same secondary data source (if it is a panel data set, regular waves might be collected).

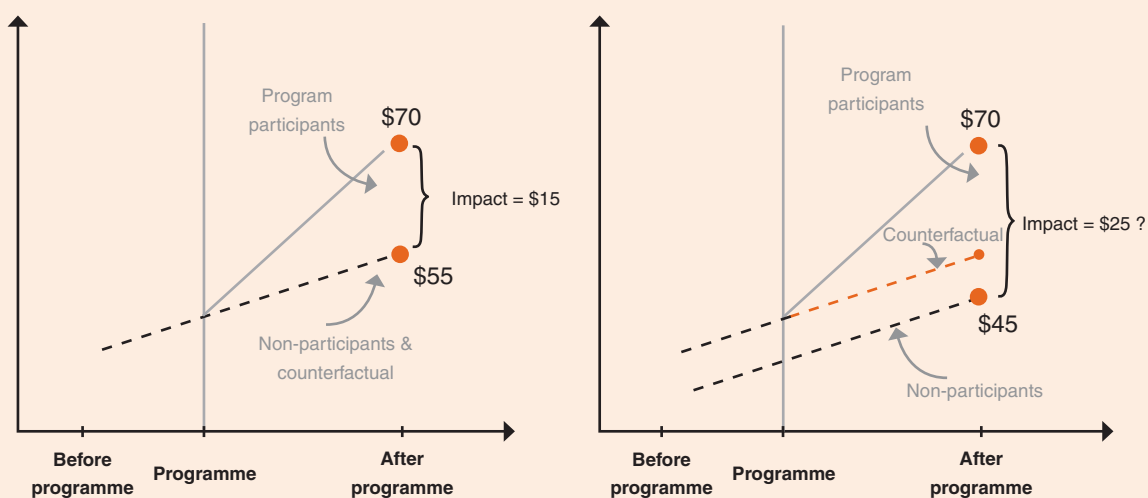
Box 5.6: Comparing participants and non-participants

Sometimes, although we might be able to identify a comparison group, we might only have available data on a few key outcome variables and no information on covariates, such as socio-economic background, knowledge, skills, etc. In these cases, we can use a simple impact evaluation methodology and compare outcomes of participants and non-participants. Thus, the counterfactual is estimated by the outcome of people who did not participate in the programme. However, this method is unlikely to yield either credible results or useful information about the true effect of our programme.

In particular, if non-participants (comparison group) differ from participants (treatment group) in ways that are relevant to the outcomes, this type of comparison will not be valid and will feature selection bias. More precisely, this method relies on two strong assumptions. First, we need to assume that programme participants and non-participants had, on average, similar outcomes at the beginning of the programme.

The right-hand side of figure 5.8 depicts a situation where participants already had a higher income at the beginning of the intervention than non-participants. This case leads to an overestimation of the true impact of our intervention.

FIGURE 5.8: COMPARING PARTICIPANTS AND NON-PARTICIPANTS



Second, we must assume that, in the absence of the intervention, both groups would have developed similarly over time. This requires the assumption that, on average, participants would have reacted in the same way as non-participants to all external factors. Note that in the situation described on the right-hand side of figure 5.8 this assumption holds true. The black dotted line, which describes how non-participants developed over time, and the red dotted line, which describes how participants would have developed in the absence of the intervention, move parallel over time. In order to obtain accurate impact estimates through this method, both assumptions must hold.

A key drawback of this method is that if we only observe outcomes after the intervention we cannot test either of the two assumptions directly and, in many cases, they might not be true. Consider our criterion for selecting young people for the intervention. It may be on a first come, first served basis. In this case, those with better access to information about the existence of the programme, those who live nearby, those who are encouraged by their parents or simply those who are more motivated to participate are more likely to end up being part of the intervention.

To summarize, comparing participants and non-participants at the end of an intervention without extensive knowledge of background variables that would allow more sophisticated techniques, such as PSM, to be employed is not advisable for an impact evaluation.

WHEN CAN I USE PSM?

PSM is a particularly useful method when large and rich amounts of secondary data are available, as these are necessary to define a good propensity score and to match sufficient numbers of participants and non-participants with similar scores, i.e. to find a large enough region of common support. Furthermore, PSM relies on the assumption that only observed factors influence both participation and outcomes (*conditional independence*

assumption). Thus, PSM should only be applied if there is a good understanding of the drivers of programme participation and the outcomes of interest, and should be avoided if unobservable characteristics can be expected to affect those variables. In any case, careful consideration is needed before the decision can be made on how many, and which specific variables to select for the estimation of the propensity score.

ADVANTAGES

PSM is a robust impact evaluation methodology which, if its assumptions are met, can help to remove selection bias and provide internally valid results. As in the case of the other quasi-experimental methods, it can be applied based on existing data sources and no random assignment of the intervention is necessary.

By matching on the propensity to receive treatment, PSM reduces the number of dimensions on which to match participants and comparison units to one, and thereby makes matching relatively straightforward.

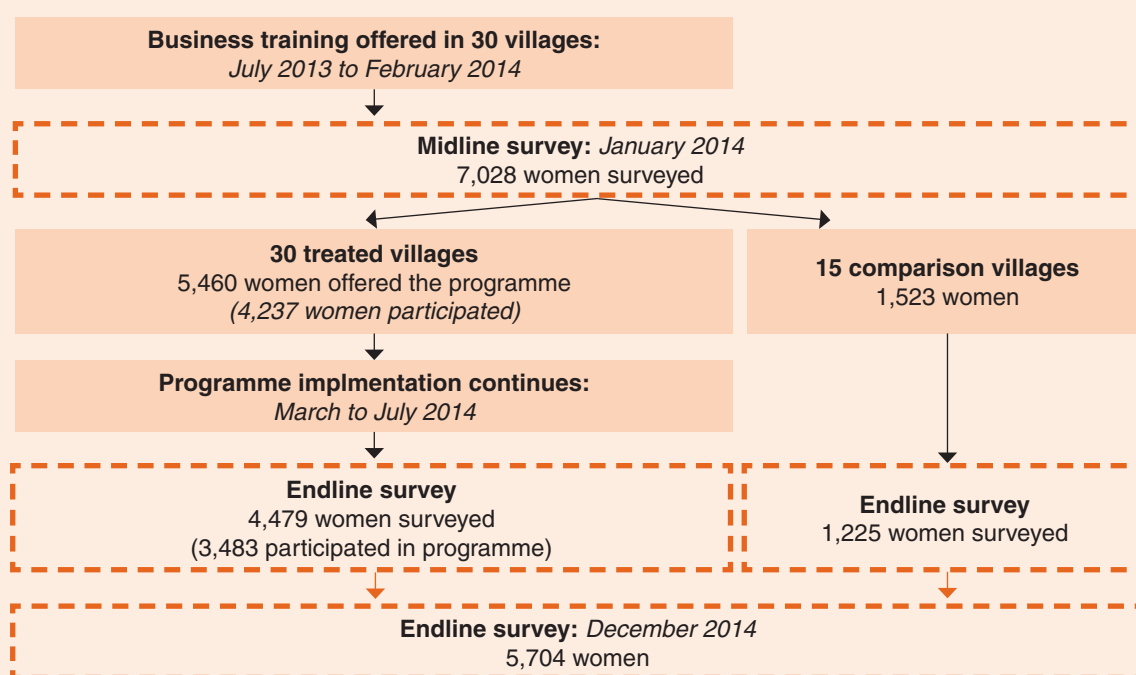
DISADVANTAGES

- ▶ The application of PSM usually requires large data sets.
- ▶ Matching can only be conducted on observable characteristics. Hence, the risk remains that selection bias due to unobservable characteristics driving programme participation can affect the evaluation results.
- ▶ The application of PSM is statistically complex and requires a corresponding level of expertise.

Box 5.7: Empowering young women through business and vocational training in rural Upper Egypt – the Neqdar Nesharek programme

From 2013 to 2014 the Population Council implemented the Neqdar Nesharek (meaning “We can participate”) programme in rural Upper Egypt. The programme targeted 4,500 young women aged 16–29 years old, adopting the “safe spaces” livelihood approach by addressing community-specific needs of vulnerable women. The intervention aimed to empower young women by providing them with business and vocational skills training and supporting them in starting a business or seeking employment. The training programme consisted of three main components: (1) business skills training, (2) vocational training and (3) life skills, legal rights and civic education.

FIGURE 5.9: IMPACT EVALUATION DESIGN (SIMPLIFIED)



The intervention was accompanied by an impact evaluation to assess the effect of Neqdar Nesharek on young women’s labour market outcomes and social empowerment measures. The evaluation used a PSM design. Impacts were calculated by matching women who participated in the programme with women with similar socio-economic characteristics from villages in the comparison group and comparing key programme outcomes between them (see figure 5.9).

FIGURE 5.10: IMPACTS ON LABOUR MARKET OUTCOMES



The evaluation found a significant impact of the programme on the economic empowerment of programme participants, as measured by their engagement in income-generating activities. Programme participants were 4.5 percentage points more likely to be engaged in an income-generating activity than women in the comparison group. As shown in figure 5.10, most of the positive impact was driven by an increase in participants' engagement in self-employment activities. In contrast, the level of participation in wage work did not significantly change for women in the treated group.

Source: ILO, 2017

Regression discontinuity design (RDD)

Regression discontinuity designs (RDDs) are often used when eligibility for a labour market intervention is based on some form of

continuous ranking of potential beneficiaries or applicants, for example a cut-off age.

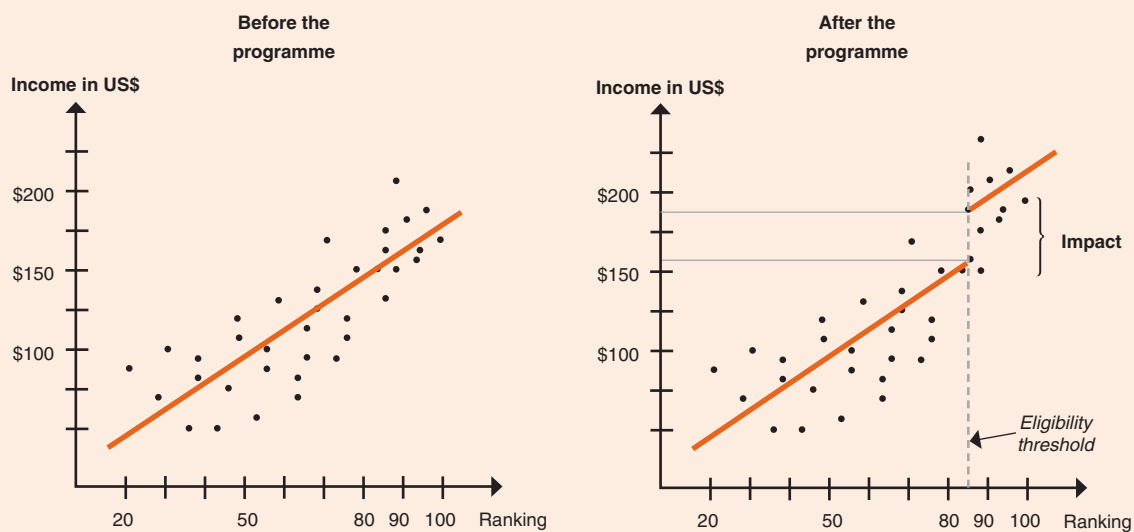
HOW IT WORKS

The premise of discontinuity (or eligibility-index) evaluation designs is that the people who score just above and just below a defined threshold are not very different from one another, or at least the difference may be continuous across the scores. For instance, 25-year-olds, who may be eligible for a youth skills training programme, are not likely to be very different from their 26-year-old peers, who may no longer be eligible. If we have a situation in which some of those youth who receive the programme (those just above the threshold)

and some of those who do not (those just below the threshold) are not fundamentally different from one another, then comparing the outcomes of these two groups, in turn, would allow us to analyse programme impact.

Figure 5.11 illustrates what we might find when analysing the impact of a youth microcredit initiative. The left-hand graph indicates that, at the time of applying to the programme, those who achieved better scores already tended to have higher incomes. There may be many reasons for

FIGURE 5.11: SAMPLE DISCONTINUITY CHART



this, for example, that those with a slightly higher level of education are already earning more and that their education also helped them to secure better scores. Or those who are more motivated to start a business are already more entrepreneurial, reflected in their higher incomes, and that motivation also helped them to convince the jury to support them. Many other explanations are possible, which we do not necessarily need to understand to apply this method.

When starting the programme, the local micro-finance bank decided that the threshold for receiving a loan was 85, and all applicants

were accepted or denied support according to their ranking relative to that threshold. Now we would like to establish whether the microcredit programme had any impact on incomes. As illustrated in figure 5.11 (right-hand graph), we assume that those who received a score below 85 have the same outcomes as previously, while the income of those with a score of 85 and above increased across the board. From this information, it is possible to identify the impact of the programme, which will be represented by the difference in outcomes (that is, the discontinuity of the linear relationship) near the cut-off point.

WHEN CAN I USE RDD?

In many cases we are not able to plan the evaluation during the programme design. Sometimes, however, we may be able to use the targeting rules of the programme to obtain a good comparison group ex-post. Some programmes use a continuous ranking of potential beneficiaries, such as test scores, credit scores

or poverty index, and have a cut-off point for acceptance into the programme. In the case of youth labour market interventions, there is often a binding age cut-off. Only youth under a specified age are eligible for the programme. This eligibility rule can be used for conducting an impact evaluation based on RDD.

ADVANTAGES

- ▶ The RDD can be applied ex post, if sufficient administrative data are available.
- ▶ It can take advantage of an existing rule for assignment to construct a valid comparison

group and thereby does not require the exclusion of an eligible group from the intervention.

DISADVANTAGES

- ▶ The main requirement for using discontinuity designs is that programme participation is determined by an explicitly specified targeting rule; in other words, by a continuous scale or score. For this method to work, we need many observations in the region immediately above and below the cut-off point in order to have sufficient numbers of youth to compare with one another. Unless the evaluation is done without baseline data or can take

advantage of existing programme records, a discontinuity design requires similar data collection to a lottery design, and therefore bears a similar cost.

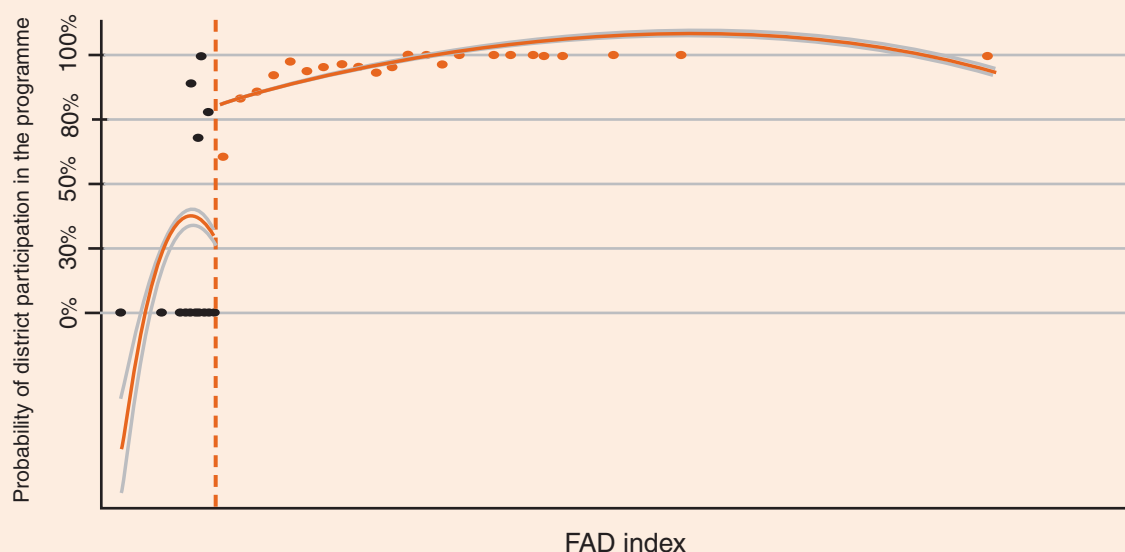
- ▶ The informative value of the results is limited to the sample around the cut-off point. This might be relevant, for example, in discussions regarding whether a programme should be scaled-up to include other age groups or regions.

