

# MONITORING AND EVALUATION OF YOUTH EMPLOYMENT PROGRAMMES

A LEARNING PACKAGE





# MONITORING AND EVALUATION OF YOUTH EMPLOYMENT PROGRAMMES

A LEARNING PACKAGE

Copyright © International Labour Organization 2013 First published 2013

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 Permissions), International Labour Office, CH-1211 Geneva 22, Switzerland, or by email: pubdroit@ilo.org. The International Labour Office welcomes such applications.

Libraries, institutions and other users registered with reproduction rights organizations 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.

Monitoring and evaluation of youth employment programmes: A learning guide / International Labour Office, Youth Employment Programme, Employment Policy Department. - Geneva: ILO, 2013

ISBN 978-92-2-128074-3 (Print)

ISBN 978-92-2-128075-0 (Web pdf)

International Labour Office; Employment Policy Dept

youth employment / programme evaluation / data collecting / measurement / evaluation technique

13.01.3

ILO Cataloguing in Publication Data

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.

ILO publications and electronic products can be obtained through major booksellers or ILO local offices in many countries, or direct from ILO Publications, International Labour Office, CH-1211 Geneva 22, Switzerland. Catalogues or lists of new publications are available free of charge from the above address, or by email: <a href="mailto:pubvente@ilo.org">pubvente@ilo.org</a>

Visit our website: <a href="https://www.ilo.org/publns">www.ilo.org/publns</a>

Printed in Switzerland ATA

#### Acknowledgements

The learning package on monitoring and evaluation of youth employment programmes is one of the products developed under the Knowledge Management Facility on Youth Employment and Migration. This facility was sponsored by Spain through the Achievement Fund of the Millennium Development Goals (MDG-F).

This package was developed to support managers and practitioners engaged in the implementation of youth employment programmes. The material contained in the package was validated by managers of 15 joint programmes on youth employment and migration during two international workshops that took place in March and April 2011.

The development of the package was coordinated by Mr Gianni Rosas, Senior Youth Employment Specialist of the ILO. This included the review of literature, material and existing tools. Ms Valli Corbanese, Senior Youth Employment Expert, drafted the different modules and finalized the package.

The helpful information collected by Mr Diego Rei and Ms Mariela Buonomo Zabaleta of the ILO's Programme on Youth Employment is gratefully acknowledged. Gratitude is also expressed to the several colleagues of the ILO and the Secretariat of the MDG-F for their useful comments and for the support of the latter during the validation workshops.

# TABLE OF CONTENTS

## ACKNOWLEDGEMENTS INTRODUCTION

#### PARTI

MONITORING OF YOUTH EMPLOYMENT PROGRAMMES



#### PART II

IMPACT EVALUATION OF YOUTH EMPLOYMENT PROGRAMMES





#### Objectives of the learning guide

Countries around the world are increasingly introducing public programmes to promote youth employment. Many interventions that appear promising, however, fail to deliver results. Methods that enhance understanding of whether youth employment programmes actually work and measure their impact on beneficiaries, are therefore of importance.

The literature emphasizes three main weaknesses of youth employment programmes: First, they can be rather costly; Second they need to be well targeted (and it is often difficult to develop effective targeting mechanisms). Third there is little available evidence on which programmes work for young people outside high and middle-income countries. Evaluations carried out on youth employment programmes in many countries have found moderate, mixed and sometimes even negative results for young beneficiaries.

A good monitoring and evaluation system can minimize these weaknesses, but needs to be an integral part of the design of the youth employment programmes. This does not mean that ongoing or completed youth employment programmes cannot be evaluated. This is done on a regular basis, but requires complex statistical methods to

measure results. A more practical approach is to incorporate impact evaluation within a logical framework. This should clearly set out how the youth employment programme will produce outputs and outcomes, and combine impact evaluations with monitoring approaches to obtain a full picture of performance.

The aim of this learning material, therefore, is to introduce readers to the methods and tools commonly used to monitor and evaluate programmes for the promotion of youth employment. The package is designed primarily for self-learning and as a guide for managers of youth employment programmes responsible for carrying out monitoring tasks and making arrangements for impact evaluation.

The content of this learning material can also be used to develop training sessions to introduce the basic concepts of monitoring and evaluation of employment programmes to different audiences. It builds on the literature developed over the years on methods to measure the outcomes and impact of public policies and programmes, and shows how they are applied in practice. As such, it can also be used by readers with little knowledge about applied research methods.



This learning guide can be used as a reference tool by all those involved in the implementation of youth employment programmes. This may include, whether officials of national labour market institutions, project managers of international, national, governmental or non-governmental organizations or other labour market practitioners. It is not intended to enable readers to actually carry out monitoring and impact evaluation themselves. Rather, it aims to increase understanding of how these processes work, the results they generate, and the use programme managers and policy-makers should make of the findings to improve welfare gains for individuals and society.

# Youth employment programmes

For the purposes of this learning guide, youth employment interventions refer to programmes designed to improve the employability and employment prospects of young people aged 15 to 24 (or 29). These programmes are usually promoted by national ministries in charge of labour and/or employment and implemented by the Public Employment Services (where they exist) or by other public or private organizations (e.g. vocational training centres, private employment agencies, civil society organizations, and enterprises). These programmes may also be financed and implemented as part of donor-funded technical cooperation projects aimed at building national capacity or providing direct services to young individuals.1 As these programmes are aimed at improving the employment situation of individuals, public authorities need to ensure that results - either positive or negative - are reported in a credible, reliable and transparent manner. This is irrespective of the source of funding and modalities of implementation of the programme itself.

#### Youth employment programmes are usually grouped into five categories:



Each of these categories includes several types of interventions, summarized in the following list.

<sup>&</sup>lt;sup>2</sup>As this package focuses on how to monitor and evaluate youth employment programmes, the principles that underpin a sound youth employment programme design are only touched upon in the learning modules, and only when this serves the learning objectives of the material.

Labour market information (LMI): includes the provision of information on the labour market (e.g. economic sectors, enterprises creating jobs, geographical areas where labour demand is higher, education streams to attend to acquire the skills of an occupation on). It may include various sources and methods of information dissemination (e.g. one-to-one advice, internet-based information, job banks, leaflets, publications and brochures).

Counselling and guidance: encompasses advice provided over a period of time for employment purposes. Providers may be school teachers, employment service staff, activists of youth/community-based organizations, employer or worker organizations, peers and so on.

Job search skills: includes advice, guidance and training (of short duration) on how to approach employers, identify enterprises and economic sectors that have open vacancies, write a curriculum vitae, and manage job interviews. Again, the providers of these services may range from practitioners (e.g. employment service staff) and teachers responsible for career education to activists/members of youth/community-based organizations, and employer and worker organizations.

Job placement: is the matching of youth looking for work with available job openings. This activity is often conducted by the public employment services (or a department of the Ministry of Labour/Employment) or by private employment agencies. Other organizations may also provide job placement services, including youth offices, schools and community-based organizations.

Adult literacy programmes: are aimed at providing individuals (over compulsory school age) with basic numeracy and literacy skills. They are also called functional literacy and numeracy programmes.

Adult vocational training (off-the-job): varying in length from few weeks to one to two years, provides to individuals (of any age above the minimum age for employment and who are not enrolled in the formal education system) with the skills to perform occupational tasks. "Off-the-job" indicates that these programmes are predominantly classroom-based. Providers may be training centres, schools, community-based organizations, non-

1 <<<<< Employment services

2 <<<<< Skills development governmental organizations, trade unions, employer organizations or chambers of commerce.

Adult vocational training (on-the-job) including apprenticeship schemes: these programmes differ from off-the-job vocational training programmes in that learning occurs through the performance of tasks in a real work setting, usually an enterprise.

The concept of apprenticeship varies widely from country to country. However, here it refers to programmes that blend training (which can be on- or off-the-job) with work. Normally, apprenticeship programmes are implemented for specific trades (such as welding or carpet weaving) and envisage a minimum period of time spent under the supervision of a mentor/expert worker. These programmes may lead to a formally recognized certificate (formal apprenticeship) or not (non-formal apprenticeship).

#### Second chance programmes (and equivalency programmes):

target individuals that missed out on educational and labour market opportunities. This includes youth who dropped out from school and need a diploma to enter the labour market, young offenders about to re-enter society and so on. These programmes may include counselling, training, housing assistance, employment subsidies and public works, among others.

Financial incentives for training: are monetary incentives provided either to enterprises or training providers (in cash or as tax rebates) or to beneficiaries (generally in the form of training vouchers or reimbursement of training expenses). They may target both unemployed individuals and young workers (in the formal or informal economy).

>>>> 3 Employment

Wage subsidies: aim to reduce the cost of labour for enterprises. They may take the form of tax relief (i.e. enterprises recruiting beneficiaries from the target group receive a discount on the profit or labour taxes to be paid) or monetary grants to the enterprise to cover part or all of the total labour cost of the person hired? They can also take the form of vouchers that employers can redeem when creating a new position for a young person.

<sup>&</sup>lt;sup>2</sup> Total labour costs include the direct remuneration that goes to the worker (take home pay), bonuses and allowances, and indirect costs paid by the employer, such as social security contributions and taxes relating to employment.

Public works (or community-based) programmes: allow a central/local government donor agency, or non-governmental organization to finance and/or implement a programme that creates temporary job opportunities. The output of these programmes is jobs of short duration, which generate income for individual participants (smoothing consumption/poverty reduction), and the creation of public goods in the form of new or improved infrastructure, or delivery of services.

Self-employment programmes (training, grants and loans): aim to provide the necessary support for young people to open up a business. They may include: advice on viable business ventures; training courses on how to establish and manage a small business; non-refundable grants; (refundable) credits; and specialized services (marketing, finance, export markets, supply chains and so on).

Entrepreneurship development (information, education and training): aims to provide information and training on what it means to be an entrepreneur. Generally these programmes target students or school leavers.

Integrated programmes: are comprehensive, multiple-service interventions that mix the whole range of programmes outlined above (e.g. employment services, skills development and employment creation) and below (e.g. work-experience programmes and grants) to address the multiple barriers a person might face in entering the labour market. Different types of programmes (e.g. counselling, job search skills, vocational training and public works) may be combined either in sequence or be delivered at the same time. Some countries implement Youth Guarantee Schemes, where young unemployed people (usually registered with the Public Employment Service) are provided with intensive counselling, vocation guidance and individual employment planning, and are offered a job, a study place (either academic or vocational) or other activation measure to help increase their employability. Youth Employment Guarantees are characterized by an obligation on the part of the employment services to engage each young person who remains unemployed for a pre-established period of time (usually three to four months).

4 <<<<< Integrated programmes >>>> 5 Other types of programmes/ services Work experience programmes: aim at providing individuals with (paid or unpaid) work experience in a public or private sector enterprise. They include internships organized by education institutions for their students.

Grants for transport, childcare and others allowances: are provided either in cash or as a reimbursement of the costs associated with programme participation (e.g. training or public works).

Other monetary and non-monetary entitlements: include any type of entitlement provided to young people to address specific barriers to job searches or work (e.g. telephone use for contacting employers, new clothing for homeless jobseekers and so on).

#### Structure and content of the learning package

This learning material is divided into two parts. The first part offers a guide to monitoring the performance of youth employment programmes and measuring short-term programme outcomes (e.g. monitoring and evaluation principles, performance monitoring, indicators, target and standard setting and establishment of performance monitoring systems for youth employment programmes).

The second part deals with evaluation approaches and techniques to estimate the impact of a programme (e.g. experimental and quasi-experimental approaches, randomized controlled trials, difference-in-differences, regression discontinuity design and matching techniques).

The content of each part is organized into learning modules. Each module discusses the topic of interest and offers a number of additional resources:

- Additional reading for those who would like to learn more about the topic;
- Examples that show how principles, concepts and methods are applied in practice;
- ✓ Glossary of the key terms used in the module;
- Additional tips to understand the concepts discussed in the module (Good to know...);
- Exercises to allow the reader to check his/her understanding of the topic (suggested answers can be found at the end of the exercises); and
- Questions and answers (Q&A) on the key areas discussed in each module.

The toolkit attached to this learning material includes additional tools for programme managers and practitioners responsible for monitoring the performance and evaluating the impact of youth employment programmes. They include terms of reference for contracting baseline surveys and an impact evaluation, sample survey questionnaires and a tabulation plan.



# Introduction to monitoring and evaluation principles

1

- 1.1. Monitoring and evaluation principles
- 1.2. Monitoring and evaluation (M&E) systems
- 1.3. Theory of change and result chain (logical framework)

Introduction to monitoring and evaluation principles

# Introduction to monitoring and evaluation principles

Learning objectives

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

Distinguish between traditional and results-based monitoring and evaluation;

Outline the rationale for establishing sound monitoring and evaluation (M&E) systems; and

List different types of approaches that can be used to evaluate a youth employment programme.

What fits?

Employment programmes, like any other type of public policy intervention, are designed to change the current situation of the target group and achieve specific outcomes (results), such as increasing employment, rising earnings or reducing unemployment. The key policy question is whether the planned results were actually achieved. Often, in fact, the attention of policy-makers and programme managers is focused on programme inputs (e.g. the human and financial resources used to deliver a training or a wage subsidy programme) and outputs (e.g. number of graduates of a training programme or number of workers placed in subsidized employment), rather than on whether the programme is achieving its intended outcomes (e.g. more individuals employed or with the skills needed to get productive jobs).

Monitoring and evaluation are the processes that allow policy-makers and programme managers to assess how an intervention evolves over time (monitoring). They examine how effectively a programme was implemented and whether there are gaps between the planned and achieved results (evaluation). Lastly, they determine whether the changes in well-being are due to the programme and to the programme alone (impact evaluation).<sup>3</sup>

In the last two decades there has been a considerable shift in the design and implementation of monitoring and evaluation (M&E) systems. Today, monitoring and evaluation form the basis of evidence-based policy-making, where the focus is on results (outcomes), rather than just inputs and outputs. Monitoring and evaluation are complementary processes. This means that both need to be carried out to provide a complete picture of how a programme performs. Monitoring and evaluation frameworks provide the basis for decision-making and designing development initiatives. Monitoring links resources with results (accountability) and helps to identify and address bottlenecks in implementation (effectiveness). Evaluation provides decision-makers and programme managers with objective evidence on what works and for whom, and allows them to plan strategically.

1.1
Monitoring and evaluation principles

<sup>3</sup> Monitoring and evaluation can be conducted using a variety of tools, methods and approaches. These include, among others: performance monitoring, the logical framework approach, theory-based evaluation, rapid appraisal methods such as key informant interviews and focus group discussions, public expenditure tracking surveys, rigorous impact evaluation, and cost-benefit and cost-effectiveness analysis. Some of these methods can be combined, while others are interchangeable. The remaining modules of this training material offer the sequence of approaches that were found through practice to be best suited to monitoring and evaluating youth employment programmes.

1

## Introduction to monitoring and evaluation principles

# What is monitoring?

Monitoring is a continuous process of collecting and analysing information about a programme, and comparing actual against planned results in order to judge how well the intervention is being implemented. Using the data generated by the programme itself (e.g. characteristics of individual participants, enrolment and attendance, end of programme situation of beneficiaries and costs of the programme), monitoring tracks performance against expected results. It makes comparisons across individuals, types of programmes and geographical locations, and analyses trends. The existence of a reliable monitoring system is essential to evaluate a programme. Without the information collected through monitoring, it is almost impossible to appraise the relevance, effectiveness and sustainability of the results of a programme, or to measure whether it had an impact on the well-being of beneficiaries.

## What is evaluation?

**Evaluation** is the assessment of an intervention to determine its relevance, efficiency, effectiveness, impact and sustainability. It is a process that systematically and objectively assesses all the elements of a programme (e.g. design, implementation and results achieved) to determine its overall worth or significance. <sup>4</sup> The objective is to provide credible and useful information for decision-makers to identify ways to achieve more of the desired results.

4 For the purposes of this learning material, a "programme" refers to an intervention (see the Introduction) aimed at improving the employability and/or employment prospects of a young individual.

#### Broadly speaking, there are two main types of evaluation:

Performance evaluations focus on the quality of service delivery and the outcomes (results) achieved by a programme. They typically cover short-term and medium-term outcomes (e.g. student achievement levels, or the number of welfare recipients who move into full-time work). They are carried out on the basis of information regularly collected through the programme monitoring system. Performance evaluation is broader than monitoring. It attempts to determine whether the progress achieved is the result of the intervention, or whether another explanation is responsible for the observed changes.

Impact evaluations look for changes in outcomes that can be directly attributed to the programme being evaluated. They estimate what would have occurred had beneficiaries not participated in the programme. The determination of causality between the programme and a specific outcome is the key feature that distinguishes impact evaluation from any other type of assessment.

**Performance evaluation:** This technical cooperation project aimed at piloting school-to-work transition programmes (apprenticeship, internship, in-company training and wage- and self-employment measures) targeting disadvantaged youth.

The programmes yielded a placement rate of 46 per cent, with most beneficiaries employed in service and light manufacturing enterprises. The measures having the highest placement rates were internships (52 per cent placement rate, partially due to the higher educational level of beneficiaries), followed by wage subsidies (50 per cent), in-company training (45.8 per cent) and lastly apprenticeship (40 per cent). While internship, wage subsidies and in-company training worked equally well for both sexes, internship and apprenticeship had higher placement rates for young men. Compared to the vocational training programmes offered by public training centres, the piloted programmes provided better results in terms of placement and overall costs. The placement rate of participants to vocational training programmes was approximately 26 per cent, with an average cost per young person employed of €1,000. The average cost of the school-towork transition measures per participant placed was €640.

#### Example <<<<<

Skills
development
for the
recovery and
reconstruction
of Kosovo:
Phase II
(2004–07)

Impact evaluation: The objective of this programme was to provide young people (aged 18–24) with a work experience period to increase their employability. This subsidized employment programme placed participants in both private and public sector enterprises. To minimize displacement effects, participants were assigned tasks that otherwise would not have been carried out and asked to allocate four to eight hours a week to job search. All unemployed youth aged 18–24 could apply to the programme, the only eligibility criterion being the duration of unemployment (at least two months).

The programme was evaluated for young people aged 21 to 24 on three criteria: post-programme annual earnings, probability of re-employment and re-entry into regular education. The control group comprised young people in the same age group, registered as unemployed during the period of the programme, but who did not participate. A probability score matching approach was used to identify individuals with similar characteristics of participants.

Example <<<<

Youth practice, Sweden (1992-95)

2 3 4 5

1

Introduction to monitoring and evaluation principles

#### >>>> Example

Youth practice, Sweden (1992-95)

The results of the evaluation indicate that the programme had negative effects on earnings and employment in the first year following the end of the programme. Two years after, however, the estimates were positive, but not statistically significant. Also, there was no significant effect on the probability of re-entry into regular education.

Source: Youth Employment Inventory (www.youth-employment-inventory.org).

Monitoring and evaluation usually include information on the cost of the programme being monitored/evaluated. This allows users of this data to judge the benefits of a programme against its costs and to identify which intervention has the highest rate of return. The tools generally used for this are cost-benefit and cost-effectiveness analyses.

A *cost-benefit analysis* estimates the total benefit of a programme compared to its total costs. This type of analysis is normally used ex-ante, to decide among different programme options. The main difficulty is to assign a monetary value to "intangible" benefits. For example, the main benefit of a youth employment programme is the increase of employment and the earning opportunities for participants. These are tangible benefits to which a monetary value can be assigned. However, having a job also increase people's self-esteem, which is more difficult to express in monetary terms as it has different values for different persons.

A cost-effectiveness analysis compares the costs of two or more programmes in yielding the same outcome. Take for example a wage subsidy and a public work programme. Each has the objective to place young people into jobs, but the wage subsidy does so at the cost of \$500 per individual employed, while the second costs \$800. In cost-effectiveness terms, the wage subsidy performs better than the public work scheme.

There are a number of reasons why monitoring and evaluation should become an integral part of the design of youth employment programmes. They help managers and decision-makers to understand which programmes were successful and why, and which were not. As such, monitoring and evaluation are organizational learning tools which help to design programmes that respond to the needs of the

6 7 8 9 10

target group and take decisions on resource allocation. A more rational use of human and financial resources, in turn, improves accountability and transparency in public service delivery.

A growing number of countries are working to improve the delivery of public services by creating systems to measure performance. These monitoring and evaluation (M&E) systems are used to measure the quantity, quality and targeting of the goods and services (outputs) provided by the government, and appraise their outcomes and impact. There are several reasons for this. Fiscal pressures and rising expectations from citizens drive governments to provide more services with higher standards of quality. Parliaments and civil society are increasingly requiring governments to publicly report and explain the performance of the goods and services provided (accountability). Donors increasingly require beneficiary countries to establish systems able to identify leakages in government funding and fight corruption. Governments build M&E systems because these directly support core government activities (e.g. the budget process, national planning, policy and programme development, and the managements of public institutions) and provide information for accountability purposes. This is why the introduction of M&E systems is often linked to public sector reforms such as performance-based budgeting, evidence-based policy-making and results-based management.

In the field of active labour market programmes, and especially youth employment programmes, the rationale for setting up a M&E system centres on:

1.2 Monitoring and Evaluation (M&E) systems

1 <<<<<

Improving the quantity, quality and targeting of services and programmes

2 <<<<

Justifying the use of resources and leveraging for more

Enhancing accountability (towards beneficiaries, social partners, supervisory bodies and civil society)

Improving the management of core functions of institutions with a mandate on employment

2 3 4 5

Introduction to monitoring and evaluation principles

The experience of countries with effective M&E systems points to a number of lessons learned.

There must be demand from a government as a pre-requisite for the establishment of an M&E system. Such a system must produce monitoring information and evaluation findings that are valued by key stakeholders, used to improve government performance, and generate demand for the M&E function to ensure its sustainability and funding. A common barrier to demand is lack of knowledge about what M&E actually includes, particularly where the buy-in of government ministers is necessary to raise resources. Demand can be increased once key stakeholders understand M&E better, when they are exposed to examples of cost-effective systems, and when they became aware of the returns on investment. Senior officials tend to use high-level strategic indicators such as outputs and outcomes, while line managers and their staff tend to focus on operational indicators that measure processes and services.

Strong incentives are needed both to perform M&E well and to promote the use of its findings. The availability of M&E information does not guarantee its use by programme managers, budget officials responsible for advising on spending options, or by a parliament responsible for accountability oversight. The objective of M&E systems is not to produce a large volume of performance information and evaluation findings, but rather to effectively use whatever information exists to support core government functions.

A common feature of effective M&E systems is the lead role taken by a ministry that can design, develop and manage the system. In many countries this role falls under the purview of the ministry of finance. However, there are also examples of ministries of labour and social protection successfully taking the lead in establishing M&E systems. In Mexico, for example, the Secretariat for Social Development (SEDESOL) manages an M&E system centred on both qualitative and impact evaluations.

The M&E system needs to produce reliable and credible information. Data verification and credibility is partly a technical issue of accuracy, procedures and quality control. Related to this is the need for data to be available on a timely basis, easy to understand and consistent over time. There is also the issue of objectivity. As performance information becomes increasingly important, especially when it is used for accountability purposes, so does the need for objective appraisal. This is a concern that can be addressed

in different ways. In Latin American countries, for example, it is usual for evaluations to be commissioned by government departments, but to be conducted externally. The advantage of this approach is that it increases the perceived credibility of the evaluation. The challenge, however, is ensuring full ownership of both the evaluation process and its findings by the authorities whose programmes are being evaluated. In OECD countries it is more common for evaluations to be conducted internally. This helps to ensure that the evaluations draw on expertise of the agency's staff, and it encourages the use of findings. To ensure objectivity, a higher-level institution supervises the evaluation process (usually the ministry to which the agency reports).

Finally, most countries with well-performing M&E systems have not developed them in a linear manner (e.g. starting with a clear understanding of what the system would look like once it matured and then progressively achieving this vision). Instead, such systems have been developed incrementally and in some cases even in a piecemeal manner.



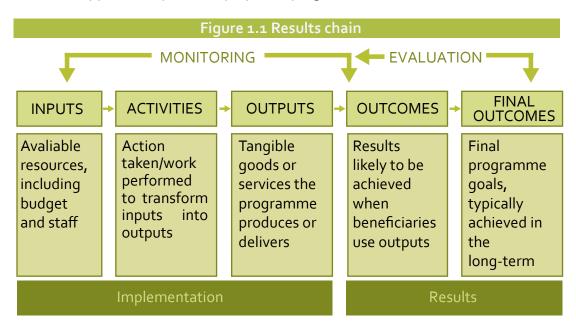
2 3 4 5

Introduction to monitoring and evaluation principles

Theory
of change
and results
chain (logical
framework)

A theory of change describes how an intervention will deliver the planned results. It maps the casual logic of why and how a particular programme will reach its outcomes. A causal chain (or logicalframework)outlineshowthesequenceofinputs, activities and outputs of a programme will attain specific outcomes (objectives). This in turn will contribute to the achievement of the overall aim. A results chain maps: (i) inputs (financial, human and other resources); (ii) activities (actions or work performed to translate inputs into outputs); (iii) outputs (goods produced and services delivered); (iv) outcome (use of outputs by the target groups); and (v) aim (or final, long-term outcome of the intervention). The results chain, together with a well-designed M&E plan, is a key monitoring tool. It details all the information needed to measure whether progress is being made towards the expected results. A results chain (or logical framework) – developed during the planning phase – also helps evaluators to understand the linkages across activities, processes and outcomes.

Figure 1.1 below shows a results chain that includes the monitoring and evaluation functions, followed by an example applied to a youth employment programme.



The Ministry of Labour is planning to modernize its adult training system to improve the employment and earnings prospects of out-of-school youth (15–24 years of age). The reform includes the introduction of a competency-based vocational training system, the refurbishing of vocational training centres, the production of new learning material and the training of trainers.

- The inputs of the programme include the staff of the Vocational Training Department of the Ministry of Labour and the vocational trainers, as well as the funds for the refurbishing of training centres, the development of new training programmes and the design of competency-based curricula.
- The activities comprise the development of competency-based vocational training curricula, the training of trainers, the development of competency-based training materials and the refurbishment of training centres.
- The outputs of the programme are competency-based curricula for 50 occupations, 5,000 trainers trained in the new methodology, 100,000 sets of training materials prepared, and 15 training centres refurbished.
- The outcomes are trainers that use the new competency-based training materials and trainees that follow the new curricula in refurbished training centres. Medium-term outcomes are improved trainees' performance on competency-based assessments.
- The final outcomes are improved completion rates and higher employment and earnings of trainees.

In the example above, the monitoring system would continuously track: (i) the resources invested in the programme (number of days/staff, amount of funds spent); (ii) the implementation of activities in the planned timeframe (refurbishing of training centres, development of competency-based curricula, production of training materials and design of training programmes); and (iii) the delivery of goods and services (delivery of training courses and materials to trainees, and refurbished training centres) in the quantity and quality planned.

#### Example <<

Results chain framework

1

## Introduction to monitoring and evaluation principles

A performance evaluation would, at a specific point of time, judge the inputs-outputs relationship and the immediate outcomes. It then assesses the design and implementation of the intervention against criteria of relevance, effectiveness and efficiency. A performance monitoring system supports the evaluation by providing information on trainees' completion rates, the share of participants that passed competency-based tests, and their employment rates and level of earnings after programme's end.<sup>5</sup>

An impact evaluation would provide evidence on whether the changes in completion rates and labour market performance of trainees were caused by the intervention and this alone. A practical example of how a monitoring and evaluation system can be set up for the adult training reform outlined above is provided in the following example.

5 Gross employment rate in the context of performance measurement indicates the overall employment found for participants at follow-up. It is termed "gross", because it is the sum of the employment gain caused by the programme (net employment rate) and the employment gained that is due to other factors outside the programme.