Box 5.8: Effectiveness of a workfare programme on labour market outcomes: Construyendo Perú

Public works programmes are an increasingly popular policy tool in developing countries. From 2007 to 2011 the Government of Peru implemented the programme Construyendo Perú with the primary objective of supporting unemployed individuals in situations of poverty. The programme provided them with access to temporary employment and skills development training through the financing of public investment projects with intensive use of unskilled labour.

The ILO evaluated the medium- to long-term effects of the programme using a regression discontinuity approach. The evaluation exploits an interesting assignment rule of the programme at the district level that consisted in selecting beneficiary districts by ranking them according to the FAD (Factor de Asignación Distrital) index. The FAD is a composite index that combines demographic information with an index of human development shortcomings and a poverty severity index. As such, districts with an FAD index above a certain threshold (i.e. those with higher poverty and development shortcomings) were allowed to participate in the programme and districts below that threshold did not participate in the programme. This is an example of a fuzzy regression discontinuity design. As shown in figure 5.12, districts just above the cut-off point were considerably more likely to participate in the programme than those just below the threshold.

FIGURE 5.12: DISCONTINUITY IN THE PROBABILITY OF DISTRICTS PARTICIPATING IN THE PROGRAMME



The reasoning behind the evaluation is therefore to estimate the causal impact of the programme by comparing outcomes of individuals around the cut-off point of the FAD index. The evaluation found that over the medium-term (three to five years) the intervention helped to increase employment and reduce inactivity for women and less-well educated programme participants. However, the programme was not able to improve the prospects of lower-educated participants in terms of job quality and, in fact, had a detrimental impact on job quality perspectives of women and more highly educated individuals (for example by increasing the probability of informal employment).

Source: Escudero, 2016

Simple comparisons: Before and after

Sometimes randomization is not possible and, moreover, the conditions for a valid quasi-experimental evaluation do not hold; for example, if we cannot find a suitable comparison group with baseline information available and/or if the common trend assumption cannot be confirmed. In these cases, it is advisable to consider whether it is worth conducting a quantitative impact evaluation at all.

If it is not possible to include a comparison group in an impact evaluation, the most basic approach relies on simply comparing the outcomes of programme participants before and after the intervention. This simple approach can give an idea of the change that occurs over the course of an intervention but should be regarded as part of a monitoring system rather than as providing evidence of the causal impact of an intervention since there is no way of knowing if an observed change should be attributed to the intervention in question or to other circumstances.

Taking the example of a training programme, we may observe that the monthly income of participants increased from \$50 before the intervention to \$60 after the intervention and therefore conclude that the impact of the intervention was \$10 per month per person (see figure 5.13, left-hand graph). However, if in the absence of the intervention the income level could have increased anyway due to a change

TIP



Before and after comparisons are sometimes also referred to as *tracer studies*, especially in the context of standardized surveys administered to graduates from secondary or tertiary educational or (Technical and) Vocational Education and Training programmes. Tracer studies have only limited means to evaluate impact but are powerful tools for measuring the employability of graduates and collecting feedback to improve the study programme. For a detailed guide to carrying out tracer studies see [Schomburg \(2016\)](#).

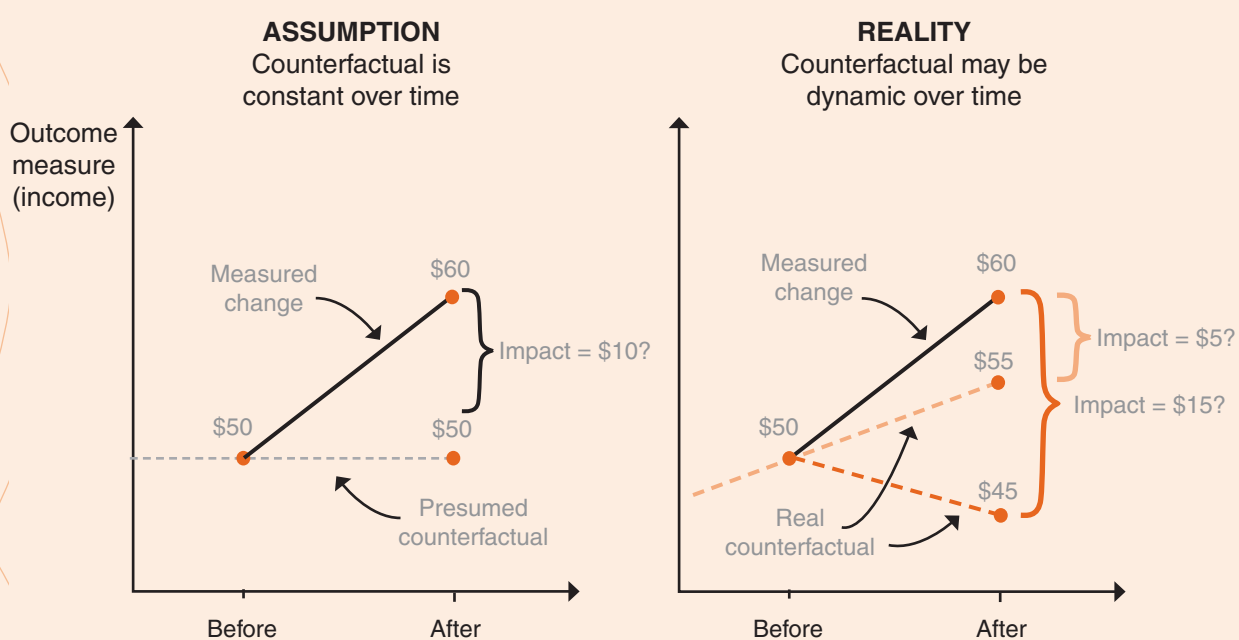
in circumstances (i.e. the situation we are facing corresponds to one of the scenarios shown in the right-hand graph of figure 5.13), we will not be able to obtain an accurate estimate of the intervention.

Since the real counterfactual scenarios (the dotted lines in the figure) cannot be observed, there is no way of knowing if the case that applies in a particular evaluation is the one shown on the left-hand side of figure 5.13 or the one on the right. It is therefore impossible to have a sense of whether the impact we estimate with this method is the true impact of our intervention or a “contaminated” one.

Performing before-and-after comparisons could make sense if there are reasons to believe that, in the absence of the treatment, outcomes would, on average, remain unchanged. This could apply to interventions that (a) are delivered over a *short period of time* (for example, short skills training interventions, job-counselling services or events that aim to change the attitudes of participants) and (b) are expected to have *effects of interest in the short term*. However, the above-mentioned limitations persist and, unlike

well-implemented experimental and quasi-experimental methods, simple before-and-after comparisons cannot be considered robust impact evaluations. Their level of robustness can be improved, first, by controlling for potential confounding factors in a regression model (instead of simply comparing the outcomes) and/or, second, by the complementary application of qualitative methods in order to work out the causal mechanisms underlying the observed change in outcomes.

FIGURE 5.13: COMPARING BEFORE-AND-AFTER OUTCOMES



Improving the relevance of quantitative impact evaluations

As shown in the preceding section, there is a range of good quantitative methods to provide an internally valid answer to the basic evaluation question, “Did the project work?”; that is, “Did it affect the outcomes of interest as defined in our intervention and learning objectives?”. The question of whether the intervention as a whole had an impact is an important one, but it is by no means the only question we may want to ask.

In order to gain a detailed and holistic understanding of how and why the effects of a youth employment programme unfold, we need to “dig deeper”. Having a clear understanding of the heterogeneity of impacts and the causal mechanisms leading to observed effects also helps us to derive valuable lessons from our evaluation and gain a better insight into whether a programme is likely to work in other settings.

This brings us back to the important issue of internal and external validity (see the beginning of this note). In order to achieve internal validity, our methods need to be robust and properly implemented. In order to achieve external validity, we need to understand the relevant contextual factors of our programme and their potential effect on the evaluation results. It is difficult to reach achieve objectives using a single method.

For example, experimental evaluations, if properly implemented, can give us credible information about the impacts that can be uniquely attributed to our project, but can tell us very little about their replicability in other settings. Critically, quantitative impact evaluations tell us “what” happened – the average treatment effect – but they do not tell us “why”. For this purpose, the complementary application of qualitative methods is required.

MEASURING A VARIETY OF IMPACTS

First, it may be useful to have a more nuanced picture of the programme’s actual impact. This can be partly achieved within the quantitative designs described above. Relevant questions to ask could be:

- ▶ Do outcomes vary across different groups of beneficiaries (e.g. young men benefit, but young women do not)?
- ▶ What is the short-term versus the long-term impact of the intervention?
- ▶ Does the intervention have positive or negative spillover effects? Are there any intended

or unintended outcomes beyond the actual target group?

Second, we may also be interested in testing cross-cutting designs, testing how the effectiveness of our intervention changes as we modify the design. These designs allow us to investigate the following questions:

- ▶ Is one intervention design more effective than another? We may want to compare alternative interventions (providing start-up grants versus start-up loans for young entrepreneurs, for example), or test the

most effective combination of programme components (training alone, training plus internship, or training plus internship and mentoring).

- ▶ What is the most effective dosage of the intervention? For example, should we provide 20, 50 or 100 hours of training (see table 5.3 for further impact evaluation questions)?

Cross-cutting designs help to identify more than just the overall impact of a project; they also evaluate specific intervention features and why these do or do not work. For example, a programme may provide vocational

and entrepreneurial skills training, such as carpentry or tailoring, along with a small amount of start-up capital for businesses. The provision of cash grants could be expensive or politically difficult, and so the programme director may wonder whether the start-up capital is necessary, or if participants are able to implement their training without the capital. A cross-cutting design can help to determine the best project design in this case. In practice, this requires us to compare the outcomes of different treatment groups to a comparison group and to each other.

COMBINING QUANTITATIVE AND QUALITATIVE APPROACHES

Furthermore, we may be interested in shedding light on the channels through which the impact operates – that is, understanding why and how an impact unfolds. For example, we might want to answer the following questions:

- ▶ How and why did things happen as observed?
- ▶ Why did a project (or part of it) not work as we expected?
- ▶ What can we learn from failure?

If such potential avenues of investigation are envisaged at the design stage, theories can be tested partially within the quantitative methods above. To achieve this end, the surveys must include questions designed to capture the different factors (intermediate outcomes) through which the impact is hypothesized to operate in order to verify if the intervention affects these intermediate outcomes.

However, many outcomes of youth employment interventions (such as mental health, empowerment or household relations) are complex and multidimensional and may not be captured with quantitative methods. Mixed methods allow for tracking qualitative indicators and provide selected case study analysis

to help develop a better understanding of the dynamics and results of the intervention. For example, structured and semi-structured qualitative interviews, in which participants are free to express real-life stories that fall outside categories of quantifiable information, can help to round out an understanding of a programme's impact (Bamberger et al., pp. 6–7; Leeuw and Vaessen, 2009).

Qualitative data collection methods might be particularly useful for collecting information about how well the intervention was implemented (see Note 4 on performance evaluations). Understanding the implementation process is crucial to discovering how the intervention implementation affected results and correctly interpreting findings to determine whether disappointing results are due to weaknesses in intervention design or in implementation. Furthermore, qualitative techniques might shed light on why specific findings transpired and, in particular, why effects differed across the target population (for example, between rural and urban young people or between young women and young men).

Table 5.3: Categories of impact evaluation questions

Question	Description	Additional data requirements	Sample evaluation result and interpretation
<p>What is the overall intervention impact on outcomes A, B, and C in group X? ...in context Y?</p>	<p>This is the standard impact evaluation question</p>	<p>n/a (standard data collection based on the method chosen)</p>	<p>The average impact of the training intervention on the income of youth is +\$20 per month. The intervention has a positive impact on participants' income</p>
<p>Do the outcomes vary across population groups?</p>	<p>Interventions often affect groups differently (heterogeneity of impacts). Measuring only average impact may hide these differences, so we need to break down impacts by population group</p>	<ul style="list-style-type: none"> Socio-demographic information of participants and comparison group (age, gender, income level, etc.) To be able to disaggregate the results, the number of people covered by the evaluation (the sample size) needs to increase with each category of information that is to be analysed 	<p>The average increase in income is \$40 for boys and \$0 for girls. Older youth benefit more than younger youth (\$30 versus \$10, on average). Therefore, the intervention is not equally effective for all participants. We need to understand why groups benefit to a different extent and possibly adapt the programme's targeting and design to accommodate particular groups</p>
<p>What is the short-term versus the long-term impact of the programme?</p>	<p>The change in outcomes may not be constant over time. Short-term effects may vanish, while long-term effects may not manifest themselves until years after the intervention has ended</p>	<p>Data over an extended period of time (in practice, this often means following treatment and comparison groups for several years)</p>	<p>At the end of the programme, we observe an average monthly income for participants of -\$5 (a loss) compared with the controls. Two years after the programme, the average increase in monthly income for the treatment group is \$20. Those who participated in the training were not able to work as much as their peers during the course of the training, so they lost income. Over time, however, the training paid off and participants were able to secure higher incomes than their counterparts who did not participate. Looking only at short-term outcomes may provide misleading results</p>
<p>Does the intervention have spillover effects?</p>	<p>The intervention may have indirect effects on non-participants (positive and negative)</p>	<ul style="list-style-type: none"> Data beyond the treatment and comparison group, to include family or community members Several treatment groups (one receives design A, one receives design B, etc.) The number of people covered by the evaluation needs to be large enough to be able to create more than one treatment group as well as a comparison group 	<p>Not only do participants have a \$20 higher average income, their neighbours also experienced a \$5 increase. Participants apparently passed on the knowledge to others</p>
<p>Is intervention design A or intervention design B more effective?</p>	<p>There is often ambiguity about the best possible intervention design. Questions can relate to comparing alternative interventions or combinations of programme components</p>	<ul style="list-style-type: none"> Several treatment groups (one receives design A, one receives design B, etc.) The number of people covered by the evaluation needs to be large enough to be able to create more than one treatment group as well as a comparison group 	<p>The average increase in income is \$5 for those who received training and \$30 for those who received training and an internship. Thus, providing practical work experience in addition to training appears to significantly improve impact</p>
<p>What is the most effective dosage of the intervention?</p>	<p>More is not always better; finding the right balance of how much service to provide is important to maximize impact on the one hand and minimize costs on the other</p>	<ul style="list-style-type: none"> Several treatment groups (one receives design A, one receives design B, etc.) The number of people covered by the evaluation needs to be large enough to be able to create more than one treatment group as well as a comparison group 	<p>The average increase in income is \$0 for those who received 1 month of training, \$20 for those who received 3 months, and \$20 for those who received 6 months. Although 1 month of training was insufficient, 6 months of training had no additional benefit compared with 3 months of training. The optimal length of the training seems to be about 3 months</p>
<p>Why did the intervention (not) work? Why did it only work for part of the target population/when the intervention had a certain duration?</p>	<p>Along with assessing the impact itself, it is crucial to understand how and why it took place as observed</p>	<p>Both quantitative and qualitative data, ideally triangulated to establish reasonable causal connections. For example, in-depth interviews with training participants, trainers and employers</p>	<p>Employers valued specific skills which could realistically be acquired by participants in 3 months' time. Remaining longer in training did not add additional value to the employees' skillsets. Employers did not want to incentivize longer training duration as they would lose employee working time during the training Training only led to an increase in income for boys as employers tend to assign different tasks to girls, in which the skills transferred in the training are less in demand</p>

Box 5.9: Process tracing

Process tracing involves the in-depth analysis of the different events linking an intervention to one or more intermediate or final outcomes, and their causal relations. Often (but not always) process tracing methods aim to develop and test theoretical mechanisms, which can be generalized to cover other interventions and contexts. In summary, process tracing is applied as follows:

1. Developing a hypothesized causal mechanism for how change happens

As a first step, we have to build the narrative of the process we are going to assess. This can be a project's theory of change, including the people and activities involved in it. It is important that the process is set out in its smallest individual elements, which should all be both essential for it to work and measurable.