#### >>>> Example

Monitoring and evaluation system (reform of the adult training system)

#### 1. Monitoring system

The Ministry of Labour has given the responsibility of implementing the reform to the Vocational Training Department. The total amount of funds available is \$120 million over three years. This amount is based on cost estimates prepared by the Ministry of Labour and the Treasury.

The Adult Training Department prepares a three-year implementation plan that details the activities to be carried out with the annual funding allocation (\$40 million). For each of the outputs envisaged by the intervention (e.g. competency-based curricula for 50 occupations, 5,000 teachers trained, 100,000 sets oftraining materials prepared and 15 training centres refurbished), the management of the Adult Training Department: (i) lists the inputs needed, the activities to be carried out, the outputs to be produced and the outcomes to be attained; (ii) develops the indicators that will measure progress at each level of the results chain; and (iii) sets the targets to be achieved for each period (quarterly, semi-annually, annually).

For example, the development of competency-based curricula for 50 occupations includes: (i) the identification of

occupations most demanded by the labour market through occupational skills surveys; (ii) the establishment of a curriculum development team for each occupation; and (iii) the design and validation of new curricula for each occupation selected.

The indicators could be: (i) the number of occupational skills surveys carried out; (ii) the number of priority occupations identified and their skills content; (iii) the number of curricula development teams set up; and (iv) the number of new curricula drafted and validated.

The targets may include: (i) two occupational skills surveys implemented in the first two years; (ii) at least 20 curriculum development teams appointed, with each team comprising education and training practitioners, representatives of employers and workers organizations; (iii) at least 20 curricula drafted; and, (iv) at least 10 new curricula validated.

For each indicator and target, the department needs to decide how, when and who is responsible to gather the information needed. For example, to measure progress towards achievement of the target "20 curriculum development teams appointed", it will be necessary for someone to organize and maintain the list of members for each team, record how many times they meet to work on curricula, and, generally, keep track of how the team is progressing on the work assigned.

The same process is repeated for each of the outputs. At regular intervals (once every quarter, semi-annually or annually), all this information, including the overall costs incurred, is systematized and checked against the original plan. This information is used to correct gaps in performance and prepare planning for the next period.

A good monitoring system requires the development of indicators and targets at each level of the results chain (inputs, activities, outputs and outcomes). Indicators should be well chosen and limited in number. For each indicator and target, it is then necessary to understand how to organize the collection of information, the costs involved, and to assign responsibilities for its continuous tracking. For example, at the level of outcomes (results), it is necessary to set up a database to store information on trainers that participated in training, trainees enrolled in the reformed programmes, dropouts and graduates, and information on labour market performance once trainees leave the programme. Information on the ultimate beneficiaries (in this example out-of school youth aged 15–24) should be disaggregated by sex and other personal characteristics to measure how different groups are benefitting from the reform (e.g. young boys vs. young girls; poor youth vs.

#### Example

<<<<<

Monitoring
and
evaluation
system
(reform of the
adult training
system)

cont.

2 3 4 5

1

Introduction to monitoring and evaluation principles

>>>> Example

Monitoring

and evaluation

system

(reform of the

adult training

cont.

system)

non-poor; rural vs. urban youth; national majority youth vs. national minority and so on).

#### 2. Evaluation system

The evaluation focuses on results over different periods: short-term outcomes (e.g. trainers using the new training material and trainees attending new training programmes in refurbished training centres); medium-term outcomes (e.g. improved completion rates, better performance of trainees in competency-based assessments); and long-term outcomes (e.g. higher employment and earnings levels for trainees).

For short- and medium-term results a performance evaluation (based on the performance monitoring system highlighted above) would reveal: (i) how many of the 5,000 trainers who were trained are actually using the new curricula, the upgraded training methodologies and the newly developed training material; (ii) how many of the trainees attending adult training programmes are following the reformed curricula and how they perform in the final assessment compared to others; and (iii) how many trainees exposed to the new methods complete their training programme compared to others (or prior to the intervention).

To be able to calculate all of the above, the monitoring system set up by the Adult Training Department needs to have reliable records on: (i) the trainers trained in the new methodologies; (ii) young people (and their characteristics) that enrolled and attended the reformed training courses, as well as those enrolling and attending the old type of courses (if any); (iii) results of the final testing of participants to the new and the old programmes; and (iv) employment and earning levels of trainees (and their individual characteristics) who enrolled, attended and completed new and old training programmes.

A simple comparison between the performance of trainees graduating under the reform programme with those that graduated under the old one (by sex, field of study attended, rural/urban divide and any other relevant characteristic) will provide indications on whether the new programmes are yielding higher, lower or the same results (see Module 2 to 5 of this learning material).

An impact evaluation, on the other hand (see Modules 6 to 10 of this learning material) will measure whether the changes (positive or negative, intended or unintended) in employment and earnings of trainees were caused by the reformed programme, and by this alone.

#### **Aditional reading**

**Gertel, P.J., Martinez, S. et al.** 2011. Impact evaluation in practice (Washington, DC, World Bank).

**Khandker, S.R., Koolwal, G.B., Samd, A.H.** 2009 Handbook on impact evaluation. Quantitative methods and practices (Washington, DC, World Bank).

**Mackay, K.** 2007. Systems to support better government (Washington, DC, World Bank Independent Evaluation Group).

Morra-Imas, L.G, Rist, R.C. 2009. The road to results: Designing and conducting effective development evaluations (Washington, DC, World Bank).

**O'Leary, C.J., Nesporova, A., Samorodov, A**. 2001. Manual on evaluation of labour market policies in transition economies (Geneva, ILO).

Wholey, J.S., Hatry H.P. 1992. "The case for performance monitoring", in Public Administration Review (Vol. 52, No 6), pp. 604–610.

**Zall Kusek, J., Rist, R.C.** 2004. Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners (Washington, DC, World Bank).



Introduction to monitoring and evaluation principles



## Good to know...

Performance monitoring and evaluation cannot serve as a substitute for impact evaluation.

Performance must be monitored on a continuous basis to judge whether the programme is being implemented as planned and outcomes are consistent with the results expected. As a management tool, performance measurement should provide rapid feedback on the operation of a programme, but it cannot measure its impact.

#### There is no "best" type of evaluation.

Much depends on the programme being evaluated and the questions that the evaluation se eks to answer. Ideally, both performance monitoring and impact evaluation should be carried out on employment programmes. However, impact evaluations are not always feasible. When this is the case, well-executed performance monitoring, combined with a cost-effectiveness analysis, may be the best approach.

#### >>>>>

#### **Glossary**

#### Performance monitoring

Performance monitoring is the continuous process of collecting and analysing information to assess how a programme is performing. It relies on implementation data to track performance against expected results, make comparisons across programmes, and analyse trends over time.

#### Performance evaluation

Evaluations are periodic, objective assessments of a programme to determine its relevance, efficiency, effectiveness, impact and sustainability. Performance evaluation assesses the quality of the service delivered and the outcomes (results) achieved. It typically covers short- and medium-term outcomes, but not casual impact.

#### Impact evaluation

An impact evaluation is an evaluation that determines whether there is a causal link between the programme and a specific outcome. An impact evaluation answers the question of

#### Glossary < <<<<

whether the programme is responsible for the changes observed (e.g. the positive/negative, intended/unintended changes are attributable to the programme and not to other factors).

cont.

#### Results chain framework

The results chain sets out the programme logic that explains how the final objective is to be achieved. It shows the links from inputs to activities, to outputs and to outcomes, to understand how the changes brought about by the programme affects the well-being of individuals.

#### **Evidence-based policy-making**

This approach helps decision-makers to make informed decisions about policies and programmes by placing the best available scientific evidence at the heart of policy design and implementation.

#### Cost-benefit analysis

This refers to the total costs and total benefits used to appraise (ex ante or ex post) an intervention. Cost-benefit can be calculated ex-post in impact evaluations if most of the benefits can be monetized and the information on programme costs is available.

#### Cost-effectiveness analysis

This analysis compares two or more programmes based on their respective costs to achieve the same (or similar) outcome.

Introduction to monitoring and evaluation principles



# **Questions** and answers

#### Why are monitoring and evaluation necessary?

Monitoring and evaluation provide objective and systematic information on the design, implementation and results of a programme to judge what was successful, what not and why. They provide managers and decision-makers with the tools to decide what should be continued or scaled up, what should be cancelled, and where to invest human and financial resources.

#### Which is the best type of evaluation?

As different types of evaluation respond to different purposes, there is no best type. The most appropriate evaluation method will depend on the programme to be evaluated and the questions that the evaluation should provide answers to.

#### When should an impact evaluation be carried out and when not?

Impact evaluations may be long and costly. Therefore, their use is generally warranted when the programme is new and needs to be tested; when there is the intention of extending its scope to affect large numbers of individuals or communities; or when the results are intended to inform key policy decisions. When these conditions are not satisfied, the costs of an impact evaluation are likely to outweigh its benefits.

### What is the difference between monitoring and performance monitoring?

Monitoring tracks inputs, activities and outputs on a continuous basis. Performance monitoring goes a little further, measuring short- and mid-term outcomes of a programme to judge whether these were achieved. Although termed "monitoring", it is in fact an evaluation, falling just short of impact evaluation because it is unable to establish causality between the changes in an outcome and the programme.

#### Is it always necessary to doperform a cost-benefit or costeffectiveness analysis?

Analysis of the costs of a programme is an important measure of efficiency and effectiveness (namely the relation between costs and outputs and between costs and outcomes). To do this, a cost-effectiveness analysis is sufficient, because it compares the cost of the intervention being evaluated against those of a similar programme (e.g. it gives a relative measure of the costs). A cost-benefit analysis, conversely, measures the costs of a programme or policy in absolute terms, by comparing costs against benefits. A cost-effectiveness exercise is relatively simple, while a cost-benefit analysis is more complex (and hence costly), as it also attempts to monetize intangible benefits (likeincluding subjective measures such as self-esteem and well-being).

During an informal meeting, the Head of the Employment Department of the Ministry of Workforce Development informs you that they are planning to introduce a job voucher programme targeting youth aged 18–24 to increase their employment levels. The scheme is a letter of guarantee to pay \$150 per month for up to six months to private sector enterprises for employing young people that have received the voucher.

The Ministry plans to pilot the voucher in one province of the country only (for a total of 600 vouchers) for a year, so as to refine the design and implementation procedures prior to scaling up the following year. The Ministry is interested to find out whether the voucher works better than the existing wage subsidy programme. This is the refund provided by the National Employment Agency (NEA) of social security contributions paid by employers for the recruitment of unemployed individuals for a maximum of six months. The Head of the Department asks your thoughts on which type of evaluation would better respond to their needs.

#### Exercise <<<

What fits?

#### Check your answer

Performance monitoring combined with a cost-effectiveness analysis will probably respond to most of the Ministry's needs. Aside from tracking inputs, activities and outputs (where the effectiveness of the voucher administration can be judged), performance monitoring also measures short-term outcomes (in this case, the employment rate of participants at the end of the voucher scheme). This can be compared with the gross placement rate of the standard wage subsidy. The same applies to costs.

To give a simple numerical example, suppose that at the end of the voucher programme 420 young people are working (this is a placement rate of 70 per cent) and the total cost of the programme is \$450,000 (600 vouchers\*\$150\*average five months). This means that the programme cost was \$1,071 per youth employed. If the average cost per person employed of the wage subsidy is higher than that of the voucher, the voucher is more effective. If it is lower, the wage subsidy is more effective. Another performance measurement that can be done is to compare average hourly earnings of participants employed to see which programme yields higher hourly earnings. For example, the wage subsidy may achieve a higher level of employment,

2 3 4 5

Introduction to monitoring and evaluation principles

but in low wage jobs, while the voucher helps participants to get jobs with higher wages.

Given the pilot nature of the voucher scheme, an impact evaluation would establish with certainty whether the programme increases the probability of employment. There are two issues, however, that may make impact evaluation impractical in this case. First, the timeframe envisaged for the scaling up: one year after programme start may be too short a time to allow the full effect of the programme to materialize. Second, the number of participants involved (600) may be too low to obtain reliable estimates (see Part 2 of this learning material).



# Performance monitoring of youth employment programmes

- 2.1. Performance management systems: and introduction
- 2.2. Performance measurement of youth employment programmes
- 2.3. Formulating the objectives of youth employment programmes

2

Performance monitoring of youth employment programmes

# Performance monitoring of youth employment programmes

Learning objectives

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

List the main processes of performance (or results-based) management systems;

Outline the key features of performance measurement of youth employment programmes; and

Specify the objectives of a youth employment programme for performance measurement purposes.

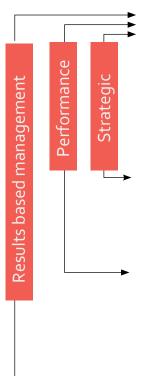
What are the objectives ...?

Performance management, also known as results-based management, is a strategy designed to achieve changes in the way organizations operate, with improving performance (better results) at the core of the system. Management systems that track outcomes, estimate programme impact and replace less effective programmes with more effective ones, have been a key factor in the development of more effective public policies in many developed countries since the late 1980s.<sup>6</sup>

**Performance measurement**, also known as results-based or performance monitoring, is concerned more narrowly with the production of information on performance. It focuses on defining objectives, developing indicators, and collecting and analysing data on results.

The basic purpose of a results-based management system is to generate and use performance information for organizational learning and decision-making purposes. Results-based management systems comprise seven stages:

**2.1**Performance management systems: an introduction



- 1. Formulating objectives: identifying in clear, measurable terms the results being sought and developing a conceptual framework for how the results will be achieved.
- **2. Identifying indicators:** for each objective, specifying exactly what is to be measured along a scale or dimension.
- 3. wSetting targets: for each indicator, specifying the expected level of results to be achieved by specific dates, which will be used to judge performance.
- **4. Monitoring results:** developing performance-monitoring systems that regularly collect data on the results achieved.
- **5. Reviewing and reporting results:** comparing actual results against the targets (or other criteria for judging performance).
- **6. Integrating evaluations:** conducting evaluations to gather information not available through performance monitoring systems.
- **7. Using performance information:** using information from monitoring and evaluation for organizational learning, decision-making and accountability.

6 Such an approach has also been deployed by development cooperation agencies since the late 1990s to increase the effectiveness of their operations. As such, result-based management is now applied by these agencies at project, country programme and agencywide level. See OECD-DAC: Results based management in the Development Cooperation Agencies: a review of experience. Background report, (Paris, OECD, 2000), www.oecd.org/dataoecd/17/1/1886527.pdf

**1 3 4 5** 

Performance monitoring of youth employment programmes

The first three phases relate to strategic planning. Stages one to five make up performance measurement, while the seven phases together constitute a results-based management system, which integrates monitoring and evaluation information and ensures that it is used for strategic planning purposes.

The most widely known type of performance measurement is found at the project level. Here, the results chain (or logical framework) plots how project inputs, activities, outputs and outcomes will contribute to the achievement of the overall aim. As a management tool, performance measurement is also used to assess the quality of service delivery and the outcomes achieved by public policies and programmes (e.g. health, education, employment and social services).

In the employment field, performance measurement is used to compute the outcomes of employment services and programmes. It usually applies to the activities carried out, either directly or through contractors, by the Public Employment Service (PES) to deliver on the employment policy objectives of a country. However, it can (and should) be used to determine the outcomes of any programme aimed at improving the labour market outcomes of individuals, irrespective of the scale of the programmes and of the organization/agency that implements them. As such, performance measurement of employment programmes may well be part of a (technical cooperation) project implemented with donor funds by an international development agency or a national or international non-governmental organization. When a technical cooperation project implements, directly or indirectly, youth employment programmes, two types of evaluations need to be planned. The first is performance measurement of the employment programmes and services implemented under the aegis of the technical cooperation project. The second is evaluation of the technical cooperation project itself (see example below).

Youth
Employment
Partnership in
Serbia (YEPS,
2007–10)

TheYEPS Project, implemented by the International Labour Office (ILO) from 2007 to 2010, had two main objectives. The first is strengthening the capacity of labour market institutions to develop and implement the youth employment policy. The second is piloting an integrated package of active labour market programmes aimed at enhancing the employability and improving employment prospects of disadvantaged youth.

The project carried out a performance measurement exercise of the pilot youth employment programmes implemented in collaboration with the Ministry of Economy and Regional Development (MoERD) and the National Employment Service of Serbia (NES) in April 2010. These included vocational training, wage subsidies and work training contracts. The exercise measured the employment rate and average monthly earnings of young people who participated in the programmes designed by the project. A follow-up survey was carried out to detect beneficiaries' labour market status (employment/unemployment) and earnings after participation. The results were benchmarked against those of other youth employment programmes implemented by the employment service in the same period.

The findings of the performance measurement exercise informed the project evaluation, which was conducted at the end of 2010. This latter aimed to assess the relevance, effectiveness, efficiency and sustainability of all components of the project (capacity building, youth employment policy development, programme design and implementation).

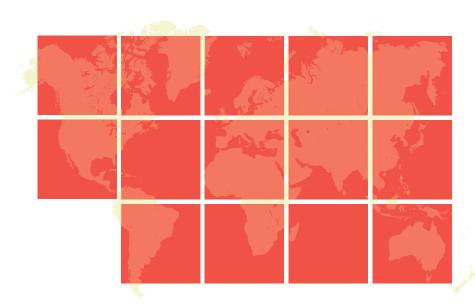
## Example

<<<<<

Youth
Employment
Partnership in
Serbia (YEPS,
2007–10)

cont.

Source: Youth employment and partnership in Serbia (Final Evaluation Summary), Project SRB/o7/o1/ITA, www. ilo.org/eval/Evaluationreports/ WCMS\_150344/lang--en/index. htm



**1 3 4 5** 

Performance monitoring of youth employment programmes

# Performance measurement of youth employment programmes

Performance measurement of youth employment programmes allows users to determine performance across individuals, types of programmes and geographical areas. Individuals are young people, usually aged 15 to 24 (or 29 depending on the national definition of "youth") who face (or are likely to face) difficulties in gaining decent and productive work. Performance measurement can be used, for instance, to compare the outcomes of a programme on young men and young women, on teenagers and young adults, on youth with lower and higher levels of educational attainment, and so on.

Youth employment programmes may take a number of different forms. What characterizes them is their labour market objective, namely improvement of employment, employment-related earnings or employability (i.e. the capacity to find and retain a job). Performance monitoring permits comparison of outcomes of different programmes on the same (or similar) groups of participants. For example, the employment rate and earnings of young people who participated in a school-based training programme may be compared with those of young people participating to on-the-job ones.

Asidefromtracking inputs and outputs, performance monitoring measures the gross (short-term) outcomes of a specific programme against its established objectives and targets, where these have been set. The findings of the performance monitoring exercise provide information to improve both the programme management system (i.e. the modalities of implementation) and its outcomes. Performance monitoring activities generally track: easy-to-collect data to determine compliance with programme implementation rules; inputs considered important to the programme's purpose (e.g. participants' characteristics); expected outputs (e.g. participation or completion rates); and short-term outcome measures (e.g. employment and earnings of participants).

Setting up a performance monitoring system for youth employment programmes involves four steps: (i) clarifying programme objectives (discussed in the following part of this module); (ii) identifying performance indicators (Module 3); (iii) setting the baseline and designing targets (Module 4); and (iv) monitoring results (Module 5).

The first step in performance monitoring is to clarify the programme's objectives. One must define precise and measurable statements concerning the results to be achieved in a given timeframe, and identify the means to meet those objectives (i.e. the resources and activities). The formulation of objectives should be done at the design phase.

#### 2.3.1. Design of youth employment programmes

The process that underpins the development of youth employment programmes and its linkages to the youth employment policy cycle are shown in Figure 2.1.

Step 1 in the process is to obtain a clear understanding of the extent of the youth employment challenge. This requires the collection and analysis of statistical data on youth labour supply, labour demand and conditions of work (situation analysis). If this information is not readily available, a survey of the youth labour market should be carried out. An analysis of key youth labour market indicators will serve to identify those factors most likely to influence employment outcomes for young people. This process enables the "profiling" of youth population groups (but also individuals) more likely to be disadvantaged in the labour market.

Step 2 in the process is to review existing policies and programmes on youth employment, and then map the institutional framework that governs the youth labour market – including the roles played by employer and worker organizations. This completes the analysis of the youth employment situation and is instrumental to identify the main challenges to be addressed.

Step 3 in the process is to identify key youth employment problems and establish the relevant cause-effect relationships between them. This process serves to devise ways and means to address them.<sup>8</sup> The cause-effect analysis identifies the root cause of the problem (e.g. a cause that can explain the effect, either directly or through a series of events, and that, if removed, would eliminate or reduce the problem). Once identified, the key youth employment problems can be ranked by priority.

2.3
Formulating the objectives of youth employment programmes

<sup>7</sup> The main sources of youth employment data are national labour force surveys, establishment surveys (on enterprises) and labour-related administrative figures (e.g. social security organizations, public and private employment offices, unemployment insurance schemes and civil service administration). For more information on the data requirements and data sources needed to conduct a youth employment situation analysis, see ILO, 2013 Youth labour market analysis: A training package on youth labour market information (Geneva, ILO).

<sup>8</sup> G. Rosas and G. Rossignotti: *Guide for the preparation of National Action Plans on Youth Employment* (Geneva, ILO, 2008).

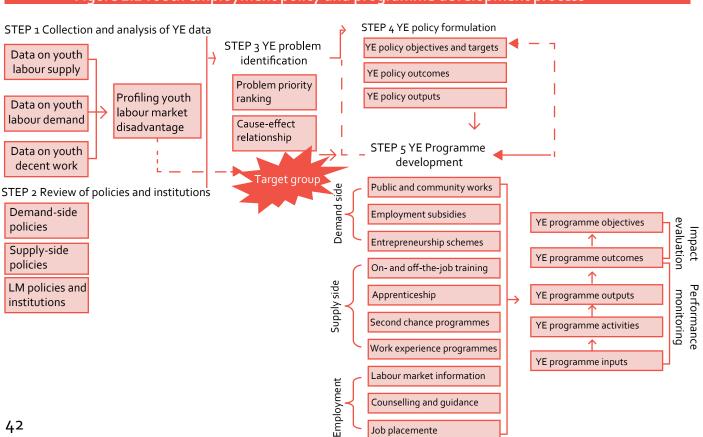
# Performance monitoring of youth employment programmes

Step 4 in the process is to formulate policy options based on the above identification and ranking of key youth employment problems. This involves several stages: (i) the formulation youth employment objectives and targets; (ii) the selection of feasible and affordable policy options to achieve the objectives; (iii) the design of outcomes, outputs and activities; (iv) the planning of the means and resources necessary to achieve the objectives; and (v) the establishment of adequate monitoring and evaluation mechanisms.

Step 5 in the process is to develop youth employment programmes. These are typically outputs of the (youth) employment policy. However, they can also stem from other types of policies (e.g. a poverty reduction or national development policy) or be formulated outside the existing policy framework (e.g. an emergency public work programme designed to provide short-term income opportunities to population groups affected by a natural disaster).

Irrespective of the policies that youth employment programmes form part of, their design builds on a throughout knowledge of the youth labour market and institutional setting (Steps 1 and 2) and the identification of key youth employment problems (Step 3).

#### Figure 2.1 Youth employment policy and programme development process



# Specifically, programme design comprises at the very least four stages:

Analysis of youth employment data to understand the cause-effect relation of youth employment problems (on the supply and demand side);

Profiling of labour market disadvantage (e.g. the identification of groups most affected by the employment problems);

Targeting approaches (e.g. the matching of groups/individuals most affected by the employment problems to the type of programmes most likely to be effective). It is at this stage that the type of programme, duration and compensation levels are determined; and

Formulation of programme objectives, outcomes, outputs and activities and allocation of resources.

# Performance measurement

Programme design

Ideally, the design phase should also indicate the monitoring and evaluation approaches (this would constitute the fifth and final step of programme design), without however, going into detail. The role of performance monitoring is to break down the elements of stage four of the programme design (objectives, outcomes, outputs, activities and inputs) into measurable statements.

# 2.3.2. Formulating objectives of youth employment programmes

In many instances, the objectives of a youth employment programme are implied rather than expressly stated. In such cases, the first task of performance monitoring is to articulate what the programme intends to achieve in measurable terms.

Performance monitoring of youth employment programmes

# >>>> Example`

Consequences
of poorly
formulated
youth
employment
programme
objectives

The government has decided to invest \$3 million in public works for poor youth in regions hit by floods. The programme design specifies: (i) the eligibility criteria for individual participants (youth aged 15–24, without work, living in areas affected by the floods and with a monthly household income below \$75); (ii) the type of works that are eligible for financing (only reconstruction of dwellings destroyed by the floods); and (iii) the average duration of employment for participants (three months) and compensation levels (minimum wage). The programme, however, is silent on the results it intends to achieve.

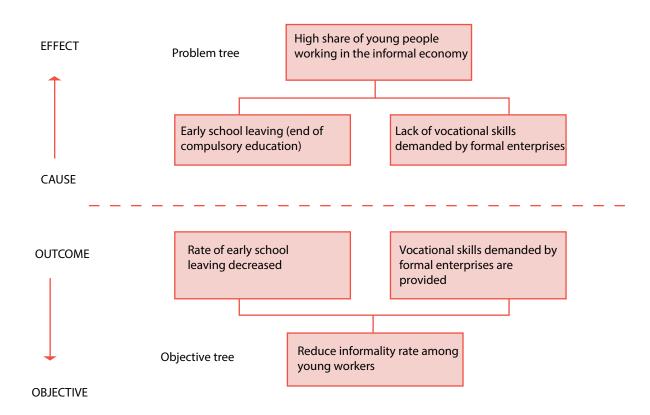
Unless otherwise stated, a performance monitoring system typically uses two measures to gauge the success of employment programmes: (i) the employment rate of participants at follow-up, and (ii) total costs. If the public work scheme were to be judged only against these two measures (with employment being the most important), the results will likely be disappointing. This is because public works normally have a limited effect on the re-employment prospects of beneficiaries and can be expensive.

If the intention is to provide temporary earnings opportunities for young people affected by floods and to reconstruct dwellings, other measures would be more appropriate. These include: (i) the number of person/days of work generated; (ii) the share of households living in reconstructed dwellings (over total dwellings destroyed by the floods); (iii) the consumption levels/household income of participants in the public work scheme (during the programme); and (iv) the share of youth employed in the public work (over total youth affected by floods) and so on.

A useful tool for the development of objectives is the graphic tree diagram. It can help identify the causes and effects of youth employment problems (problem tree) and show how these problems can be addressed (objective tree). Once the main youth employment problems (at national, regional or local level) and their cause-effect relations are clear, the identified problems – the effect of a number of causes – are reversed and become the objectives to be achieved. This process should be as participatory as possible and involve employer and worker organizations, as well as representatives of young people themselves.

**Figure 2.2** provides a simplified example of the relationship between cause and effect identified in the problem analysis, and objectives and outcomes.

Figure 2.2 Problem and objectives trees



The above example identifies a clear cause-effect relationship between work in the informal economy, early school living and lack of skills demanded by formal enterprises. A youth employment programme offering young people vocational training in skills demanded by formal enterprises would contribute to reducing informality. The objective of such a programme could be "provide young people with the vocational skills required for employment in formal enterprises". Two measures, in particular, can be used to judge the success of this programme. The first is the type and level of vocational skills acquired by participants (measured in terms of the pass rate of participants in standardized, competency-based tests). The second is the employment rate of participants in formal enterprises at programme's end.

+ 5

2

# Performance monitoring of youth employment programmes



**FAQ** 

#### How does a performance management system work?

Performance management is a system that tracks programme outcomes and estimate impacts to improve the effectiveness of programme (or policy) interventions.