For example: “Teachers conduct skills training for unemployed youth”; “Students attend skills training”; “Students acquire new knowledge and skills about how and where to look for jobs”; “Students search more and more efficiently for jobs”; “Students have a higher probability of finding a job”.

In order to make plausible claims regarding the causal linkages between the different parts of this mechanism, it is necessary to identify possible alternative explanations for the occurrence of each individual element and to look for evidence to confirm or rule out those explanations. For example: “Students acquired the knowledge on how and where to look for jobs independently”; “Students find jobs because of an improvement in the local labour market situation”.

2. Defining and collecting the required evidence

After defining the mechanism, or our theory, we need to define the empirical evidence required to analyse each link in the causal chain. This applies both for “our” mechanism and for the competing alternative hypotheses. Consequently, the previously identified evidence will be gathered through primary and/or secondary data collection. Sources for this evidence can be stakeholder interviews, programme documents, survey data, meeting minutes, and statistics, among others. Evidence should be collected in such a way that it can either confirm or refute the different competing hypotheses. It is good practice to triangulate methods, i.e. to use different methods to assess the same element from different angles.

3. Assessing the evidence and drawing a conclusion

The collected evidence is then examined in a procedure similar to that used in a criminal trial. In process tracing, we aim to establish a case that offers sufficient proof to reasonably assume that each element of the mechanism took place due to another element and that together they caused certain outcomes.

There are different tests for assessing the strength of the evidence for each hypothesis. For example, the “smoking gun” test refers to convincing evidence directly referring to the mechanism in question. So, for example, a statement from a skills training participant such as: “Thanks to the things I learned in the training, I have much more confidence to apply for jobs and I send out more applications than before” can make us fairly confident that this participant did not increase his or her job-search behaviour – which could be an intermediate outcome variable measured quantitatively – due to other reasons.⁶

When assessing the evidence for different competing hypotheses, it is important to bear in mind that the strength of the overall evidence for a certain mechanism is always only as robust as the weakest evidence for one individual link. Finally, based on the conclusions from this exercise, the hypothesized mechanism, as well as the alternative hypotheses will be confirmed or ruled out.

⁶ For more tests and details on their application, see Bennett, 2010.

Rather than being substitutes for a quantitative impact evaluation, several of the abovementioned evaluation strategies can contribute to assessing a specific intervention. Using a mixed-methods approach therefore allows us to combine the strengths and offset the weaknesses of both qualitative and quantitative evaluation tools, allowing for a stronger evaluation design overall.

Employing a mixed-methods design for an impact evaluation in practice implies collecting both qualitative data, for example through field visits, key informant interviews or focus group discussions, and quantitative data, relying for instance on administrative data, surveys or secondary data sources, such as household surveys (see also table 3.5 in Note 3).

Mixed-method evaluation designs are also closely related to and inform theory-based impact evaluations. As [White and Phillips \(2012\)](#) observe, theory-based impact evaluations aim to establish causal links “by collecting evidence to validate, invalidate, or revise the hypothesised explanations, with the goal of documenting the links in the actual causal chain”. They often seek to combine all available quantitative and qualitative evidence to establish beyond reasonable doubt that an intervention impacted its participants. Box 5.9 introduces the methodology of process tracing as a theory-based impact evaluation methodology.

KEY POINTS

1. **Impact evaluations answer cause-and-effect questions to determine whether and to what extent an intervention caused observable change.** Understanding impact requires us to isolate the effects of the intervention from other factors influencing beneficiary outcomes.
2. **Quantifying impacts of interventions requires estimating the counterfactual;** that is, what would have happened to beneficiaries in the absence of the intervention. To this end, most quantitative impact evaluation designs rely on having a comparison group that shares as many characteristics with the beneficiaries as possible.
3. **Observational impact evaluation designs include difference-in-differences and matching methods.** They can be applied in a broad range of contexts and based on secondary data sources, but for some interventions these methods might not be able to estimate impacts credibly. Experimental designs that rely on some degree of randomization can produce highly credible impact estimates but can be costly and difficult to implement for certain interventions.
4. **To maximize learning about “why” interventions worked, or did not work, used mixed-method approaches** which build on qualitative and quantitative data and make use of several methodologies for analysis.

KEY RESOURCES



- ▶ Card, D.; Ibararán, P.; Villa, J.M. 2011. *Building in an evaluation component for active labor market programs: A practitioner's guide*, IZA Discussion Paper No. 6085 (Bonn, Institute for the Study of Labour (IZA)).



- ▶ 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).



- ▶ Duflo, E.; Glennerster, R.; Kremer, M. 2006. *Using randomization in development economics research: A toolkit*, BREAD Working Paper No. 136 (Bureau for Research and Economic Analysis of Development).



- ▶ Khandker, S.R.; Koolwal, G.B.; Samad, H.A. 2010. *Handbook on impact evaluation: Quantitative methods and practices* (Washington, DC, International Bank for Reconstruction and Development (IBRD) and World Bank).

REFERENCES

- ▶ Bamberger, M.; Rao, V.; Woolcock, M. 2010. *Using mixed methods in monitoring and evaluation: Experiences from international development*, Policy Research Working Paper No. 5245 (Washington, DC, World Bank).
- ▶ Bennett, A. 2010. "Process tracing and causal inference", in H. Brady and D. Collier (eds.) *Rethinking social inquiry: Diverse tools, shared standards* (2nd edn) (Lanham, MD, Rowman & Littlefield), pp. 207–220.
- ▶ Card, D.; Kluve, J.; Weber, A. 2009. *Active labour market policy evaluations: A meta-analysis*, IZA Discussion Paper No. 4002 (Bonn, Institute for the Study of Labour (IZA)).
- ▶ Card, D.; Ibararán, P.; Villa, J.M. 2011. *Building in an evaluation component for active labor market programs: A practitioner's guide*, IZA Discussion Paper No. 6085 (Bonn, Institute for the Study of Labour (IZA)).
- ▶ Duflo, E.; Glennerster, R.; Kremer, M. 2006. *Using randomization in development economics research: A toolkit*, BREAD Working Paper No. 136 (Bureau for Research and Economic Analysis of Development).
- ▶ Escudero, V. 2016. *Workfare programmes and their impact on the labour market: Effectiveness of Construyendo Perú*, Research Department Working Paper No. 12 (Geneva, ILO).
- ▶ Fiala, N. 2015. *Access to finance and enterprise growth: Evidence from an experiment in Uganda*, Employment Working Paper No. 190 (Geneva, ILO).
- ▶ 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. *Empowering young women through business and vocational training: Evidence from rural Upper Egypt*, Taqueem Impact Brief Series, Issue 10 (Geneva).
- ▶ Khandker, S.R.; Koolwal, G.B.; Samad, H.A. 2010. *Handbook on impact evaluation: Quantitative methods and practices* (Washington, DC, International Bank for Reconstruction and Development (IBRD) and World Bank).
- ▶ Leeuw, F.; Vaessen, J. 2009. *Impact evaluations and development: NONIE guidance on impact evaluation* (Washington, DC, Network of Networks on Impact Evaluation).
- ▶ Schomburg, H. 2016. *Carrying out tracer studies: Guide to anticipating and matching skills and jobs*, Vol. 6 (Luxembourg, ETF, CEDEFOP, ILO).
- ▶ White, H.; Phillips, D. 2012. *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework*, Working Paper 15 (New Delhi, International Initiative for Impact Evaluation).



Case study

ASSESSING RURAL MICRO-ENTERPRISE GROWTH THROUGH DIFFERENT EVALUATION METHODS

Disclaimer: This is a fictional case study.
All information contained within has been invented for learning purposes.

Learning objectives

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

- ▶ identify impact evaluation methods without being told the specific employed method
- ▶ explore the problem of producing estimates of the causal impacts of a development programme, and the various ways of estimating the impacts using comparison group designs
- ▶ develop an intuitive understanding of when and how impact evaluation methods will produce biased results by learning about the concept of selection bias and how comparison group designs are only as good as their ability to eliminate selection bias.

Introduction and case study context

Micro-enterprises are vital in rural areas with limited formal employment options, both for providing informal employment and for ensuring household economic security for business owners. However, once a business has been started, there are a number of challenges to growth.

What can be done to help develop rural businesses? The Training for Rural Economic Empowerment (TREE) programme tests some of these constraints to understand what kind of financial and training services have impacts on enterprise growth, for whom and why.

The International Labour Organization (ILO) conducted the training tested here using

their TREE methodology, a development approach which ensures that women and men living in poverty gain the skills and knowledge they need to improve their incomes and take a more active role in shaping their communities. Moreover, a local microfinance organization delivered loans to individuals of US\$200 at a discounted annual interest rate of 20 per cent (reduced from the standard 25 per cent).

This case study focuses on 400 rural micro-enterprise owners who were offered the chance to participate in a skills training programme and to receive loans. In total, 144 out of the 400 business owners took part in the training and received the loan.

Comparing different impact evaluation methods: Did TREE work?

Did the TREE programme work? Did the programme improve business profits? What is required in order for us to measure whether a programme worked, or whether it had an impact?

In general, to ask if a programme works is to ask if the programme achieves its goal of changing certain outcomes for its participants, and ensure that those changes are not caused by some other factors. We need to simultaneously show that, if the programme had not been implemented, the observed changes would not have occurred (or would be different). But how do we know what would have happened? Measuring what would have happened in the absence of the programme requires entering an imaginary world in which the programme was never offered to these participants. The outcomes of the same participants in this imaginary world are referred to as the counterfactual. Since we cannot observe the true counterfactual, the best we can do is to estimate it by mimicking it.

The key challenge of programme impact evaluation is constructing or mimicking the counterfactual. We typically do this by selecting a group of people that resemble the programme participants as much as possible but who did not take part in the programme. This group is called the comparison group, which ideally differs from the group of beneficiaries only insofar as they did not participate in the programme.

We then estimate “impact” as the difference observed at the end of the programme between the outcomes of the comparison group and the outcomes of the programme participants.

Importantly, the impact estimate is only as accurate as the comparison group is successful at mimicking the counterfactual. Therefore, the method used to select the comparison group is a key decision in the design of any impact evaluation.

That brings us back to our questions: Did the project work? What was its impact on the outcome being evaluated, namely business profits?

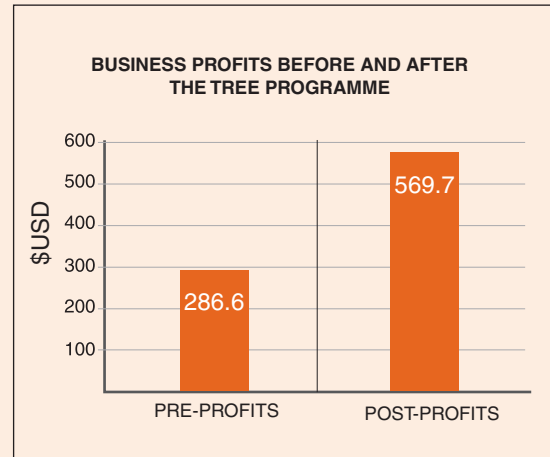
In our case, the intention of the programme is primarily to “improve enterprise growth”, and profits (measured in US\$) are the key outcome indicator. So, when we ask if this project worked, we are asking if it improved business profits. The impact is the difference between profits after the businesses have been exposed to the intervention and what their profits would have been if the intervention had never existed.

What comparison groups and impact evaluation methods can we use? The following (fictional) experts illustrate different methods of evaluating impact. Table 5.2, at the beginning of this note, presents an overview of different evaluation methods for your reference.

Part I: News release: “Training for Rural Economic Empowerment” programme helps businesses grow

TREE celebrates the success of its programme. It has made significant progress in its goal of helping businesses grow through provision of loans and skills training. The achievement of the TREE programme demonstrates that providing skills training to business owners, combined with loans to ease capital constraints, can produce significant gains.

Just before the programme started, businesses were making profits of \$286, on average. *But after spending just a few months in the programme, profits for these businesses doubled!*



Discussion topics

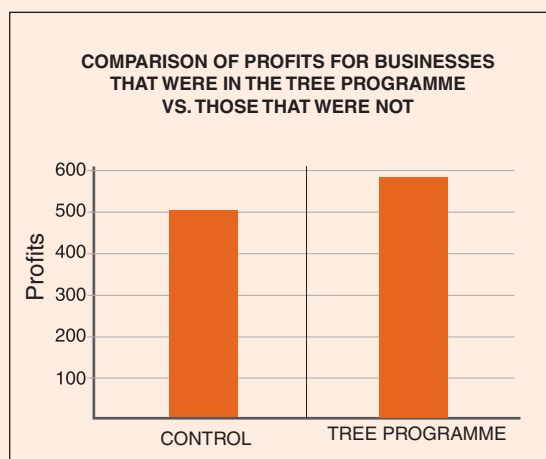
1. What type of evaluation does this news release imply?
2. What represents the counterfactual?
3. What are the challenges with this type of evaluation?

Part II: Opinion: The “Training for Rural Economic Empowerment” project not up to the mark

With an estimated outreach of 6 million trainees, a continuously growing network of more than 17,000 trainers and 200 master trainers in 2,500 partner institutions, TREE is one of the biggest training systems used for the support of micro- and small enterprises (MSEs) currently on the market. But do the profits of its businesses actually double in size, as suggested in the first example? Recent evidence suggests otherwise.

► An independent team of evaluators was hired to verify these findings. The team compared profits of TREE businesses to

profits of other businesses in nearby villages. They found that TREE businesses grow their profits by only a meagre \$64, and not \$286 as originally estimated. That's only a 12 per cent increase in profits after 6 months of the TREE programme paired with loans. It seems that income estimates were severely overestimated and that ILO's assurances about the successes of the programme were false.



Discussion topics

1. What type of evaluation does this opinion piece imply?
2. What represents the counterfactual?
3. What are the challenges with this type of evaluation?

Part III: Letter to the Editor: Independent evaluators should consider evaluating fairly and accurately

There have been several unfair reports in the press concerning programmes implemented by the ILO. A recent article by an independent evaluator claims that TREE is, in reality, not helping businesses grow. However, their analysis uses the wrong metric to measure impact. It compares the profits of TREE business with other businesses in the village – not taking into account the fact that TREE targets those whose profits are particularly low initially. If TREE simply recruited the biggest businesses into their programmes, and compared them to their smaller counterparts, they could claim success without conducting a single training session or providing a single loan. But TREE does not do this. And, realistically, TREE does not expect its smaller businesses

to overtake the bigger businesses in the village. It simply tries to initiate an improvement over the current state.

Therefore the indicator should be *improvement* in profits – not the final profit level. When we repeated the analysis using the more appropriate outcome measure, the TREE businesses improved at twice the rate of the non-TREE businesses (US\$283 profit increase compared to US\$162). Had the independent evaluators thought to look at the more appropriate outcome, they would recognize the incredible success of TREE. Perhaps they should enrol in some TREE training themselves.



Discussion topics

1. What type of evaluation does this letter imply?
2. What represents the counterfactual?
3. What are the challenges with this type of evaluation?

Part IV: Designing your own evaluation to assess the impact of TREE

As discussed earlier in this case study, there are challenges and reservations with respect to all three of the evaluation methods detailed above. It is now your turn to design an impact evaluation for the TREE programme, assuming that the programme is yet to be implemented.

To begin, assume that your research team has surveyed several thousand micro-enterprises from a census of businesses and selected 1,600 business owners to be included in the evaluation. All of these business owners have expressed an interest in receiving the ILO training and in participating in the loan programme. However, due to resource constraints, your project manager tells you that

the training and loan programme can only be offered to a maximum of 800 businesses.