#### What is performance measurement/monitoring?

Performance measurement is a monitoring system that continuously tracks inputs, activities, outputs and (short-term) outcomes of a youth employment programme on the basis of measures called performance indicators. Performance monitoring of youth employment programmes requires a number of steps: (i) the clarification/setting of programme objectives; (ii) the identification of performance indicators; (iii) the setting of targets; (iv) the measurement of results; and (v) the reporting of findings.

#### How are youth employment programmes designed?

The design of youth employment programmes typically involves five stages. In the first stage, labour market data are screened to identify the cause/effect relations of youth employment problems. The second stage serves to identify those groups among the youth population most affected by employment problems (profiling). The third stage involves the matching of target groups with the features of (available) employment programmes most likely to work for the target group. The fourth stage centres on formulating outcomes, outputs and activities of the programme and allocate human and financial resources. The final stage identifies monitoring and evaluation approaches to measure the achievement of programme outcomes.

# Why should youth employment programmes have clearly articulated objectives?

The objectives of a youth employment programme are the basis for the development of performance indicators. These measurements will be used to track progress and decide whether the programme is delivering the expected outcomes. If the objectives are missing or are unclear, this will be reflected in the indicators. This may lead to important aspects of the programme not being tracked and measured.

# **Aditional reading**

Cockx, B. 1998. The design of Active Labour Market Policies. What Matters and What Doesn't? (IRES and Department of Economics, Université Catholique de Louvain).

**Cockx, B.** 2000. "The design of active labour market policies: Building in effective incentives", in International Labour Review (Vol. 139, No. 4), pp. 457–480.

**Calmfors, L.** 1994. *Active Labour Market Policy and unemployment* – A framework for the analysis of crucial design features, OECD Economic Studies No. 22 (Paris, OECD).

**Grubb, D.** 2004. "Principles for the performance management of Public Employment Services", Public Finance and Management (Vol. 4, No. 3), pp. 352–398.

**OECD-DAC.** 2000. Results based management in the Development Cooperation Agencies: a review of experience. Background report (Paris, OECD), www.oecd.org/dataoecd/17/1/1886527.pdf

You are about to carry out a performance measurement of a new on-the-job training programme for young people with less than lower secondary education. Your agency has been implementing this programme over the last two years. To benchmark the results, they have decided to carry out a similar exercise on a government-managed active labour market measure entitled First Work Experience (FWE). This started at the same time as your programme and targets a similar group. The FWE features are contained in the "Call for Application" advertised in the national media:

- The FWE is available to secondary and university educated youth (aged 18–29), who graduated maximum two months prior to the publishing of the Call for Applications in any educational stream.
- The FWE offers fully subsidized employment with a public or private sector enterprise for a period ranging from six months (for secondary graduates) to 12 months (for university graduates) to gain work experience in the field of study.
- Participants receive net monthly wages ranging from \$100 (for secondary school graduates) to \$150 (for university graduates).

<<<<<

**Exercise** 

What are the objectives?

Performance monitoring of youth employment programmes

## >>>> Exercise

# What are the objectives?

• Partner enterprises are responsible for certifying the skills acquired on the job by participants. There is no obligation on employers to retain young workers once the programme period has expired.

The operational procedures developed by the employment service to manage the FWE do not provide additional information on the programme design. Young people are referred to FWE on a first-come, first-served basis, and there is presently no indication that the programme will be evaluated.

You need to develop performance indicators to measure the performance of the FWE, so as to compare its results with those of the on-the-job programme of your agency. But first, you will have to clarify the programme objectives.

#### >>>>>

# **Check your answer**

Three pieces of information provided in the text characterize this programme: (i) the programme targets young people that have recently graduated (pointing to a lack of work experience as a barrier to labour market entry); (ii) youth are placed in enterprises to practice the skills of their occupation; and (iii) enterprises are required to certify the skills acquired.

Therefore, the programme appears to have two key objectives. The first is to provide young graduates with the work experience required by the labour market. The second is to equip participants with the job skills required to perform the chosen occupation. Two main indicators, therefore, will measure performance, namely: (i) the share of participants employed in the occupation of choice; and (ii) the share of participants certified as competent in their chosen occupation.



# Development of performance indicators for youth employment programmes

- 3.1. Indicators of performance: basic principles
- 3.2. Process indicators
- 3.3. Outcome indicators



Development of performance indicators for youth employment programmes

# Development of performance indicators for youth employment programmes

Learning objectives

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

Develop process and outcome indicators to measure the performance of youth employment programmes.

Learning exercises

Which process indicator?
Identifying performance indicators

Performance indicators are concise quantitative and qualitative measures of programme performance that can be easily calculated and tracked on a regular basis. The design of indicators determines the data to be collected and analysed (see Module 5). Quantitative indicators measure changes in a specific value (number, mean or median) and a percentage. Qualitative indicators provide insights into changes in attitudes, beliefs, motives and behaviours of individuals. Although important, information on these indicators is more time consuming to collect, measure and analyse, especially in the early stages of programme implementation.

Users should develop indicators themselves. As already mentioned, the success of a monitoring and evaluation system is measured not in the quantity of data and information generated, but rather in terms of the use that is made of the information made available. Thus, it is normal to start with a large number of indicators, and to progressively focus only on those that are used for programme design and management.

Performance indicators should be SMART, namely: specific (to measure the information required as closely as possible); measurable (to ensure that the information can be readily obtained); attributable (to ensure that each measure is linked to the programme's efforts); realistic (to ensure that the data can be obtained in a timely fashion, with reasonable frequency and at practicable cost); and targeted (to the programme population).

#### Box 3.1. Tips for the development of indicators

**Relevance**. Indicators should be relevant to the needs of the user and to the purpose of monitoring. They should be able to clearly indicate to the user whether progress is being made (or not) in addressing the problems identified.

**Disaggregation**. 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. For services and/or programmes the disaggregation is normally done by type of service/programme.

3.1 Indicators of performance: basic principles

Development of performance indicators for youth employment programmes

#### Box 3.1. Tips for the development of indicators

**Comprehensibility.** Indicators should be easy to use and understand and data for their calculation relatively simple to

collect.

Clarity of definition. A vaguely defined indicator will be open to several interpretations, and may be measured in different ways at different times and places. It is useful in this regard to include the source of data to be used and calculation examples/methods. For example, the indicator "employment of participants at follow-up" will require: (i) specification of what constitutes employment (work for at least one hour for pay, profit or in kind in the 10 days prior to the measurement); (ii) a definition of participants (those who attended at least 50 per cent of the programme); and (iii) a follow-up timeframe (six months after the completion of the programme). Care must also be taken in defining the standard or benchmark of comparison. For example, in examining the status of young people, what constitutes the norm – the situation of youth in a particular region or at national level?

The number chosen should be small. There are no hard and fast rules to determine the appropriate number of indicators. However, a rule of thumb is that users should avoid two temptations: information overload and over-aggregation (i.e. too much data and designing a composite index based on aggregation and weighting schemes which may conceal important information). A common mistake is to over-engineer a monitoring system (e.g. the collection of data for hundreds of indicators, most of which are not used). In the field of employment programmes, senior officials tend to make use of high-level strategic indicators such as outputs and outcomes. Line managers and their staff, conversely, focus on operational indicators that target processes and services.

**Specificity.** The selection of indicators should reflect those problems that the youth employment programme intends to address. For example, a programme aimed at providing work experience to early school leavers needs to incorporate indicators on coverage (how many among all school leavers participate in the programme), type of enterprises where the work experience takes place and the occupation, and number of beneficiaries that

#### Box 3.1. Tips for the development of indicators

obtain a job afterwards by individual characteristics (e.g. sex, educational attainment, household status and so on).

**Cost.** There is a trade-off between indicators and the cost of collecting data for their measurement. If the collection of data becomes too expensive and time consuming, the indicator may ultimately lose its relevance.

**Technical soundness.** Data should be reliable. The user should be informed about how the indicators were constructed and the sources used. A short discussion should be provided about their meaning, interpretation, and, most importantly, their limitations. Indicators must be available on a timely basis, especially if they are to provide feedback during programme implementation.

**Forward-looking.** A well-designed system of indicators must not be restricted to conveying information about current concerns. Indicators must also measure trends over time.

**Adaptability.** Indicators should be readily adaptable to use in different regions and circumstances.

When choosing performance indicators, it is important to identify indicators at all levels of the results chain of a youth employment programme, and not just at the level of outcomes. Information on process is useful for documenting programme implementation over time and explaining differences across programme sites. Information on individual participants (e.g. sex, age group, national origin, medical condition, educational attainment, length of unemployment spell and so on) allows users to judge compliance with targeting criteria. Some examples of the most common process (also called implementation) indicators, calculation and disaggregation methods are shown in Table 3.1. Data sources, data collection and analysis are discussed in detail in Module 5.

cont.

Source: adapted from Canadian International Development Agency (CIDA), 1997. Guide to Gender-Sensitive Indicators (Ottawa, CIDA).

3.2 Process indicators

# Development of performance indicators for youth employment programmes

Table 3.1. Example of process (implementation) indicators (measurement and disaggregation)					
	Process indicators	Calculation method	Disaggregation		
1	Composition of entrants/ participants/ completers	Number of entrants in period t*100 = Total number of entrants in period t	- by type of programme  - by characteristics   of individuals  Programme: training, subsidy, selfemployment, public work   scheme  Individuals by sex, age group, education level, unemployment duration, type of disadvantage, prior occupation/work experience		
2	Stock variation of entrants/ participants/ completers	Number of entrants in period t = Number of entrants in	As above		
		period t-1			
3	Inflow of entrants (or participants)	Number of new entrants in period t =Stock of entrants end of period t-1	As above		
4	Degree of coverage of target population (entrants/participants /completers)	Number of programme entrants*100 = Total targeted population	As above		
5	Implementation	Number of implemented actions = Number of planned actions	As above		
6	Average cost per entrant/ participants/ completer	Total cost of programme = Total number of entrants	By programme: training, subsidy, self-employment, public work scheme.		

Note: \* Entrants are all individuals who start a specific programme. Participants are all individuals who entered and attended the programme for a minimum period of time (usually determined by the rules of the programme as the minimum period required to produce changes, for example 50 per cent of the programme duration). Completers are those who completed the whole programme. Dropouts, usually, are those who left the programme before a minimum period of attendance established by the rules of the programme (e.g. the difference between entrants and participants).

The indicator in the first row, for example, serves to determine whether the targeting rules of the programme are being complied with. For instance, in a youth employment programme targeting individuals with less than primary education, the share of entrants by this level of education over the total will determine if eligibility rules are being followed and allow tracking of sites with the best/worst compliance. The indicators in the second and third rows serve to measure the evolution of the programme's intake. It is normal, in fact, to see increases in intake as the programme matures. The time t may be any time interval (yearly, quarterly or monthly). The indicator in the fourth row serves to measure the overall coverage of the programme.

Depending on its scope, the denominator can be the total number of youth (in a country, region, province or town) or only those who have certain characteristics (e.g. only those who are unemployed, workers in the informal economy, individuals with a low level of education and so on). The indicator in the fifth row serves to measure the pace of implementation compared to the initial plan, while the indicator in the last row is used to calculate overall costs.

You have been assigned to manage a new programme targeting young people not in education, employment or training (NEET) in a region bordering the European Union. A survey carried out recently shows that the share of youth who are NEET in that region increased from 25 per cent two years ago to 30 per cent (or 20,000 young people) today. The national average is 15 per cent. The programme offers beneficiaries (who volunteer) a voucher that can be spent to (re)enrol in formal education, attend a vocational training course, or obtain subsidized employment in a private company. After the second quarter of implementation, you have to report on progress. Which indicators would better show the progress of the programme, how would you calculate them, and to which level of disaggregation?

#### Exercise

<<<<<

Which process (implementation) indicator?

Development of performance indicators for youth employment programmes

>>>>> Check your answer

The most telling indicator in this instance is the degree of coverage of the target population (total number of programme entrants\*100/total number of NEET in the region), disaggregated by each of the programme options available (school, training and employment) and individual characteristics (especially sex and highestlevelofeducationattained prior to programme participation). This indicator alone offers a number of insights, for example: (i) the option the target group prefers (education, employment or training); (ii) the relation between sex and educational attainment of entrants and selected programme option (e.g. girls may prefer to return to school, while boys prefer training or employment); and (iii) programme uptake (e.g. if only 200 youth received the voucher in the first two quarters, it is necessary to investigate why uptake is low).

#### Two other indicators also warrant attention:

- 1) The first is the stock variation of entrants (number of entrants in the second quarter over the number of entrants in the first quarter, by option available and individual characteristics).
- 2) The second is the average cost per programme entrant (to determine whether the programme is efficient compared to another intervention with the same objective/target group).

To give a simple numerical example, in the first quarter the programme assigned 800 vouchers (300 for school enrolment, 150 for training and 350 for employment) and 1,200 in the second quarter (400 for school, 200 for training and 600 for employment). The degree of coverage at the end of the second quarter is 10 per cent of the target population. Intake in the second quarter increased by 50 per cent compared to the first quarter. If intake maintains the same pace, the entire target group will be included in the programme within 18 months. The preferred option is employment (47.5 per cent of participants), followed by school re-enrolment (35 per cent of participants) and training (17.5 per cent).

Since the overarching objective of youth employment programmes is to help young people get a job, the most significant outcome indicators are: (i) the gross placement (employment) rate by individual characteristics and type of programme; (ii) average cost per young person placed; and (iii) the level of earnings of youth participants employed.

3-3
Outcome
indicators

The more disaggregated the data, the better, as this allows comparison across individuals, programmes and geographical locations. Calculation methods and disaggregation are shown in Table 3.2 below.

	Table 3.2. Outcome indicators (measurement and disaggregation)				
	Outcome indicators	Calculation method	Disaggregation		
1	Gross placement rates (individuals)	Number of placement*100 = Total number of participants (including dropouts)	<ul> <li>by type of programme</li> <li>by characteristics of individuals</li> <li>by type of job</li> <li>Programme: training, subsidy, self-employment, public work scheme</li> <li>Individuals by sex, age group, education level, unemployment duration, type of disadvantage, prior occupation/work experience</li> <li>Jobs by economic sector and size of the enterprise, occupation, contract type, and contract duration</li> </ul>		
2	Earnings	Number of individuals placed in a job and earning (hourly) wages over the minimum*100 =Number of placements	<ul> <li>by type of programme</li> <li>by characteristics of individuals</li> <li>by type of jobs</li> </ul>		
3	Cost per placement	Total cost = Number of placements	– by type of programme		

5

3

Development of performance indicators for youth employment programmes

The above-mentioned disaggregation also allows data users to judge the "quality" of the results achieved.

The use of total placement as an indicator of performance, in fact, has two main shortcomings:

1)The first is the likely prevalence of short-term employment and the likelihood that beneficiaries re-enter unemployment soon after the end of the programme.

2)The second is the lack of distinction between "easy-to-place" youth (who would eventually get a job also without the programme) and "disadvantaged" youth (who are likely to experience long spells of unemployment if they are not helped).

The first issue results in "gaming" behaviour, for example, administrators may be tempted to "cheat" the system by focusing on short-term placement (with no attention to quality) to achieve programme targets (see Module 4). The second gives rise to "creaming" (or cream-skimming), namely the selection for programme participation of those youth most likely to succeed, as compared to those who most need the programme.

The disaggregation proposed in Table 3.2 corrects these shortcomingsbyrequiringcollectionofinformation on the characteristics of individuals employed and the type of jobs they perform. Where possible, disaggregation by work in the formal/informal economy is also useful to measure the attainment of decent and productive jobs. Finally, calculation of hourly wages helps to measure the welfare gains more accurately than total earnings, as young workers may have higher earnings only because they work longer hours. Examples of the performance indicators used in employment programmes implemented in the United States are provided in the example below.

Performance measures were first established by the US Department of Labour in the late 1970s under the Comprehensive Employment and Training Act (CETA). The Job Training Partnership Act (JTPA), the main workforce development programme in the 1980s and 1990s, required comparison of employment and earnings performance to overall costs. The indicators included employment rate in unsubsidized employment, employment retention for six months, increase in earnings and/or the wage rate, reduction in welfare dependency and skills acquisition.

In practice, the measures focused on the labour market status of participants 13 weeks (three months) after the end of the programme. For youth programmes, the common measures use a broader concept of successful outcome by also including training and education (the measures include attainment of a certificate or degree and literacy and numeracy gains). The indicators used are:

# Employment rate:

<<<<

Number of participants employed in the three months following the programme over total participants;

# Employment retention rate:

<<<<

Among those employed, the number of those employed for six months and longer after the end of the programme; and

#### Average earnings:

<<<<

For participants employed for at least nine months after the end of the programme, average earnings in the previous six months.

## Example

<<<<<

United States
Workforce
Development
Programmes

Source:

Barnow, B.S. 2009. The role of performance management in workforce investment programs (Washington, DC, Johns Hopkins University).

3

Development of performance indicators for youth employment programmes



Cost is another important measure: it allows users to decide whether a programme is cost-effective (e.g. whether the rate of return in terms of placement justifies the resources invested). Usually, the overall costs of a programme are compared to those of other programmes with similar objectives and target groups).

#### Overall costs include:

- 1) The disbursements made to service providers (e.g. the payment made to a training centre to conduct a vocational training course) or to other agencies (e.g. the cost of insuring participants during programme participation);
- 2) Payments made to individual participants (e.g. the reimbursement of transport costs incurred to reach the site of training, subsidies for living costs and so on); and
  - 3) The administrative cost of running the programme.

If the programme is assigned to a private provider, administrative costs are included in the service fee paid to the provider and are not calculated separately. When a programme is managed by a specific agency, a simple method of calculating administrative costs for programme implementation is to keep track of how much time staff members usually spend on each participant (treatment hours). The average cost of staff time can be calculated on the basis of payroll data. To approximate the total administrative costs of the programme the unit staff cost is then multiplied by the total treatment hours and the total number of participants to a given programme.

When the programme is managed by the Public Employment Service (PES) it may be difficult to ask the staff to keep track of the time they spend with beneficiaries to calculate administration costs. An alternative way to approximate such costs is to divide all administrative costs of the PES at the end of a given year (excluding unemployment benefits and the disbursements for active labour market programmes) by the total number of unemployed registered to obtain average total administrative costs per individual unemployed (on an annual, monthly and daily basis). This daily amount is multiplied by the average number of treatment days spent on each participant. The total obtained is added to the disbursements made for each participant to approximate the total cost per individual treated (see example below).

g In terms of youth employment programmes, cost-efficiency refers to the simple relation between cost and results (e.g. cost of the programme vs. number of individuals placed). Cost-effectiveness relates also to the quality of placement, and not only to quantity (e.g. cost of the programme vs. individuals placed in "good" jobs).

The Public Employment Service (PES) is implementing an on-the-job training programme targeting youth with a low level of education. The programme has envisaged a monthly compensation for young beneficiaries for three months and a training grant for enterprises (lump sum). The total disbursement for the programme in the last year was €84,000. The total number of participants was 100.

The total number of unemployed registered at the end of last year was 8,400, while the total budget of the PES for the same year was  $\in 18,700,000$  (excluding the amount allocated for the unemployment benefit and active labour market programmes). This means that the average (administration) cost per each unemployed registered was  $\in 2,226$  ( $\in 18.7$  million/8,400) per year, with a daily cost per unemployed of  $\in 6$ . This amount is multiplied by the number of estimated treatment days for each participant (for a programme lasting three months the average is ten days). The disbursement per participant was  $\in 840$  ( $\in 84,000/100$ ), while the administrative costs were  $\in 60$  ( $\in 6*10$  days). The total cost/participant for the on-the-job training programme is  $\in 900$  ( $\in 840+\in 60$ ). If 50 per cent of participants found a job at the programme's end the average cost per individual employed is  $\in 1,800$  (e.g. overall cost of the programme divided by participants employed, i.e.  $\in 90,000/50$ ).

# Example

<<<<<

How to compute PES administrative costs for performance monitoring

The local government of one of the poorest region in the country has decided to invest part of its tourism revenues to promote youth employment. Half of the funds available will be used for wage subsidies, as these programmes yield good results (70 per cent placement rate). The remaining funds (\$5 million) will be invested in training. The authorities are not satisfied with the current training programmes, run mostly by schools and training centres with gross placement rates ranging between 35 and 40 per cent, and want to organize on-the-job training (OJT) in private and public enterprises.

The programme will target unemployed youth aged 20–24, with less than secondary education. The labour market figures available, in fact, show that the employment rate of this group has remained fairly stable in the last five years at 50 per cent, compared to 60 per cent for youth with secondary education and 85 per cent of those with tertiary education.

# Exercise

<<<<

Identifying performance indicators

4

5

3

Development of performance indicators for youth employment programmes

### >>> Exercise

# Identifying performance indicators

cont.

The share of low educated youth in the population aged 20–24 is 55 per cent (or 150,000 young people). Of these, 30 per cent are inactive, 50 per cent are employed and 20 per cent are unemployed (e.g. their unemployment rate is 28.5 per cent).

The programme envisages that partner enterprises will receive a training grant to cover the cost of training (a lump sum of \$150 per trainee), while participants will receive a sum equal to 50 per cent of the monthly minimum wage (\$100) for the training period (maximum six months). The objective of the programme is to promote employment opportunities for low-educated youth by providing them with the skills demanded in the local labour market.

The authorities ask your advice in designing performance indicators for this new programme.

#### >>>>

## Check your answer

To monitor implementation, the following process indicators are relevant (see Table 3.1):

- Composition of individual entrants/participants/completers (by sex, education level, length of unemployment spell, other labour market barrier);
- Composition of partner enterprises (by size, ownership and economic sector);
- Coverage of target population (number of entrants and participants as a percentage of the total eligible population by sex, education, length of unemployment spell, other labour market barriers);
- OJT entrants/participants by occupation (by individual characteristics);
- ✓ OJT entrants leaving the programme prior to completion by individual characteristics; and
- ✓ Average monthly cost per programme entrant and participants.

# Check your answer

<<<<<

To measure results, the following outcome indicators could be considered (see Table 3.2).

cont.

- ✓ Share of participants employed in regular, non-subsidized jobs (by individual characteristic and type of jobs), including:
  - the percentage employed in the same company of training;
  - the percentage working in the occupation of training; and
  - the percentage using the skills learned during the programme.
- ★ Earnings (percentage of participants employed earning hourly wages above the minimum) by individual characteristics and type of job; and
- Average cost per participant employed (calculated as total disbursement for the training grant, plus total disbursement made to participants, plus administrative cost divided by number of participants employed).

As outcome indicators depend on the specific objectives of each youth employment programme, each programme has its own. Some examples of outcome indicators for the most common employment programmes are provided in Table 3.3. When selecting outcome indicators, attention must be paid to the costs and time involved in collecting the data needed (see Module 5). For this reason, it is always better to focus on a few, but well-chosen indicators.

Development of performance indicators for youth employment programmes

Table 3.3. Examples of performance indicators for youth employment programme				
Vocational training				
	= number of individuals who complete the training programme/number of entrants*			
Completion rate	= number of individuals who passed standardized testing at the programme's end/number of entrants			
Graduation rate	= number of individuals who left the course in the first (30,60, 90) days of programme/number of			
Drop-out rate	entrants * For training programmes, it is necessary to distinguish between those who entered the course (entrants) and those who attended a minimum period (participants). In some programmes, the term "completers" is used to denote those who complete the whole programme.			
Proportion of participants in regular (unsubsidized) employment at follow-up, including:	= number of participants employed at follow-up/number of entrants/participants			
<ul> <li>(For on-the-job training): share of trainees employed in same enterprise offering training;</li> </ul>	= number of employed trainees in same enterprise/number of trainees employed			
(For all types of training): share of trainees employed in the occupation of training; and	= number of trainees employed in occupation of training/number of trainees employed			
(For all types of training): share of trainees using skills acquired during the training.	= number of trainees employed who use skills learnt/number of trainees employed			
Average earnings	= total earnings of trainees employed/number of trainees employed			
	= number of trainees employed earning hourly wages over minimum/number of trainees employed			
Average cost per participant/completer	= total cost /number of participants/completers			
Average cost per participant/completer employed at follow-up	= total cost /number of participants/completer employed			

Employment subsidy		
Proportion of subsidized workers (participants) in regular employment at follow-up, including:	<ul><li>= number of subsidized workers employed at follow-up/number of participants</li><li>= number of subsidized workers employed</li></ul>	
Share of subsidized workers still employed at follow-up in partner enterprise	at follow-up in partner enterprise/number of subsidized workers employed at follow-up	
Average earnings	= total earnings of subsidized workers/number of subsidized workers employed at follow-up	
	= number of subsidized workers earning hourly wages over minimum/number of subsidized workers employed at follow-up	
Average cost per subsidized worker employed at follow-up	= total cost of subsidy/number of subsidized workers employed at follow-up	
Average cost of subsidy per subsidized worker	= total cost subsidy/number of participants	
Self-employ	ment assistance	
Proportion of persons still self-employed at follow-up	= number of self-employed at follow-up/number of participants	
Average earnings	= total earnings of self-employed/ number of individuals still self-employed at follow-up	
	= number of self-employed earning incomes over the minimum/number of individuals who are still self-employed at follow-up*	
	* to compare self-employment earnings, one can use either the level of the statutory minimum wage, or the average earning for self-employed, if available (this is usually calculated by the statistical office)	
Average cost of assistance per person still self-employed at follow-up	= total cost of assistance/number of self- employed at follow-up	
Average cost per participant	= total cost of assistance/number of participants	
Average added employment generated by assisted self-employed	= number of additional jobs created (individuals employed) by self-employed individuals assisted by the programme	

Development of performance indicators for youth employment programmes

Public (community) works		
Proportion of public work participants in regular employment at follow-up	= number of public work participants who are employed at follow-up/number of participants	
Average earnings	= total earnings of participants employed at follow-up/ number of employed at follow-up  = number of participants employed earning hourly wages over minimum number of participants employed at follow-up	
Average cost per participant employed	= total cost for public work/number of participants employed at follow-up	
Average cost of public work	= total cost public work/number of participants	

The outcome indicators listed for public (community) works are only indicative. As these programmes typically aim at improving living standards/welfare of household and communities, the selection of indicators is determined on a case-by-case basis (see example below). Often, public works programmes rely more on process rather than on outcomes indicators (Table 3.4).

Table 3.4. Examples of process indicators for public (community-based) works		
Person-days of employment generated	= number of workers engaged in public work (community-based) project *number of days each person worked	
Earnings of individual participants (during public work)	= daily wage *days of work in public work	
Training days	= number of individuals who attended training *number of days of training (one day of training amounts usually to six hours of training)	

In 2008, the Government of Liberia launched a Cash for Work programme with the support of the World Bank. The programme aimed to mitigate the negative impact of the economic crisis on vulnerable households and communities. It provided 17,000 vulnerable individuals with temporary employment opportunities. The objective was to improve the standard of living and health of household and community members.

The programme used two process indicators to monitor programme implementation. The first is the number of person-days of employment generated. The second is the number of communities that received improved infrastructure.

Outcome indicators were formulated separately for households and communities. For households, the measures included increases in expenditures for school, food, health care, farm and asset investments. For communities, the measures included decreases in: (i) the incidence of malaria and cholera; and (ii) the travel time to the nearest market, school and clinic.

Source: World Bank, 2010. Evaluating impact: *Turning promises into evidence. Liberia Yes Project*, (Washington, DC, World Bank), http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1256762343506/6518748-1275681412812/7130285-1275925241249/Impact\_Evaluation\_of\_Liberia\_YES\_Project.pdf

# Example <

Outcome indicators for public works programmes: Liberia YES Project

<<<<

## **Additional reading**

**Barnow, B.S.** 2009. The role of performance management in workforce investment programs (Washington, DC, Johns Hopkins University).