We also assume that the key outcome of interest remains “business profits”.

1. How would you design the evaluation? In particular, how would you select a comparison group?
2. When would your research team collect data and from which businesses (all 1,600 or only a subset)?
3. Why do you think this is a reliable impact evaluation method that overcomes some or all of the shortcomings of the three methods discussed above?

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. 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 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.

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

Printed in Switzerland

Contents

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

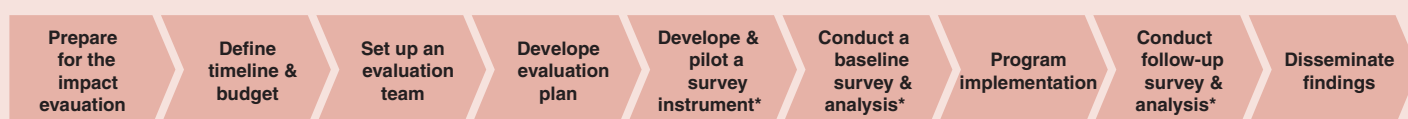
Before initiating an impact evaluation, the following questions need to be asked:

- ▶ **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\$)
A. Staff salaries												
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
A. Consultant fees												
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
C. Travel and subsistence												
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
D. Data Collection*												
Surveying					Youth	40	2,000	80,000	Youth	60	2,000	120,000
E. Dissemination												
Report, printing										5,000	1	5,000
Workshop (s)										5,000	1	5,000
Total cost per stage				28,250				110,750				166,250
Total evaluation cost												305,250

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

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/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"> • Anticipate possible consequences for the children and youth involved. Do not proceed unless potentially harmful consequences can be prevented or mitigated • 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 • 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 • 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 • Do not raise unrealistic expectations about the benefits or rewards of participation • If any, provide only modest rewards or incentives to participate that are in line with local living standards
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"> • 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.) • When working with minors, always seek informed consent from parents or guardians, • 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
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"> • Prior to interviewing young people, try to collect as much information as possible from alternative indirect sources (adults, administrative records, etc.) • 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) • When necessary, acknowledge that questions can be sensitive, and anticipate and address the concerns of parents and participants • 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
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"> • Always ensure the privacy and confidentiality of responses from parents and young participants, which will also strengthen the reliability of the information provided • Never release information about the respondent without the express approval of the respondent and his or her parent • 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 • 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

TROUBLESHOOTING: DEVELOPING AND PILOTING A SURVEY INSTRUMENT

- ▶ **Measurementmania:** 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.

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.

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.

- ▶ **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*, 3rd 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 Negdar 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?

Guide on Measuring Decent Jobs for Youth

Monitoring, evaluation and learning in labour market programmes



Note

7

Evidence uptake in policy formulation

A large, intricate, light gray decorative pattern covers the entire page. It features a central vertical axis of symmetry with complex, interlocking floral and geometric motifs, including pointed arches, scrolls, and leaf-like shapes.

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

NOTE 7.
Evidence uptake in policy formulation

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. 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 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 7: Evidence uptake in policy formulation/ International Labour Office. Geneva, 2018.

ISBN : 978-92-2-131668-8 (print)
978-92-2-131669-5 (web pdf)

Also available in Arabic: دليل قياس الوظائف اللائقة للشباب. الرصد والتقييم والتعلم في برامج سوق العمل النشطة - الملاحظة 7: اعتماد الأدلة في صياغة السياسات، ISBN 978-92-2-630799-5 (print), 978-92-2-630800-8 (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.

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

Printed in Switzerland

Contents

Evidence uptake in policy formulation	1
Types of evidence-based research	3
Diagnostic research	4
Descriptive research	4
Causal research	6
The system shaping evidence uptake	8
The supply side: Providers of evidence	8
The exchange function: Communicating research	11
The demand side: The users of evidence	13
Tools to boost evidence uptake in youth employment	18
Conduct a context analysis	18
Build a plan for evidence uptake	20
Synthesize and repackage research	23
Capacity development	24
Conclusion	27
Key points	27
Key resources	28
References	28
Case study: Uptake of evidence on the effects of skills training on young people's financial behaviour and employability in Morocco	31
Tables	
7.1: Types of policy impact objectives to be achieved through impact evaluation	7
7.2: Strategies to increase demand for evidence	17
Figures	
7.1: Supply-and-demand factors that generate evidence-informed policy	9
7.2: Examples of factors which determine level of demand for research evidence	15
7.3: The rapid framework	18
7.4: Examples of stakeholder mapping	22





Evidence uptake in policy formulation



Prerequisites:

This chapter requires no prior knowledge. It introduces readers to tools to maximize the likelihood that evidence-based research will be used to inform the development and implementation of youth employment policies.



Learning objectives:

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

- ▶ understand the different types of evidence-based research: diagnostic, descriptive and causal
- ▶ appreciate the complexity of the system shaping evidence uptake, including supply-side and demand-side factors, as well as the politics and incentives driving evidence generation and use
- ▶ use key tools to communicate and disseminate evidence to policy-makers and synthesize and repackage research for different audiences.



Keywords:

Communications, context analysis, evidence supply, influence mapping, knowledge management, policy briefs, evidence uptake plan, rigour, stakeholder engagement, synthesizing research, working papers.

International development aid donors spend billions on development-related research with the goal of diagnosing, monitoring and evaluating their programming efforts. Yet, how much evidence-based research plays a role in decision-making and policy formulation processes is often questioned. This situation also pertains to the case of youth employment, where policy formulation processes are complex, involve a variety of stakeholders and interest groups, and can become highly politicized.

Evidence is therefore only one piece, albeit an important one, in the policy formulation puzzle. Direct effects of evidence-based research on the policy formulation process are difficult to measure. However, some believe that research affects policy, not so much through immediate and direct impact on the design of public policies, but rather mainly through a process of “gradual sedimentation” of insights, theories, concepts and ways of looking at the world (Weiss, 1977).

The formulation of evidence-informed policies on youth employment is constrained because of the small, albeit growing, base of evidence on “what works”. A recent systematic review, which assessed the breadth and depth of impact evaluations, focused on active labour market programmes (ALMPs) for young people and identified 113 impact evaluation studies on the topic globally. Of these, nearly half were published after 2010, with 21 studies published in 2014 alone (Kluve et al., 2017). As investment in research is becoming increasingly concentrated on improving the evidence base for effective labour market programmes and policies for youth, this note aims to provide guidance for evaluators and researchers to ensure that findings and recommendations are integrated into employment-related policy frameworks, national strategies and other policy formulation processes involving young people.

Types of evidence-based research

The development and delivery of evidence-based research and analysis is an important function of the International Labour Organization (ILO) and its partners for shaping thinking on employment globally and influencing policy-making. Over the past years, the ILO has initiated several major research initiatives aimed at filling knowledge gaps and providing evidence essential for informed policy-making at the country level. Of particular note are a series of reports providing new quantitative estimates of youth employment, child labour, forced labour, domestic workers and migrant workers. These efforts were instrumental in the successful inclusion of decent work objectives into the Sustainable Development Goals, many of which will be monitored and assessed by the ILO going forward.

The ILO's work in evidence-based research can be described in three overarching categories:

DIAGNOSTIC RESEARCH

As seen in Note 1, employment diagnostic analyses provide a means to comprehensively analyse the labour market and employment situations of youth. An employment diagnostic is usually the first step in supporting the development of national employment policies and strategies, serving as a basis for policy dialogue and often leading to policy design.

Diagnostic research: Understanding barriers and opportunities for programming and policy-making. Examples include:

- ▶ employment diagnostic analysis
- ▶ value chain analysis
- ▶ skills forecasting.

Descriptive research: Normative and observational in nature, whose main goal is to monitor outputs and outcomes of programmes. Examples include:

- ▶ performance evaluations
- ▶ observational analysis, qualitative focus.

Causal research: Evaluations that establish causality between interventions and their impact. Examples include:

- ▶ randomized controlled trials (RCTs)
- ▶ quasi-experimental evaluations.

Recent examples of employment diagnostics informing the formulation of national employment policies include the School-to-work transition survey (SWTS) in Samoa (see box 7.1) and the employment diagnostic of Bangladesh (see box 7.2).

Similar examples of the SWTSs contributing to national youth employment policy development

can be observed in Uganda (Ministry of Labour and Training Authorities' National Action Plan for Youth Employment) and in Ukraine

(Ministry of Social Policy of Ukraine and the Institute for Demography and Social Research's new law on "Employment of the Population").¹

DESCRIPTIVE RESEARCH

Descriptive research on the results of individual programmes and projects is generated through monitoring and analysis, as well as programme performance assessments. In many organizations, the performance assessment of projects and policy is the function of an evaluation unit, which is responsible for reporting on results and effectiveness of investments. While the structure of such approaches does not allow researchers to definitively determine the causal linkages between programmes and outcomes, these reviews ensure that programmes are achieving their goals, and in cases where outcomes are not positive, allow readjustment of programme implementation.

Many of the findings of these evaluations and performance monitoring have been made public. The Donor Committee for Enterprise Development (DCED) Standard for Results Measurement requires programmes to publish their monitoring data in the form of an annual report. In the case of the ILO, performance evaluation summaries are made public through the i-Track database (see box 7.3). Over the past decade, this database shows 89 evaluations concentrated on youth employment, an example of a systematic and user-friendly approach to access information about what works in youth employment.

¹ For more information, see [Klein, 2016](#).

Box 7.1: School-to-work transition survey (SWTS) in Samoa

The Samoan SWTS, implemented in 2012, was a joint project of the Samoa Bureau of Statistics and Ministry of Commerce, Industry and Labour and the ILO. The results of the study, published in 2014, reveal a high youth unemployment rate of 16.7 per cent, as well as a high youth labour underutilization rate, which is a measure of the skills mismatch in the labour market, of 52.2 per cent.

Dialogue and knowledge transfer between the ILO and the Samoan Government has had a direct link to policy development on youth employment. The SWTS coincided with a government plan to implement a national youth action plan, the Samoan National Action Plan on Youth Employment (SNAP). SNAP utilized the original country report as well as supporting a re-analysis of the SWTS information to inform the development of the action plan. Subsequently, the SNAP approach was incorporated into the Samoa One United Nations Youth Employment Programme, which is currently being implemented.

CAUSAL RESEARCH

As set out in Note 5, impact evaluations are the major method for generating cause-and-effect knowledge; that is, for determining whether observed changes in the economic

or social well-being of beneficiaries can be attributed to a particular intervention, project or programme.

Box 7.2: Employment diagnostic in Bangladesh

In the Bangladesh national employment policy formulation example, particular attention was paid to the specific population of young people. In addition to youth unemployment, the level of labour underutilization is very high among the youth in Bangladesh. Nearly 38 per cent of the country's youth are neither in the labour force nor in education or training. A further 20 per cent are in irregular employment, while 4.6 per cent were unemployed (ILO, 2014). A revision to the Bangladesh Labour Act was approved in 2013 in the wake of a number of major industrial incidents in the ready-made garment sector. Policy improvements were made in the areas of freedom of association, collective bargaining and safety in the workplace, all areas where young women in factory settings were disproportionately affected.

Despite these improvements, the Government of Bangladesh sought a deeper understanding of the economic situation facing youth. Therefore, at the request of the Ministry of Labour in Bangladesh, the ILO and the Asian Development Bank completed an Employment Diagnostic Analysis in 2015. Of particular importance for the Government of Bangladesh, and the Bangladesh economy as a whole, is the issue of migrant and overseas employment, a major source of employment for the young and growing labour force and an important source of foreign exchange earnings. The diagnostic analysis called for better coordination mechanisms to manage migratory flows, focusing on reducing exploitation of migrant workers and closer collaboration with recruitment agencies. In 2015, the Government of Bangladesh passed the Overseas Employment and Migrants Act, enacting many of the recommendations generated by the employment diagnostic analysis.

Box 7.3: i Track database and i-eval Discovery

All evaluation reports produced by the ILO are systematically scheduled and stored in the i Track database. This includes mandated independent or internal evaluations, as well as joint, external, impact and high-level evaluations that cover the ILO's work. The knowledge generated from these evaluations in terms of lessons learned, emerging good practices and recommendations is also stored in the i Track database and made available through i-eval Discovery. This information is meant to support organizational learning and can be used to inform the design and implementation of ILO programmes and projects.

The purpose of i-eval Discovery is to encourage the use of evaluations. The application visually displays all of the ILO's evaluations, recommendations, lessons learned and good practices through a user-friendly mapping feature. Information can be tailored to meet specific criteria by applying various filters, such as by year, country/region, theme, evaluation type, timing and nature (see <http://www.ilo.org/ievaldiscovery>).

An impact evaluation can take place at different levels, either at the level of an individual intervention, a specific policy area, such as youth employment, or a whole policy regime or system. An impact evaluation can also cut across all three levels, as in the example of conditional cash transfers (CCTs) in Mexico and Brazil (see box 7.4).

While impact evaluations are often criticized for being too academic in nature, there has been a push in recent years for impact

evaluation experts to engage in policy debates and processes more directly, changing their roles from “doing research for development” to “doing research as development”. However, given the technical nature of causal research, it is often difficult for researchers to communicate their findings effectively to policy-makers.

Impact evaluation is a particularly important aspect of policy-oriented research because the nature of its design provides for an

Box 7.4: Impact evaluation of conditional cash transfers (CCT) programmes

Conditional cash transfer programmes have spread rapidly over the past decade in the developing world. CCT programmes provide cash transfers to poor families, which are contingent on children’s educational and health investments, typically school attendance and regular medical check-ups, with the goal of breaking the intergenerational cycle of poverty. As of 2010, all but two countries in Latin America and over 15 countries in Asia and Africa had a CCT programme as part of their social protection systems.

Impact evaluation has long been an integral part of the design of CCT programmes. Much of the popularity of CCT programmes can be attributed to the results of the evaluation, in terms of the culture of policy evaluation, the delivery of social protection and how it affected government regimes directly.

The first generation of CCT programmes and associated impact evaluation studies produced considerable evidence suggesting that these programmes demonstrably helped to lift many families out of poverty and have improved short-term educational, nutritional and health outcomes of millions of children worldwide.

The Nicaraguan CCT programme reduced the fraction of participating households below the poverty line (i.e. the poverty rate) by 5 percentage points after two years, and the Colombian CCT programme reduced the poverty rate by 3 percentage points over four years. The evidence from programmes in Mexico and Honduras, however, suggests no discernible impact on the poverty rate among programme participants.

More recent evidence from a wider array of CCT programmes in Latin America indicates that CCTs may contribute to reducing poverty rates at the national level.

Evidence using household data from 13 Latin American countries suggests that, relative to an internationally comparable poverty line of US\$2.5 per day in 2005 purchasing power parity (PPP) terms, national poverty rates would be 1 to 2 percentage points higher (approximately 13 per cent higher relative to average baseline rates) in the absence of CCTs.

Source: Saavedra and [García, 2012](#); [Saavedra, 2016](#).

accurate assessment or estimation of causality in the context of evaluating a programme or project. Impact evaluations are structured around the careful identification of a counterfactual, a means of estimating what outcomes for beneficiaries would have been, had the programme or project not been available to them.

However, given their ability to attribute causality, or at least provide a targeted estimate of a programme or policy's direct impact on beneficiaries, impact evaluation can be particularly useful in bringing order and rationality to the making of policy. While it is recognized that no single piece of research will result in

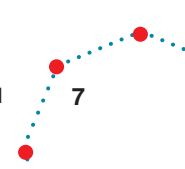
any particular policy change, impact evaluation's ability to determine results and attribute changes to public interventions is unique in (a) helping governments decide whether to continue or terminate particular policy initiatives; (b) expanding and institutionalizing successful programmes and policies and cutting back unsuccessful ones; and (c) determining which programmes to modify and which components of the programme were in need of modification (Weiss, 1999).

Table 7.1 identifies additional types of policy measures that can be achieved through impact evaluation and other types of causal research.