**Canadian International Development Agency (CIDA).** 1997. *Guide to Gender-Sensitive Indicators* (Ottawa, CIDA).

Card, D., Ibarrarán, P. and Villa, J.M. 2011. Building in an evaluation component for active labor market programs: A practitioner's guide (IZA working paper No 6085).

**Cockx, B.** 2000. "The design of active labour market policies: Building in effective incentives", in International Labour Review (Vol. 139, No. 4), pp. 457–480.

**Gertel, P.J. Martinez, S. et al.** 2011. *Impact evaluation in practice* (Washington, DC, World Bank).

**Grubb, D.** 2004. "Principles for the performance management of Public Employment Services", Public Finance and Management Vol. 4, No. 3), pp. 352–398.

**1 2 4 9** 

Development of performance indicators for youth employment programmes

# >>>>> Additional reading

cont. labour

**O'Leary, C.J.** 1995. "Performance indicators: A management tool for active labour programmes in Hungary and Poland", in International Labour Review (Vol. 134, No. 6), pp. 729–751.

**O'Leary, C.J., Nesporova, A., Samorodov A.** 2001. *Manual on evaluation of labour market policies in transition economies* (Geneva, ILO).

**Schochet, P.Z., Burghardt, J.** 2008. "*Do Job Corps performance measures track program impacts?"*, in Journal of Policy Analysis and Management (Vol. 27, No. 3).

**World Bank, 2010.** *Evaluating impact*: Turning promises into evidence. Liberia Yes Project, (Washington, DC, World Bank)

**Zall Kusek, J., Rist, R.C.** 2004. *Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners* (Washington, DC, World Bank).

## Glossary

### Gaming

>>>>

Gaming describes a situation where the specification of performance may lead to negative incentives for programme providers. For example, if the measure "total employment/placement rate at follow-up" does not disaggregate by type of placement, it may lead providers to focus on short-term jobs (the easiest to find) in order to achieve a high outcome. Gaming normally occurs in systems where good performance is rewarded (e.g. higher funding allocation) and poor performance is sanctioned.

#### **Creaming (or cream-skimming)**

"Creaming" describes a situation where programme administrators/ managers select, from a larger pool of eligible participants, only those who have the best chance to succeed. This may occur when standards are unadjusted and the system is based on incentives (e.g. good performance is rewarded and poor performance is punished). More generally, "creaming" occurs when the individuals selected to participate in a programme are those more likely to have positive labour market outcomes (thus nullifying the purpose of the programme). For example, in the case of an on-the-job training programme, programme administrators might select young individuals with vocational education attainment, rather than youth with primary or less than primary education. Establishing clear targeting criteria in the programme implementation protocol can offset creaming (i.e.

# Glossary <

clearly defining the individual characteristics of those who can participate). Any deviation from these criteria will establish the level of "creaming" that occurred during programme implementation.

cont.

#### Both process and outcome indicators should be developed for youth employment programmes.

Information on processes is useful for documenting programme implementation over time, explaining differences across sites, and understanding how targeting rules were complied with. Such information informs programme managers on whether adjustments are needed. Process indicators are collected and analysed regularly (they determine the "monitoring" part of the performance system).

Outcome indicators are collected and analysed at specific points in time (usually at the end of the programme). They serve to measure the short-term results of a youth employment programme and, if carefully designed, can provide invaluable information to managers on ways to improve the efficiency and effectiveness of programmes.

It is important to include in the disaggregation those variables that will minimize gaming and creaming practices.

Lack of disaggregation of process and outcome indicators may lead to gaming and creaming practices. Disaggregation of outcome indicators by individual characteristics (e.g. educational attainment, prior work experience, length of unemployment spell) provides indications of how the programme served the most disadvantaged groups. Disaggregation of the placements achieved by type of job and average hourly earnings will minimize gaming incentives.

Measurements of efficiency (costs) are important, but too much emphasis on efficiency can divert attention from other outcomes.

Performance across programmes targeting the same population group is measured on the basis of average cost per individual placed. This measure alone, however, does not reflect the welfare gains of participants. For example, a programme may yield better quality jobs for participants, but at higher costs than another programme where placement occurs mostly in economic sectors with a high staff turnover or poor conditions of work. The decision to allocate resources to the cheaper programme, based on pure efficiency considerations, would, in the long term, decrease welfare gains for youth.



5

3

Development of performance indicators for youth employment programmes



**FAQ** 

#### What are performance indicators?

Performance indicators are concise measure of programme performance.

#### How are performance indicators designed?

First, clearly articulate (or identify) the objective(s) of the programme to be monitored. Once these are clear, it is possible to develop measures that will help to ascertain whether the programme is achieving the expected results. For example, an employment programme aimed at improving the skills of low educated people will require an indicator on the (re) employment rate of participants (namely, how many participants are employed following participation in the programme), as "employment" is the overall aim of intervention. Another indicator will be required to measure the skills acquired (this can be done by calculating how many participants attained a vocational training certificate).

#### Is there a minimum/maximum number of indicators per programme?

There is no "appropriate" number of performance indicators. It is important to have indicators at all levels of the results chain (input, activities, outputs and outcomes). But those that receive the most attention are those that show how the programme is developing (process indicators) and the results it achieved (outcome indicators). Two considerations will determine the requisite number of indicators: (i) the objectives of the programme, and (ii) the resources needed to collect and analyse the data required by the indicators. For process indicators, the data sources are mostly figures drawn from the administration of the programme itself. For outcome indicators, however, data may need to be collected through surveys, which increases costs considerably.

#### What is the difference between process and outcome indicators?

Process or implementation indicators measure a programme's inputs, activities and outputs. They are analysed to inform management as to whether the programme is on track. They verify that the inputs of the programme lead to the planned activities, the implementation of which should produce the intended outputs. Outcome indicators measure the short-term results that the programme achieves as a result of the production of outputs. They are measured at specific points in time to provide a snapshot of results achieved at that date. For example, in the case of a second chance programme targeting unemployed youth, process indicators would measure the amount of human and financial resources invested to produce a certain number of second chance courses (inputs and activities). These activities result in a number of young unemployed with incomplete primary education attending second chance courses (outputs). In this case, outcome indicators would measure: (i) graduation rates; (ii) employment rates; and (iii) the average wages of employed participants; and (iv) average costs per each graduate and per each employed individual.

### Can performance measures be used as proxy for impact?

Using programme outcomes to measure programme impact could lead to a faulty assessment of programme effect. However, the link between measured performance and long-term impact could be improved in two ways. First, measures could be adjusted for the characteristics of individuals served in different locations. Second, measurements of performance at multiple periods of time (e.g. at 6, 12, 24 and 48 months after programme participation) could strengthen the connection between short- and longer-term outcomes. However, such surveys may be unfeasible due to the costs of tracking programme participants years after the intervention and obtaining sufficient survey response rates.





# Setting performance targets and standards

4

Setting performance targets and standards

4

# Setting performance targets and standards

Learning objectives

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

- Formulate performance targets for youth employment programmes; and
- Select adjustment methods for performance standards.

Learning exercises

 Design of performance targets for a youth employment programme Two of the most problematic issues in performance monitoring are the definition of appropriate targets for the outcomes and adjustment methods to allow comparison across programmes and geographical areas. For example, the information that participants have an average 50 per cent employment rate in the year following programme participation has no value as a measure of performance per se; it needs to be compared with a benchmark.

Outcomes can be measured against *targets* (i.e. particular values specified for an indicator to be accomplished in a specific timeframe) or *standards* (values that define acceptable performance for a specific programme, usually constructed on figures stemming from the implementation of prior programmes).

Targets and standards can be formulated on the basis of data and estimates of programme performance in previous periods, and on the basis of negotiations among stakeholders (e.g. programme managers and service providers). For example, if the average placement rate of young participants in an on-the-job training programme over the last two years was 56 and 58 per cent, respectively (an annual increase of 2 percentage points), setting a 60 per cent target for the following year seems reasonable (i.e. 58 per cent plus 2 percentage points), if economic conditions are stable. If stakeholders agree on the value of the target, this becomes the benchmark against which the performance of on-the-job programmes will be measured. This method to *benchmark* outcomes is feasible when a programme has been running for some time, and is providing information on performance trends.

Performance standards are typically found in public-funded employment programmes, where decentralized public employment offices, as well as private providers, compete for limited public resources on the basis of a reward-sanction system. In the United States, for example, Job Corps Centres are rewarded with more resources if they reach or exceed the targets established. Those centres that do not perform to standards are sanctioned (less funding in the period following poor performance). An example of performance monitoring conducted by simply comparing the outcomes of different employment programmes is provided below.

**4.1**Setting performance targets)

### >>>> Example

Active labour market programmes in Ireland (2002) In 2002, the Department of Enterprise Trade and Employment of Ireland commissioned a review of the Government's active labour market programmes (ALMPs). The aim was to rationalize programmes and identify the interaction of different policy interventions from the perspective of their cumulative impact on the target groups. The review, therefore, examined the features of the various programmes offered; the eligibility criteria set; targeting approaches; and employment progression rates (unadjusted gross placement rates one year after participation). The review presents outcome data in three ways. First, it examines the evolution of gross placement rates in the period 1992–2000 for five programmes: Specific Skills Training (SST); Job Training Scheme, Community Training Workshop (CTW); Community Youth Training (CYT) and Community Employment (CE).

A. Emp	loyment ra	te of partici	pants (%) <b>,</b> 19	992-2000	
Scheme	1992	1994	1996	1998	2000
SST	56	60	75	72	75
Job Training Scheme	67		74	67	84
CTW	23	27	37	38	56
СҮТР	22	41	54	57	64
CE	14	18	36	31	44

Second, it reports the gross placement rates of different programmes (for all participants) at follow-up by type of job gained (full and part time, permanent and temporary work)

B. First j	ob aft	er program	me (% al	l participar	nts), 2001		
Scheme	%	Job within 3 months	Jobs within 1 year	FT Perm	FT Temp	PT Perm	PT Temp
SST	91	86	97	69	17	3	11
Traineeship	70	75	93	20	25	31	24
CYT	67	81	98	61	22	6	11
Local training initiative	73	75	97	57	31	3	9
Enterprise training	76	79	96	54	31	4	11
CTW	90	80	92	70	19	7	4
Skills foundation	57	70	100	52	19	14	14
Job training scheme	85	91	97	74	13	13	o
Linked work experience	75	85	98	60	15	13	12
CE	46	77	94	37	11	26	25
Job initiative	46	71	81	60	20	10	10
Total	63	79	96	50	20	15	15

Finally, it disaggregates the gross placement rates by previous employment register status (short-term unemployed and long-term unemployed).

C. Employment employment			
SST	54.7	39-3	51.7
Traineeship	73.8	61.1	70.8
Enterprise training	35.1	57.1	39.1
Skills foundation	46.5	36.3	43.8
Total	53.8	41.9	51.2

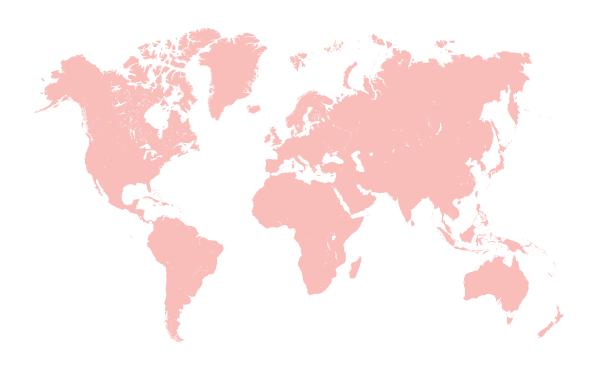
4

The considerations that can be drawn from the tables above are as follows:

First, the gross placement rates of the listed programmes have fluctuated over the years. The programme that has most improved in terms of employment outcomes is the Community Youth Training Programme;

Second, taking the totals of Table B as the average, it is possible to rank those programmes that are above (good performers) and below (poor performers) average;

Third, the same can be done with Table C, where programme effectiveness for long- term unemployed can be ranked below and above the mean total (with Traineeship being the best performer for long-term unemployed).



The setting of targets for new programmes (e.g. programmes with a new design or that have never been implemented in the target areas, or for a specific target group) comprises four steps:

**Set the baseline.** Without a baseline – that is, the value of an indicator just before implementation begins – it is difficult, if not impossible, to establish a realistic performance target. Baseline data are derived from outcomes and indicators. They may be established using existing secondary data sources or may require the collection of primary data. Since youth employment programmes are geared to improve the employment and earnings prospects of the target group, the availability of these figures before the start of the programme, is essential for monitoring and evaluation. A detailed discussion of how to conduct baseline surveys to gather baseline data is provided in Module 5.

Identify trends. This allows users to understand the historical trends in the indicator value over time (e.g. whether available data show changes, either upward or downward over time). Targets should reflect these trends plus the value that a programme is expected to add. For example, if the employment rate of youth with low educational attainment has remained at 40 per cent for the last five years, a reasonable target can be set at a value above that.

Review research findings. There is a large body of literature on active labour market programmes targeting youth (design and implementation rather than impact measures). Reviewing research findings may be helpful in setting realistic targets, especially for the most common programme (e.g. vocational training programmes). For example, the experience of various countries on training programmes targeting youth shows a gross employment rate at follow-up ranging between 50 and 65 per cent of total participants.

Benchmarking. An increasingly popular way of setting targets is to use the results of similar high-performance programmes. Also, targets can be set against the placement rate of the programme across different types of participants or the placement rate of the programme across different geographical areas. For example, if the programme targets both adults and young people, the performance of the programme targeting youth can be benchmarked against that of adults. Conversely, if the programme is implemented in multiple locations for the same target group, placement rates can be compared across sites.