Table 7.1: Types of policy impact objectives to be achieved through impact evaluation

Levels	Dimensions
<ul style="list-style-type: none"> • Project, programme or policy • Policy area • Policy regime or system 	<ul style="list-style-type: none"> • Attitudinal change • Discursive commitments • Procedural change • Policy content • Behavioral change
<p>Outcome dependent on geography (where?) and timing (when?)</p>	

Source: Adapted from Jones and Villar (2008).



The system shaping evidence uptake

While both researchers and policy-makers aim to improve the delivery of programmes and the effectiveness of policy, they are operating on different “sides” of the evidence system. On the supply-side, researchers generate knowledge, and on the demand-side, policy-makers use evidence.

Unfortunately, supply-side and demand-side actors often work without a full understanding of the context in which the other operates. There is often a lack of awareness of the technical nature of research work, on the one hand, and the complexities of policy development and implementation, including the budget cycle, on the other hand. This can often lead to misunderstandings.

Here, we outline the major factors shaping evidence uptake on the two sides of the system. Improving evidence consumption depends on both the supply (by the researcher) and demand (from the policy-maker) coming together in the policy process, as depicted in figure 7.1. We also therefore look at the important function of “exchange” – which determines how well supply-and-demand factors interact.

If constraints to evidence uptake lie mainly in supply or exchange, then approaches to improve research communication and dissemination can be adopted that will help to facilitate the communication of knowledge and enable learning on the policy side. If the problem is on the demand side, then strategies can focus on improving awareness and absorption of research inside government, expanding research management expertise and developing a culture of “policy learning” (Stone, 2009, pp. 303–315).

Many of these strategies focus on higher-level policy development; however, evaluation teams focused on smaller, non-government programmes should also be engaged in efforts to effectively disseminate findings from their programmes. At the same time, policy-makers should seek out lessons from such programmes, as much of the innovation in youth programming and policy is found therein, and initial learning in such programmes can provide important guidance in the context of scaling up similar interventions on a national level.

THE SUPPLY SIDE: PROVIDERS OF EVIDENCE

Suppliers of evidence may include government-oriented evaluation units, although they more often comprise policy-oriented international institutions, such as the ILO, economic development-oriented think tanks, academic institutions and private research firms.

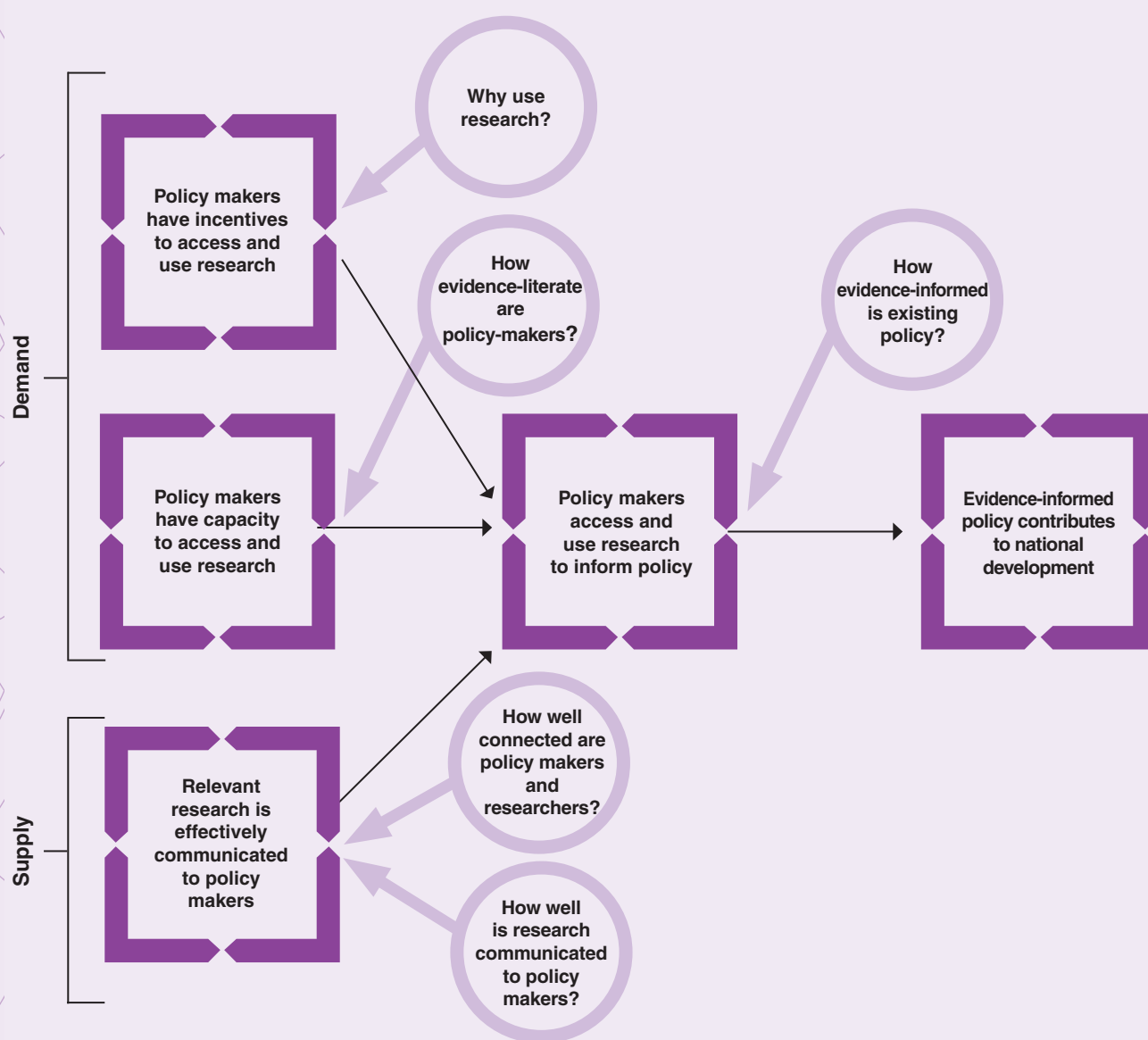
Evidence-based research can also be conducted by individual researchers. Despite the new expectations that urge researchers and evaluators to engage more deeply in knowledge transfer, many still accord it a low priority (Jacobsen et al., 2004).

There are several determinants governing the level and quality of supply of evidence-based research (Taylor, 2016), including the following:

- **Skills available to generate evidence:** This includes the technical capacity of the institution's research team, the institution's involvement in policy processes and debates, and its political weight within the local context. Here, local research institutions may be better positioned to engage

in effective dissemination within the local policy community than external actors. At the same time, impact evaluation requires a high level of specialized skills in econometrics and evaluation design – skills that local research institutions or developing country researchers may not possess. In either case, strategic partnerships can help to balance the need for technical expertise and local engagement.

FIGURE 7.1: SUPPLY-AND-DEMAND FACTORS THAT GENERATE EVIDENCE-INFORMED POLICY



Source: Newman et al., 2013.

► **Finance available to generate evidence:**

Are institutions able and willing to pay what it costs to generate the amount of evidence required? (Taylor, 2016). As outlined in Note 5, causal research can often be expensive, requiring significant resources over a number of years. However, low-cost methods for generating causal, descriptive and diagnostic evidence are available (see Notes 2, 5 and 6), meaning that access to funds is often less of a constraint than people perceive.

► **Information available to generate evidence:**

In developing economies, programmes frequently operate in data-constrained environments. As set out in Note 3, the quality of programme monitoring systems – and quantity of information collected about programme performance – is often a key determinant of whether a quality evaluation can take place. Impact evaluations are sometimes commissioned when success is predicted, since funders are more likely to make money available when they think there is a good chance for positive evaluation results. The supply of evidence is therefore highly constrained by selection bias (Taylor, 2016).

► **The enabling environment for research:**

All countries and the national research associations that operate within their borders, have rules to govern data collection and other research activities. These rules are in place to ensure confidentiality and protect identities of research respondents. These rules will directly or indirectly affect the quality and amount of evidence and data produced.

TIP



Increase evidence uptake: The proximity evaluators have to decision making structures is an important element, which may affect uptake. The location of researchers within a government agency, such as a ministry of labour or ministry of planning, or embedded in a workers' group or employers' association, may help to ensure that evaluation efforts are embedded within operations and that researchers are involved in all levels of programme design and implementation. However, if researchers are not sufficiently independent, it may affect the credibility of findings.

A key supply-side norm is what is considered “acceptable” evidence by those financing its production, rather than those demanding the evidence for use (Taylor, 2016). Heated debates are ongoing between researchers about what constitutes an acceptable level of “rigour” in evidence generation. Yet rigour – the quality of being extremely thorough and careful – is not a binary concept or the domain of one particular methodology alone.

In the youth employment setting, evidence-based research and impact evaluation are relatively new research approaches that have yet to secure much traction in the policy sphere. This is especially the case for RCT impact evaluation methods. As described in Note 5, the RCT approach provides researchers with a unique capacity to deliver a carefully constructed counterfactual to programme participation, which allows for an accurate estimate of causality. Yet, RCT is only one type of evaluation among many trying to establish causality.

THE EXCHANGE FUNCTION: COMMUNICATING RESEARCH

Evidence is only useful where it is usable, and to be usable it has to be communicated in a way that can be understood (Taylor, 2016).

Problems with communicating evidence are sometimes attributed to the “ivory tower” complex, which sees the academic community as operating aloof from the practical, real-world considerations of policy-makers and using a language that is considered unintelligible to the general public. Researchers focus, first and foremost, on delivering comprehensive assessments, and, as such, prioritize investments that ensure the implementation of methods and data collection instruments that maximize their ability to capture results. More broadly, evaluators and evaluation studies themselves are generally assessed on the quality of the study, rather than the dissemination of its results. In particular, academics engaged in impact evaluations are incentivized to undertake research that is publishable in academic journals, which may not align with the information needs of policy-makers or the effective dissemination of learning to a wider audience.

In order to bridge the worlds of evidence suppliers and users more effectively, a number of steps can be taken:

- ▶ **Develop a communication strategy:** Effective learning requires effective communication with stakeholders, based within a communication strategy carefully developed as an integral part of the project’s initial evaluation plan. The effectiveness of the institution’s communications policies and products is key, as are its staff’s communicational skills. Reinforcing the need for local presence, the institution’s networks and ongoing relationships with policy-makers and other stakeholders can also facilitate a greater degree of policy influence.
- ▶ **Align research with policy processes and evidence gaps:** In considering policy influence, researchers need to have a thorough understanding of the policy-making priorities, structures and frameworks in place and how research questions can respond to these. This includes opportunities to answer policy-makers’ questions on “what works” in youth employment. The more closely evaluations are aligned with policy-makers’ needs, the more likely they are to be used in policy-making processes.
- ▶ **Build coalitions:** Researchers and research institutions will be rewarded if they make the effort to better understand the standard processes for programme investment and policy reform within a country at the government level. This includes taking time to identify the key actors (ministers and support staff) and their roles, as well as external actors who can help in forging links with these actors. It also means being aware of the budget cycle and the timetable for its development. Being ready with the right information when ministries are beginning to prepare next year’s budget can guarantee a receptive audience looking for policy and programme solutions. Collaboratively developed research agendas, such as the impact research agenda on youth employment developed by the Taqueem Evaluation Council (see box 7.5), can be used by researchers to guide future evaluations towards the most pressing policy questions.
- ▶ **Get the timing right:** There is a fundamental tension between the time required for substantive research, particularly impact evaluations, and the information needs associated with efficient programme delivery. Striking the right balance in evaluation design – one that allows for a comprehensive assessment of outcomes while ensuring timely inputs for programme rollout or continuation – is difficult but essential.

Box 7.5: Taqueem Evaluation Council: What Works in Youth Employment

The Taqueem (meaning “evaluation” in Arabic) Initiative is a programme of the ILO, supported by Silatech and the International Fund for Agricultural Development (IFAD), whose goal is to increase the effectiveness of youth employment and enterprise interventions in the Middle East and North Africa (MENA) region through improved results measurement and evaluation practices within the community of youth-serving organizations. The Taqueem Initiative has sought to secure this goal through the provision of targeted training opportunities, technical support and small grants to facilitate the implementation of results frameworks by participating organizations.

In 2014, the ILO established the Taqueem Evaluation Council to play a strategic role in the Taqueem Initiative. The Council has a central role in the initiative’s aim of developing capacity to produce and use sound monitoring and evaluation (M&E), impact research and evaluation evidence in the MENA region. This is achieved by involving local and international evaluation institutions and youth employment experts as members in the Council and by creating knowledge-sharing modalities through online, virtual and face-to-face platforms.

The Taqueem Council includes the following institutions: Abdul Latif Jameel Poverty Action Lab (J-PAL), American University in Cairo (AUC), American University of Sharjah (AUS), Centre de Recherche en Économie et Statistique (CREST), International Development Research Centre (IDRC), International Initiative for Impact Evaluation (3ie), Issam Fares Institute for Public Policy and International Affairs (IFI), Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Silatech, Swiss Academy for Development (SAD) and the World Bank.

Source: See <http://www.ilo.org/taqueem>.

Box 7.6: Impact evaluation and youth employment in Tunisia

Tunisia’s National Observatory for Employment and Qualifications (ONEQ), under the Ministry of Vocational Training and Employment (MFPE), demonstrates a strong commitment to impact evaluation principles. Two major impact evaluations on youth employment have been championed by the Ministry. The first is a quasi-experimental approach assessing the impact of employment subsidies: “Tackling graduate unemployment through employment subsidies: An assessment of the SIVP programme in Tunisia”.

The second is an RCT focused on entrepreneurship, “Entrepreneurship and self-employment among university graduates: Evidence from a randomized trial in Tunis”. The positive experiences and opportunities for evidence creation developed under the two evaluations led MFPE to sign a memorandum of understanding with the International Initiative on Impact Evaluation.

For more information, see [Broecke, 2012](#) and [Premand et al., 2012](#).

THE DEMAND SIDE: THE USERS OF EVIDENCE

Demand for evidence is not limited to government policy-makers. Users may include other ILO constituents (workers' groups and employers' groups), as well as parastatal organizations, think tanks, non-governmental organizations (NGOs), youth associations and private sector firms. In some country contexts where data and research is constrained, officials tend to depend heavily on the evidence put forward by multilateral and bilateral donors to inform national policy processes.

Demand for evidence is shaped by a complex array of economic, cultural, political and historical factors. Policy-makers tend to be heavily influenced by their own values, experience, expertise and judgement, the influence of lobbyists and pressure groups, and pragmatism based on the amount of resources they have available. The most important determinants for the level of demand for evidence based research are listed below and then further elaborated in figure 7.2:

- ▶ **The institutional setting:** Governmental systems should have effective processes and rules of operation to ensure that evaluation evidence is part of the policy-making process. This requires ministries, agencies and public employees to be held to account for ensuring that programmes and intervention investments are aligned with documented evidence. To achieve this end, effective evidence-based programming is, more often than not, correlated

with democratic governance: the more democratic the system of government, the more open it will be to rigorous assessment, learning from mistakes and evidence of effectiveness. On the basis of its accountability to citizen interests and taxpayers, a democratic system creates space for unbiased, independent research. For non-democratic governments, ensuring this accountability is more difficult and depends largely on signalling from the top regarding the importance of evidence-based programming and policy development and the transparency of its implementation.

- ▶ **Cultural norms:** There are fundamental differences in the policy-making process between different countries, borne out by cultural and historical factors. Some prefer arguments based solely on theory and do not place much value on empirical evidence, while others are more data-driven. Even in countries that have invested heavily in evidence-based policy, such as the United Kingdom (see box 7.8), demand is far from homogenous among either ministries or policy-makers. A former chief economist for the UK Department for International Development, for example, says that their policy decisions are made on the basis of a compelling case based primarily on theory, since any evidence base, no matter how rigorous, will always be incomplete (quoted in [ILO, 2015](#)).

- ▶ **Knowledge management:** One important area for facilitating learning across an organization is having well-functioning processes for “externalizing” the tacit knowledge generated in the process of working, collecting emerging insights and turning them into something explicit and easy to share (see an example in box 7.8). This includes having systems in place (usually digital in nature) that are designed to store this and other policy-relevant knowledge so that staff around the government or policy organizations can access the knowledge and use it as and when needed.
- ▶ **Skills required to utilize evidence:** It is essential that policy-makers and support staff have the capacity to read, interpret and apply evidence in policy-making. In terms of youth employment, this includes a prior knowledge of and specialization in youth employment topics and a familiarity with impact evaluation. At the individual level, individuals need to have a broad range of capacities including: knowledge

Box 7.7: UK’s Department for Work and Pensions

The United Kingdom’s Department for Work and Pensions (DWP) is a leader in evaluating national active and passive labour market policies and committing to evidence-based policy making. This is illustrated by its historic investment in externally commissioned research, its internal teams of analysts, its commitment to publishing all evaluation results, and its active use of research and evaluation findings to inform policy development and review.

Recently, the UK’s public expenditure environment has constrained resources for policy delivery and evaluation, meaning that the DWP must ensure its investments provide best value for money. These financial pressures place an increased emphasis on evaluating what works, and at what cost, and ensuring that investments in evaluation answer the key questions they were designed to address.

The vast majority of the DWP’s evaluation activities are delivered by external contractors, and the Department commits a high level of funding and staff time to supporting its commitment to evaluation. A wide range of methodologies are employed in DWP evaluations, ranging from qualitative interviews to the use of more sophisticated quasi-experimental methods and randomized control trials to identify gross and net impacts.

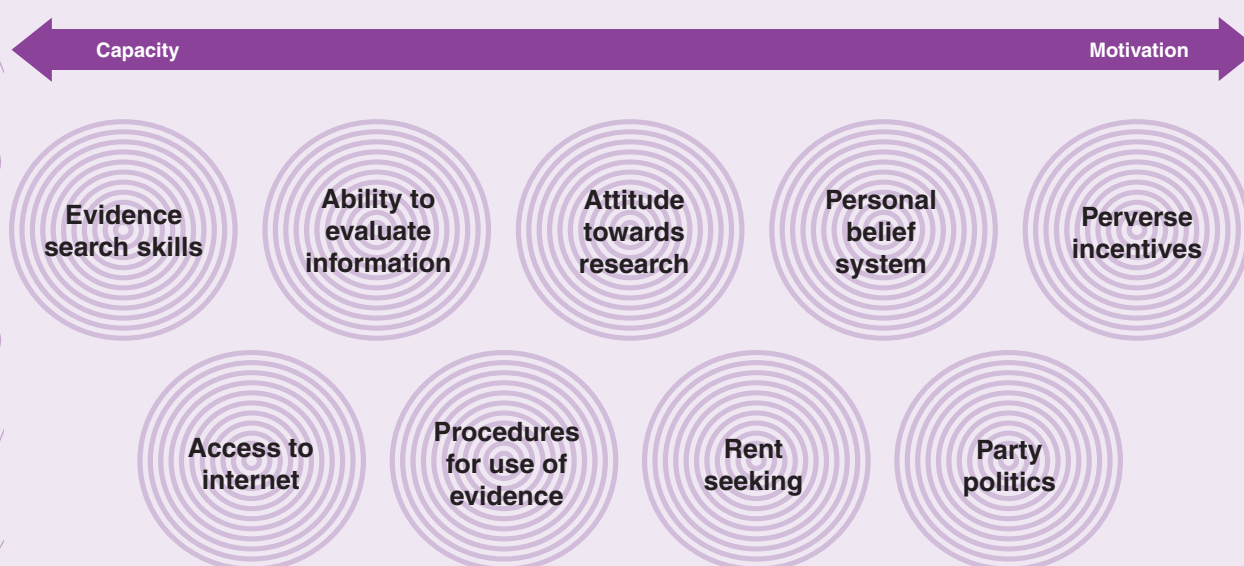
In all cases, policy implications of evaluation reports are summarized for the minister, as well as for steering meetings and project and programme management boards attended by key decision-makers. All evaluations are made available to the public, including those with negative results.

A good example of this is the “Job Retention and Rehabilitation Pilot”, a programme which initiated interventions with individuals on sick leave for between six and 26 weeks to support the return to work, with a view to rolling out the approach nationally if proven to be effective. Initial take-up was slow, although it picked up later in the piloting period, and the evaluation, which featured a randomized control trial, showed that the pilot was having no impact and so was not continued.

about what research is and how it can be used; skills in searching for and evaluating research information; critical thinking

skills to absorb, critique and amalgamate information; and a positive attitude towards research evidence.

FIGURE 7.2: EXAMPLES OF FACTORS WHICH DETERMINE LEVEL OF DEMAND FOR RESEARCH EVIDENCE



Box 7.8: Global Public–Private Knowledge Sharing Platform on Skills for Employment

The Global Public–Private Knowledge Sharing Platform on Skills for Employment (Global KSP) aims to help strengthen the links between education and training to boost the creation of productive and decent work by sharing evidence, evaluation, approaches, knowledge and experiences that governments, employers, workers and international organizations have found effective in addressing these issues of common concern across the world.

The Global KSP uses the G20 Training Strategy as its foundation and builds on it by providing evidence on how training and skills strategies, and policies and systems work, with their related requirements for resources and engagement by stakeholders and in combination with other policies and institutions. The Platform enables the exchange of ideas and experiences among policy-makers, the private sector, technical and vocational education and training (TVET) institutions, academic institutions, bilateral agencies and other international organizations that produce evidence on skills development to improve employability and productivity.

Source: See <http://www.skillsforemployment.org>

Box 7.9: Evidence-based policy-making in Rwanda

Rwanda is making tremendous strides in terms of resolving long-standing issues of poverty, health and employment. In the context of post-conflict reconstruction and redevelopment, the Rwandan Government's efforts to improve development outcomes for the country's people have garnered the support of international agencies, including the UN and the World Bank, for its efforts to increase access to housing, healthcare and work, particularly for women, and for meeting the Sustainable Development Goals laid out by the country.

Part and parcel of this success has been a commitment to evidence-based policy-making. In contrast to many low-income countries, Rwanda has taken steps to ensure that major investments in development programmes are built on solid evidence and studies that ensure learning. In 2011, Rwanda signed a programme of support with the UN, which allocated nearly US\$2 million to strengthening monitoring and building local M&E capacities. In 2015, the United Nations Population Fund highlighted the role that evidence-based data has played in facilitating development, social planning and the allocation of resources.

Rwanda's commitment to impact evaluation is particularly noteworthy. Working with international donors, NGOs and research institutions, Rwanda has supported a large number of experimental and quasi-experimental studies in recent years. Examples are listed below:

- **Young people, jobs and agricultural cooperatives:** Using a quasi-experimental design, the ILO evaluated the effects of an intervention to enhance cooperative managers' business skills, cooperatives' overall competitiveness and cooperative members' income and employment conditions, as well as to improve food security in Rwanda. The research explored the short-term impact of the intervention on the employment, organizational, marketing and financial outcomes of agriculture cooperatives (ILO, 2017a).
- **Literacy boost in Rwanda:** This two-year RCT finalized in 2016, was a partnership between the Rwanda Education Board, Save the Children and Stanford University, assessed the impact of Literacy Boost's community-based learning against school-based learning in raising literacy rates among children (Friedlander et al, 2016).
- **Life skills and work readiness:** Over the period 2013–2014, Education Development Center, Inc. undertook an RCT of the Akazi Kanoze programme, which provides youth in Rwanda with job-relevant life skills and work-readiness training, as well as links to employment and self-employment opportunities. Despite an initial decline in employment in both the treatment and control group, a higher percentage of youth in the treatment group were employed after the end of the Akazi Kanoze programme (Alcid, 2014).
- **Promoting agricultural technology adoption:** J-PAL and TechnoServe, an agri-business NGO, evaluated the impact of agricultural business and technology training in coffee-growing regions in Rwanda using an RCT. The study compared villages that received no training with villages that received low-, medium- and high-density training. Preliminary data suggests that the training helped farmers to improve their growing practices, but that the farmers are more likely to adopt those technologies and practices that require the least effort to apply (Pamuk et al, 2014).
- **Promoting father's education:** Promundo and the Rwanda Men's Resource Center (RWAMREC) launched an RCT in 2015, in collaboration with the Rwandan Ministry of Health, to evaluate the impact of fathers' group education. The study will assess the impact of fathers' group education on family planning, maternal and child health, gender attitudes, violence, risky behaviours and men's health (Doyle et al, 2014).
- **Teacher training and entrepreneurship education:** J-PAL is undertaking an RCT of curriculum reform to promote entrepreneurship education in Rwanda. This study examines the effect of a teacher-training programme on student academic, economic and labour market outcomes, on the understanding that entrepreneurship training will only be successful if teachers deliver the material effectively (forthcoming).

Where demand for evidence is deemed to be low, there are a number of strategies that can be deployed to boost prospects for uptake

(adapted from [Dhaliwal and Tulloch, 2012](#) in table 7.2).

Table 7.2: Strategies to increase demand for evidence

Problem	Potential solution
Political considerations override evidence-generated data	Target policy-makers who are open to evidence, so they use it as an input along with other factors, like political agenda, budget constraints and administrative capacity
Low policy-maker capacity to utilize, generate or institutionalize evidence	Train staff at implementing organizations, establish M&E divisions, recruit competent people and motivate them via formal linkages with academics
Short-term horizon of policy-makers	Combine short-term outcome measures with long-term outcomes and encourage phased roll-outs to offer the opportunity to evaluate programmes before major scale-up
Risk-aversion and failure-avoidance inclinations on the part of policy-makers	Set up institutions that allow innovation, space for “safe to fail” programmes and encourage a higher level of risk tolerance
Lack of pressure from civil society or legislature to conduct evaluations	Convince these institutions to demand evaluations via participation in civil society debate

Box 7.10: Evaluation clinics and communities of practice

The ILO provides technical assistance and training to partners with the objective of creating a critical mass of evaluation expertise among the social partners and country counterparts. Evaluation clinics are workshops on M&E and impact evaluation, designed to share evaluation tools and techniques, to help youth employment practitioners understand, interpret and translate evaluation evidence. The ILO offers both basic and advanced evaluation training.

Evaluation clinics are held over three to four days and cover the basics of M&E, impact evaluation methods and data collection tools. During the clinics, selected youth employment project teams serve as live case studies, whose options for stronger M&E or impact evaluation are discussed and outlined during group work sessions.

Executive evaluation courses are advanced five-day evaluation training courses which provide academics, evaluators, implementers and researchers with a thorough understanding of rigorous impact evaluation techniques in order to strengthen their capacity to understand, interpret and conduct impact evaluations.

The ILO offers technical assistance through a community-of-practice approach. The community of practice in youth employment is a collaborative approach to building capacity on M&E and impact evaluation and to foster learning and cooperation among youth-serving organizations.

Tools to boost evidence uptake in youth employment

In order to ensure that the evidence base continues to grow and is used effectively, researchers and policy-makers must proactively

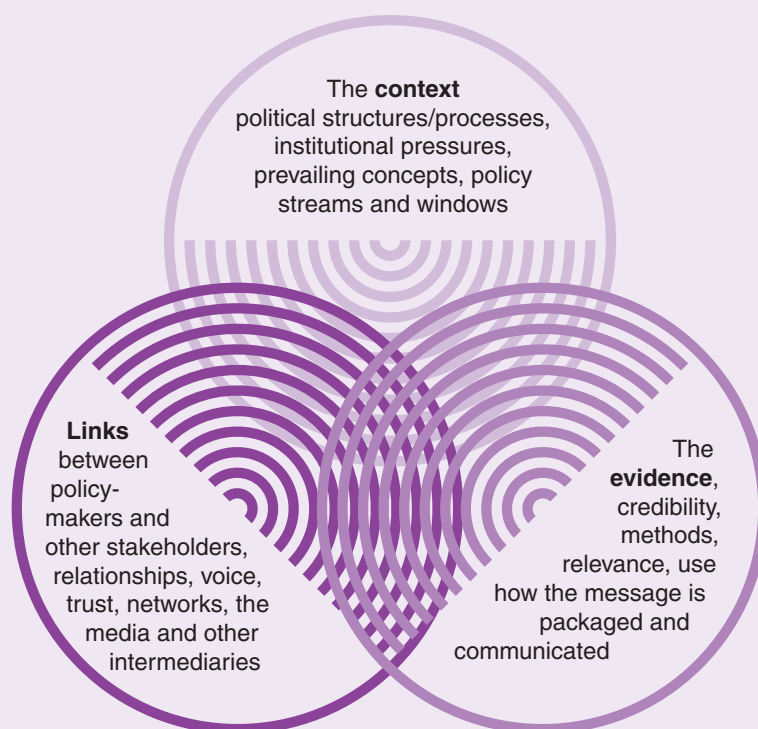
plan ways to improve research uptake. A number of tools are available to support this.

CONDUCT A CONTEXT ANALYSIS

To facilitate the use of evidence based research for promoting a youth employment-focused policy or practice, the first step is often to map the policy context surrounding that issue and identify the key factors that may influence the policy process. The RAPID Context, Evidence and Links Framework was developed by the

Overseas Development Institute as a conceptual framework to help researchers and policy entrepreneurs understand the role that evidence-based research plays, among other issues, in influencing policy (see figure 7.3). The four components of the framework can provide valuable, in-depth information regarding

FIGURE 7.3: THE RAPID FRAMEWORK



policy windows, key policy actors and networks, gaps in the existing evidence, alternative means of communication, and trends and changes in the external environment.

RAPID has developed a simple checklist of questions to accomplish this, including questions about the key external agents, the political context itself, available research-based evidence and the identification of other stakeholders who can help. This can also help you to identify where knowledge and evidence gaps exist and the strategic entry points for policy change.

Within the RAPID framework, the political economy context for evidence production is an important determinant of whether or not the evidence is likely to be taken up and used. Usually, the main driver is the political context, which includes the political system – individual actors and institutions – and the power dynamics both among and within the institutions and actors. This includes, for example, whether the society under review is a relatively open and democratic or a closed and autocratic one, or if it is a fragile and

conflict-affected society where political institutions are weak or even non-functioning. These aspects are likely to affect how knowledge circulates and how decisions are taken.

It is also important to consider which elements of knowledge, debate and decision-making are public and how many are not public. Do local communities have access to and any influence over formal power, and if so, how? As part of this analysis, one must examine the context for decision-making within social and political structures, including the role of actors within local communities, and how those determinants are likely to influence ownership, involvement and uptake, and the use of study findings in policy development.

While the political economy context of labour market interventions has always had an important influence on their development, formal context analyses are increasingly being used in youth employment as a key tool to contribute to the advancement of national employment policy and strategies. This is exemplified by the recent deployment of a context analysis in Egypt (see box 7.11).

Box 7.11: Towards evidence-based active labour market programmes in Egypt: Challenges and way forward

In Egypt, the ILO, the Population Council, the American University of Cairo, the Government of Egypt and development partners engaged in a context analysis process to advocate for evidence-based active labour market programmes (ALMPs) for young people. The objective of the exercise was to analyse the youth employment policy framework and engage key stakeholders, as well as to synthesize global evidence from impact evaluations on the effectiveness of youth employment programmes and compare the result to the situation in Egypt. The process and findings are documented in an ILO impact report presented to the Minister of Manpower and Migration and are expected to contribute to the development of a modern set of ALMPs for young people.

Source: ILO, 2017.

BUILD A PLAN FOR EVIDENCE UPTAKE

A youth employment evidence uptake plan outlines the steps for effective policy influence through the envisioned roadmap of actions, products and activities. These elements are structured in a way that leads to the ultimate goal of informing decision-making in the youth employment field. A youth employment evidence uptake plan will help practitioners to think through strategies to encourage youth employment evidence ownership by key stakeholders from the outset, strengthen the demand for information about progress and results, and help to increase the likelihood that findings will be known, understood and used to improve policy and programming.

Objectives

In setting objectives for supporting the uptake of evidence in youth employment policy formulation, it is important to be realistic about the extent of expected achievements. A good evidence uptake objective should be clear about why the changes being proposed are important, who they will affect, what needs to be done to secure these changes, and where the influencer stands in relation to others who are also trying to bring about change. Main objectives for evidence uptake should be determined in light of evidence that evaluators expect to produce and the context in which these results will be communicated. Once objectives have been set, it is sensible to focus on the more immediate objectives and intermediate outcomes that are produced by the strategies and interventions chosen. For the sake of clarity and focus, the number of objectives should be limited to not more than three.

Stakeholder engagement

At an early stage of the research, it is important to map out who the relevant stakeholders

are likely to be. Stakeholders are those with a clear interest in the outcomes of the programme or project being studied: as such, they may include policy-makers, civil society organizations, the private sector, other researchers and potential beneficiaries. It is important to recognize that employers' and workers' organizations represent formal partners in any efforts to promote better labour market outcomes, although these organizations' input is often overlooked in strategies that focus on engaging government actors. Importantly, the stakeholders to be engaged may be direct users of the research or those who can support your organization as you plan for uptake. Such "evidence intermediaries" can play an important role in communicating results upwards.

Once you have identified the relevant stakeholders, one should consider an appropriate strategy for the engagement of each. This strategy should be developed on the basis of considering what their (potential) interest is in the intervention at hand and the extent and type of engagement needed to support uptake. Each stakeholder will have different interests and perspectives, informational needs and process for dissemination of results based on those perspectives and needs. Taking time to carefully plan methods of targeting specific stakeholders will ensure the effectiveness of later outreach efforts.

If you intend to target policy-makers, you need to understand the policy-making context and norms in the country or countries in which you are working. This includes having a clear understanding of the basics of the local political system and policy-making structures. For example, one should understand the specific roles of members of parliament versus ministries and members of the government, and in this context how laws are made and programmes developed. Within government,

actors in the civil service may play different roles in programme development for the ministers. Naturally, such considerations vary widely within different governments around the globe. Beyond this, understanding the budget calendar and various steps within the formulation of the budget will help to define entry points for those seeking to provide policy-makers with evidence that can be effectively translated to action.

Once you have a basic understanding of the policy process, it is important to find out how policy on your topic of interest is made in your particular country and what relevant policy processes are ongoing. For example, you may find that there is a team within a particular ministry responsible for youth employment or that there is a parastatal organization which deals with this aspect of policy development or that responsibility for your topic is devolved to local government bodies.

At the same time, when considering specific stakeholders, one should not focus solely on the politicians themselves. Technical advisors and staff (e.g. parliamentary staff or civil servants) play an important role in guiding policy decisions and can be a good source of information about both formal and informal policy-making processes.

Influence mapping

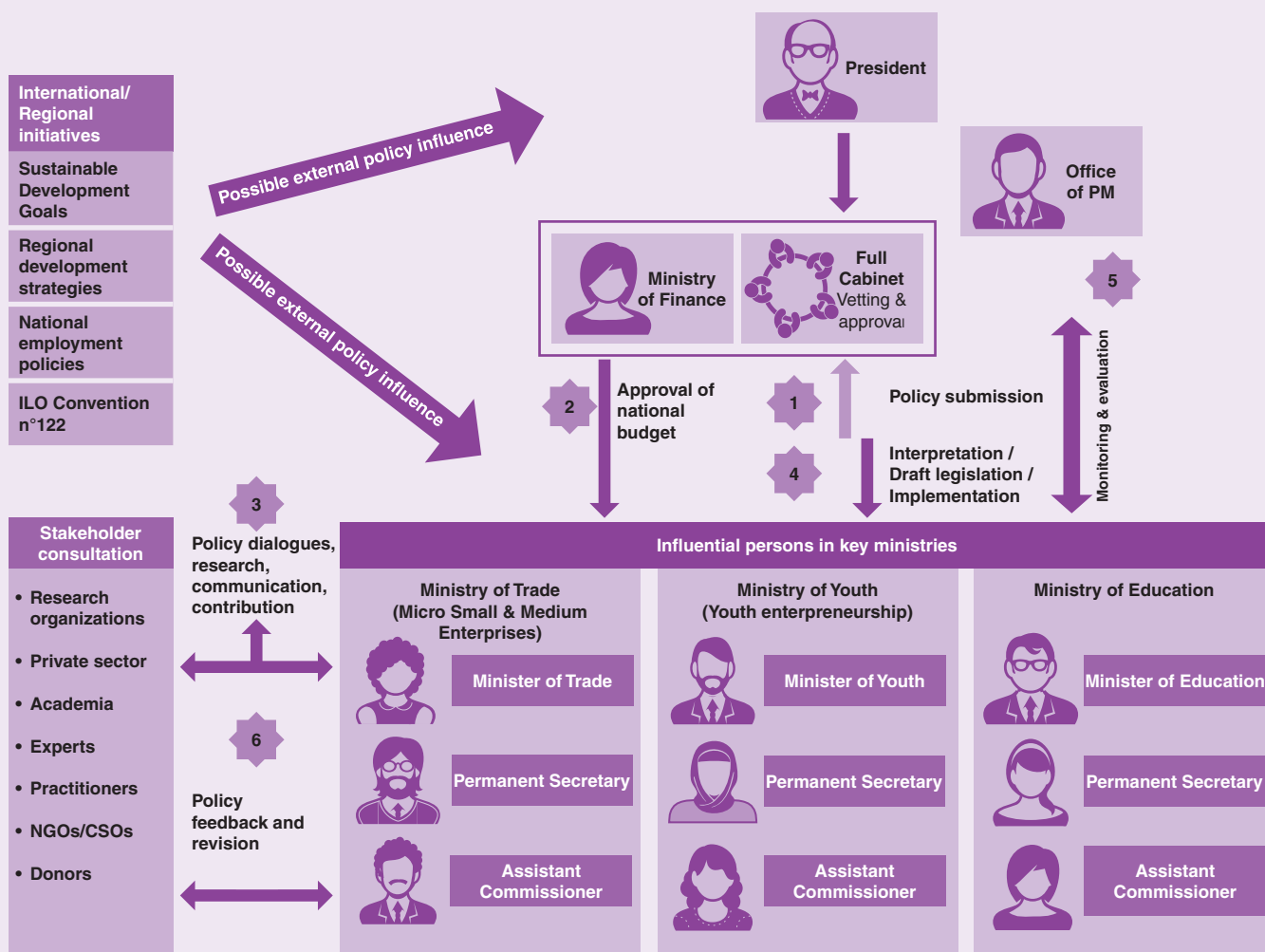
Influence mapping identifies the individuals and groups with the power to affect key decisions relevant to a particular programme or policy approach (see figure 7.4). Beyond the initial stakeholder listing, influence mapping further investigates the position and motives of each player and the best channels through which to communicate with them. The approach is also known as stakeholder influence mapping, power mapping or the arena of influence.

Continued engagement

Once research work has begun, it can be easy to forget about stakeholders, beyond those directly affiliated with the specific project or programme under consideration, until researchers are ready to communicate findings to them. Ideally, one should maintain engagement with stakeholders throughout the programme implementation and the research study. This allows them to continue to advise the research team on research implementation and keeps the research in their minds, making them more likely to pay attention to the final results. If the findings are challenging (e.g. concerning policy ineffectiveness), having existing relationships with decision-makers is likely to enable more effective discussion about the findings and, in turn, policy learning. One way of keeping decision-makers involved is to invite some of them to sit on a steering committee that meets occasionally to provide guidance on emerging issues.

Once results start to emerge, it is important that you find ways to facilitate not just results dissemination but discussion and feedback. This can be done online (using email lists or discussion fora); however, face-to-face discussions are generally more effective. It is important to go to the decision-makers rather than expecting them to come to you. If you are thinking of holding a meeting to present results, consider whether you might get a higher attendance level by holding the meeting in their “space”. For example, you could consider offering to visit a government body and give a briefing to key officials or to visit parliament to talk to members of a relevant parliamentary committee.

FIGURE 7.4: EXAMPLE OF STAKEHOLDER MAPPING



SYNTHESIZE AND REPACKAGE RESEARCH

Research results should be published in formats that are accessible to non-experts and in a style that may be more appropriate for decision-makers than peer-reviewed journal articles. This may include producing research summaries or other written outputs, such as factsheets or writing about the findings in a blog. In the past, there has been a tendency to think that research communication is all about policy briefs. These are not always appropriate for the research in question. It is important to remember that written communications, particularly for your primary stakeholders, are not an end in themselves and should be used alongside other influencing and engagement activities.

Research programmes may choose to share their findings via oral presentations at conferences and meetings. For direct policy dialogue, engagement in short, face-to-face meetings with policy-makers and stakeholders may be most effective. However, as noted above, the effort to influence policy is often a longer-term endeavour to gradually shape perspectives on policy. In this regard, active participation in conferences provides a way to tap into a wider set of stakeholders and

TIP



Writing effective briefs

- Ensure that your research findings are given in the context of the available evidence on the subject.
- Make sure you clearly outline why the research you are presenting is of relevance to policy and what the implications of your findings are.
- Make the brief attractive: policy-makers are more likely to read something visually appealing.
- Summarize the key points and put them on the first page as a clear bulleted list.
- Keep it short – ideally two to four pages.
- Spell out any acronyms and avoid technical jargon (or clearly explain it).

influence policy take-up by a less direct route. In the youth employment space, there are a number of regular conferences that bring together a diverse group of stakeholders and focus on evidence and lessons learned in this space.

Using the media

The media, including television, traditional print media and new media, offer a means not only to engage with policy-makers and

Box 7.12: Policy briefs and working papers

Policy briefs: Policy briefs help to communicate results to internal and external stakeholders. A policy brief presents the core findings of the evaluation in a plainly written format that includes visual material (e.g. graphs and charts) and that makes programmatic and policy recommendations.

Working papers: Researchers can work with the programme team to write working papers and articles for publication in academic journals and to present research findings at universities and research institutions. Working papers can then be published and disseminated through the academic associations to which the investigators belong. Being cited in academic papers is a useful way to increase the visibility of the programme and to create interest among donors.

TIP

Look out for these regularly held conferences that share youth employment evidence:

- 3ie's Annual Evidence Week
- IZA Institute of Labour Economics Conference and Seminars
- Economic Research Forum Annual Conference
- Making Cents Youth Economic Opportunities Conference
- ILO Evidence Symposium on Youth Employment

key stakeholders, but to shift popular perceptions on issues relating to youth employment. However, by and large, researchers and policy-oriented research institutions make little use of the media in their efforts to effect policy change. Journalists are keen to develop stories of interest to viewers and readers, and youth employment is an obvious issue to cover in this regard. The challenge for researchers is to make journalists aware of their work and to present it in a way that provides the journalist with a clear vision of how he or she might report on it in an interesting and intriguing manner. The checklist below is useful when reflecting on how to be camera ready (or newspaper ready) and how to build up a rapid response team within your organization which can respond to relevant media opportunities.

CAPACITY DEVELOPMENT

The requisite internal capacity for research uptake includes the knowledge, skills and attitudes necessary to access, use, create and communicate research information. Policy-makers and their support teams may lack the full range of knowledge and skills needed to assess research findings and move them towards effective implementation on a policy level. In developing nations, the capacity constraints are particularly noteworthy, given the

TIP

Synthesizing research

The most rigorous approach to synthesizing evidence is a systematic review. However, these can take a long time to produce and are not always the most appropriate synthesis method. It is important to select a synthesis method which is appropriate to your specific research. In particular, it is vital to:

- be clear about the methodology you use to search for and select literature for inclusion. This may include mentioning the databases you searched, along with the search string(s) used. You may also choose to carry out hand searching, “snowballing” (i.e. searching the citation lists of other references), applying personal knowledge and/or expert recommendations. For a systematic review the search approach needs to be agreed at the outset
- be explicit about how you will appraise research and make sure you discuss not only the quantity but also the quality of the research evidence
- ensure that you write a clear overview of the synthesis, drawing out the key messages for policy-makers and practitioners.

resource constraints faced by governments in these countries. Making an effective move from practice to policy that builds on this evidence often requires significant investment to support the building of capacity among key decision-makers and their staff. In developing a policy-influence strategy, it is therefore essential to have a capacity-building strategy founded on clear identification of gaps in capacity.

TIP



Becoming a resource for journalists

- Be available. Give reporters, especially those at news services where they work odd hours, your home and mobile numbers and tell them that you are always available.
- Seek out journalists at meetings and public events and give them your business card.
- Be ready to be quoted. Forcing a reporter to call back for authorization to use a quote will reduce the chances of the quote being used. This may necessitate loosening institutional rules regarding media engagement.
- Know the issues. Read and comment intelligently on developments relating to your cause.
- Do not always assume that journalists have received the same information that you have about topical events or relevant news releases.
- Avoid rhetoric and ideological arguments; most journalists have heard all this before.
- Know your facts; never pass on information unless you know it is true.
- Know where to find information or contacts fast and therefore gain a reputation as a good source.

TIP



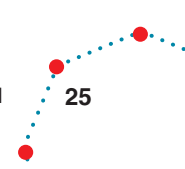
Making a report or press release more newsworthy

- Develop a short (3–10-page) executive summary.
- Put the summary on your website and include the link in any press release.
- In a press release, cover just a few main facts or statistics, most news stories are succinct.
- Use clear graphs and tables, and provide short, simply written paragraphs.
- Connect the report to a news hook to secure journalist interest.
- If possible, show a change in data from the previous year.
- Create quirky, interesting titles for trends of findings.
- If affiliated with an academic institution, issue releases with their letterhead and use their media office for press contacts.
- Make numbers more meaningful by drawing comparisons or breaking them down into familiar units.
- Inquire into publishing a short summary of the report as a guest opinion editorial for a newspaper.

For the ILO, capacity development is geared towards promoting the institutional capacity of member States, as well as representative organizations of employers and workers, to facilitate meaningful and coherent employment policy and sustainable development. The International Training Centre of the ILO in Turin, Italy plays a significant role in providing training for capacity development. The ILO is engaged in capacity development for governments in developing, implementing and evaluating national employment and labour policies targeting youth, but also in developing national skills and training schemes, employment services, labour market information

and statistical services, and social security systems.

Capacity development activities in support of employers' organizations on evidence-based research into youth employment have been limited. Building the capacity of employers' organizations in this area is essential to strengthen their policy influence, lobbying efforts and advocacy for particular approaches. Especially relevant is the growing evidence base for "what works" in youth entrepreneurship, on-the-job training and skills development.



The need is the same for workers' organizations. Workers' organizations represent the interest of employees, following a rights-based approach focusing on the sectoral and workplace level. As democratic organizations, accountable to their membership base, effective monitoring and evaluation is a priority for capacity building, but it largely lies outside traditional perceptions of their remit. To enable workers' organizations to participate more effectively in policy formulation processes on youth employment we need to (i) increase the level of knowledge within workers' organizations on the evidence base for effective youth employment strategies (ii) provide more readily available and accurate labour market information and diagnostics and (iii) facilitate assessment of existing youth skills and knowledge.

Some key areas for capacity development in evidence-based research include:

- ▶ information literacy
- ▶ basic to intermediate familiarity with research methodologies (see Note 5)
- ▶ internal communications
- ▶ internal knowledge management
- ▶ academic writing and summarizing skills
- ▶ skills in finding and appraising evidence-based research on youth employment
- ▶ thematic topic knowledge: skills development, public employment services, wage subsidies
- ▶ incentives (or disincentives) to consider evidence.

Substantial improvement in the use of research-based evidence in development policy and practice also requires effort at the partner level. The aim is to improve constituent structures, processes, resources, management and governance. At the system level, efforts should be made to improve national and regional innovation environments.

There are many approaches to achieve this improvement, including:

- ▶ establishing research partnerships between Northern and Southern research institutions/universities, as well as fostering cooperation between research institutions within different developing regions (South–South cooperation)
- ▶ providing institutional support for universities in developing countries (particularly in sub-Saharan Africa)
- ▶ providing support for national research councils
- ▶ arranging funding for developing country institutions to access the research and technical services of developing country partners
- ▶ supporting communities of practice among researchers and policy-makers working on a specific development problem or sector
- ▶ supporting policy-makers in efforts to become more aware of research-based evidence and more discerning consumers of it
- ▶ sponsoring collaborative regional master's and PhD programmes.

Naturally, for smaller research organizations and for organizations running evaluation studies, the capacity-building support needed to effectively engage policy-makers and other relevant stakeholders can be restricted by resources. With this in mind, such organizations are encouraged to tailor their communications efforts with policy-makers, describing results, methodological approaches and policy implications in clear, simple terms and making staff available to answer any technical questions that policy-makers may have. Moreover, such organizations are encouraged to reach out to larger organizations, particularly international organizations, such as the ILO and the World Bank, as partners in capacity-building efforts to support evidence-based policy engagement.

CONCLUSION

This note encourages researchers and policy specialists to diagnose, plan for and build capacity on evidence-based research on youth employment and resultant learning and knowledge development. The strategies described include employment context analyses, evidence uptake plans, communication strategies and training, with a focus on providing actors on both sides of the relationship with practical advice on effectively engaging with the other and ensuring that all research generated has positive policy implications.

Finally, researchers are advised to use synthesis products, such as systematic reviews and

rigorous literature reviews, to understand how their research fits into the existing knowledge base on youth employment. Synthesis products enable programmes to identify research questions which have not yet been answered adequately. This synthesis will help researchers not only to shape research in a way that productively fills knowledge gaps but to communicate more effectively with policy-makers and practitioners in the context of the wider body of evidence and demonstrate how the work currently being undertaken can improve their ability to shape effective youth employment policy and programmes.



KEY POINTS

1. **The evidence base supporting the design, development and implementation of effective labour market policies for young people is growing** and is likely to continue to expand in the future. Better research and better understanding “what works” in getting young people into jobs can come about by combining the following three types of evidence. Diagnostic research understands barriers to and opportunities for programming and policy-making. Descriptive research observes programme outputs and outcomes, while causal research allows to identify causal links between interventions and their impact.
2. **Improving evidence consumption depends on both the evidence supplied by researchers and evidence demand on the part of policy-makers.** Both supply and demand factors need to come together to inform the policy process. Critical factors determining the quality and quantity of supply include the skills, finance and information required to generate evidence, as well as views on what is considered “acceptable” evidence. Demand is influenced by the institutional environment, cultural norms and systems for knowledge management. The exchange of evidence is about how effectively research is communicated to bridge the two “worlds”.
3. If constraints to evidence uptake lie mainly on the supply side, then approaches to improve research communication and dissemination can be adopted to help facilitate the communication of knowledge and enable learning on the policy side. If the problem is on the demand side, then strategies can focus on improving awareness and absorption of research amongst policy-makers, expanding research management expertise and developing a culture of “policy learning”.
4. **In order to ensure that the evidence base continues to grow and is used effectively, researchers and policy-makers must deploy proactive strategies,** such as conducting a context analysis, developing an evidence uptake plan and synthesizing, packaging and presenting research findings and continuing to develop capacities.

KEY RESOURCES



- Crewe, E.; Young, M.J. 2002. *Bridging research and policy: context, evidence and links* (London, Overseas Development Institute).



- Dhaliwal, I.; Tulloch, C. 2012. "From research to policy: Using evidence from impact evaluations to inform development policy", *Journal of Development Effectiveness*, Vol. 4, No. 4, pp. 515–536.



- Byrne, J.; Albright, K.; Kardefelt-Winther, D. 2016. *Using research findings for policy-making* (London, Global Kids Online).



- Taylor, B. 2016. *The system of evidence*. (Springfield Center).



- CLEAR-AA. 2013. *Demand for and supply of evaluations in selected sub-Saharan African countries* (Johannesburg, University of the Witwatersrand).

REFERENCES

- Alcid, A., 2014. *A randomized controlled trial of Akazi Kanoze youth in rural Rwanda*. Report submitted to USAID. Waltham, MA: Education Development Center.
- Bausch, J.; Dyer, P.; Gardiner, P.; Kluve, J.; Kovacevic, S. 2017. *The impact of skills training on the financial behaviour, employability and educational choices of rural young people: Findings from a Randomized Controlled Trial in Morocco*, (Geneva, Employment Policy Department).
- Broecke, S. 2012. *Tackling graduate unemployment through employment subsidies: An assessment of the SIVP programme in Tunisia*, Working Paper Series No. 158 (Tunis, African Development Bank).
- Court, J.; Hovland, I.; Young, J. (eds). 2005. "Research and policy in international development: introduction", in *Bridging research and policy in development: Evidence and the change process* (London, ITDG Publishing and the Overseas Development Institute), pp. 3–23.
- Dhaliwal, I.; Tulloch, C. 2012. "From research to policy: Using evidence from impact evaluations to inform development policy", *Journal of Development Effectiveness*, Vol. 4, No. 4, pp. 515–536.

- ▶ Doyle, K., Kato-Wallace, J., Kazimbaya, S. and Barker, G., 2014. *Transforming gender roles in domestic and caregiving work: preliminary findings from engaging fathers in maternal, newborn, and child health in Rwanda*. *Gender & Development*, 22(3), pp.515-531.
- ▶ Friedlander, E. and Goldenberg, C., 2016. *Literacy boost in Rwanda: Impact evaluation of a two-year randomized control trial*.
- ▶ International Labour Organization (ILO). 2011. *An introductory guide for employers' organizations: Tackling youth employment challenges* (Turin, Bureau for Employers' Activities and Programme for Employers' Activities, International Training Centre of the ILO).
- ▶ ILO. 2014. *Labour market transitions of young women and men in Bangladesh*, Work4Youth Publication Series No. 13 (Geneva).
- ▶ ILO. 2015. *Foiled by randomization: Why RCTs might be the real "gold standard" for private sector development* (Geneva).
- ▶ ILO. 2017a. Evaluating the results of an agricultural cooperative support programme: business practices, access to finance, youth employment (Geneva).
- ▶ ILO. 2017. *Towards evidence-based active labour market programmes in Egypt: Challenges and way forward* (Geneva, Employment Policy Department).
- ▶ Jacobson, N.; Butterill, D.; Goering, P. 2004. "Organizational factors that influence university-based researchers' engagement in knowledge transfer activities", *Science Communication*, Vol. 25, No. 3, March, pp. 246–259.
- ▶ Jones, N.; Villar, E., 2008. *Situating children in international development policy: challenges involved in successful evidence-informed policy influencing*. *Evidence & Policy: A Journal of Research, Debate and Practice*, 4(1), pp.31-51.
- ▶ Jones, H. 2009. The "gold standard" is not a silver bullet for evaluation, ODI Opinion No. 127 (ODI, London).
- ▶ Klein, J.-F. 2016. *Work4Youth: Improving decent work opportunities through knowledge and action – Final evaluation* (Geneva, Evaluation Office, ILO).
- ▶ Kluge, J.; Puerto, S.; Robalino, D.; Romero, J.M.; Rother, F.; Stöterau, J.; Weidenkaff, F.; Witte, M. 2017. *Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services, and subsidized employment interventions*, *Campbell Systematic Reviews* 2017:12.
- ▶ Newman, K.; Capillo, A.; Famurewa, A.; Nath, C; Siyanbola, W., 2013. *What is the evidence on evidence-informed policy making? Lessons from the International Conference on Evidence-Informed Policy Making*. *Int Netw Availab Sci Publ*.
- ▶ Pamuk, H., Bulte, E. and Adekunle, A.A., 2014. *Do decentralized innovation systems promote agricultural technology adoption? Experimental evidence from Africa*. *Food Policy*, 44, pp.227-236.
- ▶ Premand, P.; Brodmann, S.; Almeida, R.; Grun, R.; Barouni, M. 2012. *Entrepreneurship training and self-employment among university graduates: Evidence from a randomized trial in Tunisia*, IZA Discussion Paper No. 7079 (Bonn, Institute for the Study of Labor (IZA)).

- ▶ Saavedra, J.E. 2016. *The effects of conditional cash transfer programs on poverty reduction, human capital accumulation and wellbeing*, paper prepared for the United Nations expert group meeting on “Strategies for eradicating poverty to achieve sustainable development for all”, New York, 1–3 June.
- ▶ Saavedra, J.E.; García, S. 2012. *Impacts of conditional cash transfer programs on educational outcomes in developing countries: a meta-analysis*, RAND Labor and Population working paper series, WR-921-1.
- ▶ Stone, D. 2009. “RAPID Knowledge: ‘Bridging research and policy’ at the Overseas Development Institute”, *Public Administration and Development*, Vol. 29, No. 4, pp. 303–315.
- ▶ Taylor, B. 2016. *The system of evidence*. (Springfield Center).
- ▶ Weiss, C.H. 1977. “Research for policy’s sake: The enlightenment function of social research”, *Policy Analysis*, Vol. 3, No. 4, pp. 531–545.
- ▶ Weiss, C.H. 1999. “The interface between evaluation and public policy”, *Evaluation*, Vol. 5, No. 4, p. 468.



Case study

UPTAKE OF EVIDENCE ON THE EFFECTS OF SKILLS TRAINING ON YOUNG PEOPLE'S FINANCIAL BEHAVIOUR AND EMPLOYABILITY IN MOROCCO

This case study is based on the impact report "The impact of skills training on the financial behaviour, employability and educational choices of rural young people" published by the ILO (Bausch et al., 2017)

Learning objectives

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

- ▶ deploy a strategic mix of tools to communicate evaluation findings with maximum impact
- ▶ use appropriate strategies to boost the demand for evaluations and match demand with evidence supply.

Introduction and case study context

As of 2015, young people aged 15 to 29 years old made up 27 per cent of Morocco's total population. An increasing number of these youth are facing severe challenges in their attempts to secure gainful employment as they transition from school to work. Youth unemployment remains high and almost 90 per cent of young women and about 40 per cent of young men who are not in school are either unemployed or out of the labour force.

Beyond the widely reported obstacles they confront in the employment sphere, Moroccan youth face a broader challenge of economic exclusion. They struggle to establish a sound financial foundation and obtain financial services that would empower them more broadly as economic actors, including savings or loans to leverage future earnings. According to the World Bank, adult Moroccans under the age of 35 have the lowest level of awareness of financial providers and their services. Another World Bank report found that 81.4 per cent of surveyed youth in Morocco identify access to finance as the key obstacle to establishing and running their own business.

The Government and NGOs have increasingly turned to youth-targeted, supply-side interventions to equip Moroccan young people with the

skills and knowledge they need as economic actors and to enter the world of work.

This case study focuses on one of these interventions, called 100 Hours to Success, which targeted youth between the ages of 15 and 25 living in Morocco's Oriental Region.² Its curriculum consisted of three main modules: financial inclusion, life skills and entrepreneurship. An evaluation was commissioned to assess the impact of 100 Hours to Success on a range of outcomes related to financial inclusion, employability and human capital acquisition. The evaluation included 1,815 youth who expressed an interest in participating in training. Using a randomized controlled trial (RCT) design, two groups were created – a treatment group of 915 youth and a control group of 900 youth – that, on average, shared identical characteristics and only differed with respect to programme exposure. A baseline and follow-up survey were carried out three years apart.

² The 100 Hours to Success course was a primary component of the larger YouthInvest project, implemented by the Mennonite Economic Development Associates with funding from the MasterCard Foundation.

Part I: Communicating evaluation results

Evaluation findings

100 Hours to Success had a strong positive and highly significant impact of 27 percentage points on participants' likelihood to maintain a savings account, measured more than two years after the end of intervention. This effect was consistent across gender and age groups and household asset levels. The effect on maintaining a savings account was stronger among women (32 percentage points) than men (21 percentage points). This suggests that women without exposure to the training are less inclined to maintain an independent savings account (due to cultural norms).

Older individuals also seemed to benefit more from the financial knowledge and awareness training, showing a pronounced and significant impact of 0.6 standard deviations on the financial literacy index. There is weak evidence that this was also true, to a smaller extent, for men and youth from more affluent households.

There is no evidence that the effects on maintaining a savings account and financial literacy translate into impacts on actual savings, nor is there any statistically significant increase in self-reported use of a budget in maintaining personal finances.

There is evidence that participants from more affluent households were more likely to have borrowed since the start of the training, perhaps encouraged to see borrowing as a viable option to leverage future earnings and attain financial goals. And although participating may have encouraged youth from less affluent households to seek out loans, a lack of collateral or reputational credit might have constrained higher rates of borrowing.

The study finds no evidence of long-term effects on participants' self-efficacy and self-reported capacities for leadership, teamwork, problem solving and willingness to take risks. The timing of the follow-up survey (a year later than initially planned) is likely to be relevant: three years after the baseline survey, any effect is likely to have faded or been overcome by other factors of influence, including the struggle to secure employment or to achieve other long-term goals.

There were mixed effects on labour market participation and educational choice, with male participants, older participants and those from more affluent households significantly more inclined to stay in education while at the same time remaining outside the labour force. For all three subgroups, the results are driven by two trends: participants tended to remain longer in education and, if they were in education, were less likely to look for a job or to work. This behaviour seems to be consistent with youth investing more in education, both through longer attendance and by devoting less time to labour market activities. There is some rationale for considering that the subgroups' exposure to the training led some of their members to consider that investing in education would help them meet their long-term goals more readily than entering a difficult labour market.

There is no evidence that participating in the training systematically affected long-term labour market outcomes or choices related to educational attainment for women, younger and less affluent training participants.

A high level of attrition in the follow-up survey, which essentially halved the sample size compared with the baseline survey, reduced the study's statistical power. This

factor severely limited options to disaggregate findings by relevant socio-economic, demographic and geographical categories.



Discussion topics

1. The Government counterpart for 100 Hours to Success was the Ministry of Interior, and particularly their National Human Development Initiative (INDH). The head of the INDH wants a three-line summary on the impact evaluation findings and their implications for Moroccan youth. What would you write?
2. The INDH team want to know which tools they should use to increase uptake of the evaluation evidence. Can you map out a step-by-step process for them?

Part II: Building demand

Soon after the evaluation was completed, the Head of the INDH read an article in a popular development newsletter which said:

“Experimental and quasi-experimental methodologies require a ‘dosing’ intervention model, where a standard one-size-fits all output – like a training course – is delivered to all participants, regardless of their individual situation. It is like giving out an identical pill to all patients who are thought to have a particular disease. This requires a plausible counterfactual (finding a group large enough to represent a case similar to those receiving the intervention). On these terms, it is easier for some sorts of programmes to demonstrate beneficial impact, and harder for others. It suits, for example, the provision of vaccines and school dinners, or cash transfers and training.

Work in other areas is less amenable to such approaches. This includes communication and advocacy, where many complex,

interacting factors produce change (rather than any single intervention), and sector-wide approaches, where it is impossible to identify a plausible comparison group to represent what would have happened without the intervention. In a climate where policy decisions are made based on ‘evidence’ alone, these types of complex areas may come under unwarranted pressure, or lose funding, as they cannot be so easily evaluated. This bias is not only unjustified, it could generate incentives that go against key practices and hard-learned lessons about aid effectiveness, accountability and learning, and how sustainable change happens. Evidence-based approaches to impact therefore endanger learning for development interventions.”⁴

The newsletter was widely distributed in the INDH. INDH staff are now worried that the experimental evaluation, while showing interesting results, could lead to biased policy-making.

⁴ Adapted from [Jones, 2009](#).



Discussion topics

1. What would you do to build demand within the INDH for the evaluation evidence?
2. Do you agree with the article? Is there a risk of this type of policy-making bias?



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

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