Source: Indecon International, 2002. Review of Active Labour Market Programmes (report commissioned by the Department of Enterprise, Trade and Employment of Ireland), www.djei.ie/publications/labour/2004/reviewactivelabmarket.pdf

## Setting performance targets and standards

4

### >>> Exercise

Design of performance targets for a youth employment programme<sup>10</sup> The local government of one of the poorest region in the country has decided to invest part of its tourism revenues to promote youth employment. Half of the funds available will be used for wage subsidies, as these programmes yield good results for youth (70 per cent placement rate). The remaining funds (\$5 million) will be invested in training. The authorities are not satisfied with the current training programmes, run mostly by schools and training centres with gross placement rates ranging between 35 and 40 per cent. Instead they want to organize on-the-job training (OJT) in private and public enterprises.

The programme will target unemployed youth aged 20–24, with less than secondary education. The labour market figures available, in fact, show that the employment rate of this group has remained fairly stable in the last five years at 50 per cent, compared to 60 per cent for youth aged 20–24 with secondary education and 85 per cent of youth with tertiary education. The share of low educated youth in the population aged 20–24 is 55 per cent (or 150,000 youth). Of these, 30 per cent are inactive, 50 per cent are employed and 20 per cent are unemployed (e.g. the unemployment rate for this group is 28.5 per cent).

The programme envisages that partner enterprises will receive a training grant to cover the cost of training (a lump sum of \$150 per trainee), while participants will receive a sum equal to 50 per cent of the monthly minimum wage (\$100) for the training period (maximum six months). The objective of the programme is to promote employment opportunities for low-educated youth by providing them with the skills demanded by the local labour market.

The authorities ask your advice to design **realistic targets** for this new programme.

10 This exercise builds on the same situation used for the development of outcome indicators in Module 3. The outlook is repeated for the convenience of readers.

### Check your answer

<<<<<

Baseline data are the labour market figures reported in the text, presumably stemming from the Labour Force Survey or a similar household-based survey. The target group has shown a stable employment rate of 50 per cent in the last five years. This is the lowest possible value for the target to be developed. The text indicates that a good "achiever" in the region is the wage subsidy scheme (70 per cent placement rate). This may be taken as the highest possible value of the target. However, the wage subsidy may also involve youth with secondary and tertiary educational levels, which may influence the overall placement rate upward. Since participants to the OJT are low educated youth, it is unlikely to yield the same type of results. Benchmarking with similar programmes is not possible in this case, as the prior training programmes achieved lower placement rates compared to the labour market rate.

If benchmarking with other programmes implemented at national level is not feasible (the text is silent on this), a review of research findings is called for. \*The Irish example, provided above, shows that on average, placement rates stand at around 63 per cent. The US job training partnership programme averages 62 per cent. Other experiences show that enterprise-based training programmes targeting disadvantaged youth typically yield gross placement rates at around 60 per cent. Hence, a range target (rather than a fix value) may be appropriate for the first cycle of implementation (a 55 to 60 per cent placement rate). The 55 per cent stems from the current employment rate of youth aged 20–24 with less than secondary education (e.g. 50 per cent) plus 5 percentage points that represent the value that is expected to be added by the programme. The 60 per cent is the rate achieved by similar programmes conducted elsewhere around the world with the same typology of target group.

The **benchmarking method** for setting performance targets, however, needs to be adjusted to take into account the characteristics of participants, local conditions and other factors that may influence programme outcomes, and which fall outside the control of programme managers. For example, a geographical area that has a high unemployment rate can be expected to have a lower placement rate compared to an area where unemployment is lower. If the performance standard does not use an adjustment method and the placement rates of participants are simply

Note: \*If a similar programme is implemented in other regions of the country, it is possible to use statistical methods to control for differences in participants' characteristics, local conditions and other factors that may have an influence on outcomes. This will help arrive at a target that may be realistic

# Setting performance targets and standards

4

compared at the mean, performance measurement may provide a distorted picture of a programme's outcomes. The same applies to the individual characteristics of participants. For example, the simple benchmarking of the outcome of a programme for low educated youth against those attained for highly educated young people may give disappointing results. This is because youth with low education typically face more difficulties in finding employment compared to individuals with higher educational attainment.

A properly designed adjustment methodology provides a level playing field for the comparison of regional performance or performance across individuals, discourages creaming, and encourages the targeting of programmes to the most disadvantaged individuals. The most common adjustment method is to develop a system of weights based on a national departure point. The weights are the amount by which different sites (or groups) deviate from the national average in the value of "adjustment factors". These factors include demographic data on young people and local labour market indicators (e.g. the unemployment rate, the index of average monthly earnings, the percentage of individuals with low education, population density and so on).<sup>11</sup>

11 For an example of how to develop adjustment weights, see: O'Leary, C.J., Nesporova A., Samorodov, A. 2001. Manual on evaluation of active labour market policies in transition economies, op.cit.

### >>>> Example

# United States Job Corps performance measurement system

The Job Corps performance measurement system comprises measures, standards and weights. Measures are the specific student outcomes to be achieved, as follows:

**Programme achievement:** measures include reading gains, math gains, the rate of attainment of a General Education Development (GED) degree, and the vocational completion rate.

**Placement:** measures include the placement rate, the average wage at placement, and the percentage of quality placements (defined as the percentage of placements in jobs that matched the area of training).

The following table lists measures and describes the pool of students whose outcomes are counted for each measure. Standards are the specific levels of each measure that a Job Corps Centre is expected to achieve. For most measures, the national office sets a uniform standard that applies throughout the United States. For two measures – the GED attainment rate and the average wage at placement – a regression model is used to establish each Centre's standard. This is done to "level the playing field" in situations where Centres face circumstances that could affect the outcome but cannot be controlled. For example, the

wage model adjusts for differences in prevailing wages for entrylevel workers in different locations, and the GED model adjusts for differences in state laws governing the GED awarding. Weights determine how each Centre's score on the various measures are combined.

Examp	e	<<<<
LAGITIP		, , , ,

cont.

	Job Corps Centre per	formance measurement s	ystem	
Measure	Pool	Measure	Standard	Weight
Programme accon	nplishments			
Reading gains	Scored < than 8.5 on reading test at programme entry	% of students in pool who gain two grades or score 8.5 on follow-up test	30%	.056
Math gains	Scored less than 8.5 on math test at programme entry	% of students in pool who gain two grades or score 8.5 on follow-up test	33%	.056
GED rate	Without high school diploma and scored 6.3 or above on reading test	% of students in pool who obtain GED/high school degree	Model	.056
Vocational completion rate	Stayed at least 60 days and participated in vocational programme	% of students in pool who complete vocation at completer or advanced	56%	.167
Placement —				
Placement rate	All teminees plus Job Corps advanced training or career training transfers	% of students in pool placed in jobs or school with bonus for transfers	69%	.111
Average wage	Students placed in a job	Average wage	Model	.111
Quality placemten	Vocational completers with a placement record	% placed in a job- training match	51%	.111

4

Setting performance targets and standards

### >>>>

### **Additional reading**

**Barnow, B.S. 2009.** The role of performance management in workforce investment programs (Washington D.C., Johns Hopkins University). Canadian International Development Agency (CIDA). 1997. Guide to Gender-Sensitive Indicators (Ottawa, CIDA)

**Card, D., Ibarrarán, P. Villa, J.M.** 2011. Building in an evaluation component for active labor market programs: A practitioner's guide, (IZA working paper No. 6085).

**Gertel, P.J. Martinez, S. et al.** 2011. *Impact evaluation in practice* (Washington, DC, World Bank).

Indecon International. 2002. Review of Active Labour Market Programmes (Report commissioned by the Department of Enterprise, Trade and Employment of Ireland) www.djei.ie/publications/labour/2004/reviewactivelabmarket.pdf

**OECD-DAC. 2000.** Results based management in the Development Cooperation Agencies: a review of experience. Background report (Paris, OECD), www.oecd.org/dataoecd/17/1/1886527.pdf

**O'Leary, C.J. 1995.** "Performance indicators: A management tool for active labour programmes in Hungary and Poland", in International Labour Review (Vol. 134, No. 6), pp. 729–751.

O'Leary, C.J., Nesporova, A., Samorodov, A. 2001. Manual on evaluation of labour market policies in transition economies (Geneva, ILO).

**Schochet, P.Z., Burghardt, J. 2008.** "Do Job Corps performance measures track program impacts?", in Journal of Policy Analysis and Management (Vol. 27, No. 3).

**Zall Kusek, J., Rist, R.C., 2004.** *Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners* (Washington, DC, World Bank).

### Adjust performance standards to take account of programme goals, participant characteristics and environmental conditions.

Performance is generally a function of many factors, so it is likely that a programme carried out in different locations, or targeting different groups, will have different performance. For example, entry-level wages may differ from site to site, as they are determined by local economic conditions (including the cost of living). Individual characteristics are also a powerful factor in determining employment outcomes after participation. A programme targeting low-skilled youth may show different outcomes for youth coming from the national majority compared to youth of minority groups.

### Different programmes do not need to have the same performance measures or standards.

Programmes often differ in significant ways, and there is no reason why programmes with different participants, activities and/or implemented under different economic conditions should have identical measures. For example, placement in an unsubsidized job is not a good measure of performance for a programme that provides community service opportunities to poor youth who would like to work.

### Be cautious in establishing performance measures with large rewards and/or sanctions.

There is strong evidence that programmes with reward and sanction mechanisms may lead managers to cream-skimming and reducing services to those most in need with a view to achieving the standard set and gaining the reward.

Including efficiency measures can be useful, but too much emphasis on efficiency can lead to the exclusion of individuals who require expensive services.

Research on the use of cost measures indicates that these may lead to deleterious consequences. Efficiency is, however, a very important goal, particularly when only a small fraction of those eligible for programmes can be served at current budget levels.



### >>>> Glossary

### **Target**

A target specifies a particular value for an indicator to be achieved by a given date in the future. It describes the intended achievement of the programme in concrete terms. Once this is set, it becomes the standard against which programme performance will be judged. An example of a target is "40 per cent employment rate of targeted youth one year after programme's end".

#### Standard

A standard is a target that specifies acceptable performance to be achieved in all sites where a programme is implemented. An example of a standard is "the programme achieves at least 50 per cent placement rate".

#### **Baseline**

A baseline establishes the value of each indicator at a programme's start. Contrary to a situation analysis, which aims at detecting youth employment problems, their cause-effect relation and the groups most affected by the problems, a baseline is geared to measure the specific situation of the targeted individuals in relation to the specified outcome indicators, before programme's start. The figures of the baseline serve to construct targets that will indicate in which direction the outcome indicator should move.

### Benchmarking

This method is used to set targets by using the results of high-performance programmes as a yardstick. The yardstick can also be the performance gained by the programme for other target groups or in other geographical areas.

### Adjustment methods

When benchmarking is used to set targets or standards, adjustment methods are used to account for those factors known to influence outcomes outside the control of programme managers (e.g. differences in the individual characteristics of a target group or the level of development of different geographical locations). Each factor is assigned a weight (constructed on the difference between the national average and the situation of the individual or site). This is used to level out the difference the factor is likely to have on the final outcome.



# Monitoring of results: Data gathering and analysis

5

- 5.1. Building the baseline
- 5.2. Measuring the results of a youth employment programme
- 5.3.Performance measurement of youth employment programmes: a practical example

4

Monitoring of results: Data gathering and analysis 5

# Monitoring of results: Data gathering and analysis

Learning objectives

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

Build a baseline to monitor outcome indicators;

Select sources and collection methods for baseline and follow-up data; and

Measure the results of a youth employment programme.

Learning exercises

None

The first measurement of a performance indicator is the baseline. A performance baseline is information – quantitative and qualitative – that provides data on the outcome indicators directly affected by the programme at the beginning of the monitoring period. A baseline is different from a situation (youth labour market) analysis conducted to design youth employment policies and programmes. A situation analysis draws correlations between the characteristics of individuals and their labour market outcomes to identify the groups that need to be targeted. The baseline collects information on the current situation of individuals in the target group relative to the outcome(s) to be achieved.

**5.1**Building the baseline

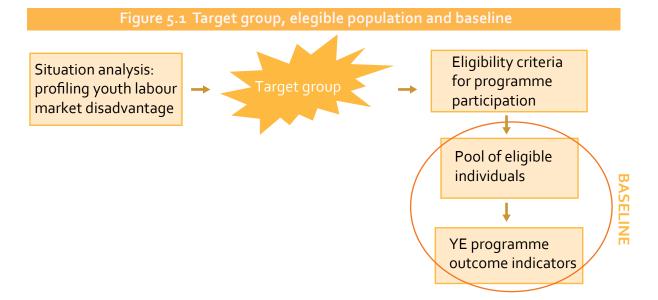


Figure 5.1 illustrates the sequence leading from the situation analysis to the identification of the target group. This serves to define the eligibility criteria for the youth employment programme. Once the pool of the eligible population group has been identified, the outcomes and indicators of the programme guide the collection of information for monitoring. If the programme also envisages an impact evaluation, the evaluation approach selected also determines its data requirements (see Part 2 of this learning package). An example of how this process is carried out in practice is provided below.

4

### Monitoring of results: Data gathering and analysis

5

### >>>>> Example

In search of more and better jobs in Serbia: profiling youth labour market disadvantage

In 2007, the ILO project Youth Employment Partnership in Serbia (YEPS) carried out research on Labour Force Survey data. The twofold aim was to gather quantitative and qualitative information on the transition from school to work of young Serbians (aged 15–29) and provide a detailed situation analysis for the development of the youth employment policy and action plan.

This research found that low educational attainment was the most important factor in predicting poor labour market outcomes. In fact, youth with primary education or less were more likely to be inactive and out of school, unemployed (and for longer periods) or employed in temporary and casual forms of work, compared to young people with secondary and tertiary education. The research also found that young people belonging to Roma groups, and internally displaced and refugee individuals had lower educational attainment than the general youth population. This placed them at a higher risk of unemployment and work in the informal economy.

The findings of the youth situation analysis were used to determine the group to be targeted by active labour market programmes and to define eligibility criteria for programme participation. The data of the unemployment register maintained by the National Employment Service (NES) were screened to identify the geographical districts where unemployed youth with low educational attainment had the highest share among registered youth. To further target the youth employment programmes, the length of unemployment spell (minimum three months) and lack of prior work experience were used to screen programme entrants.

The baseline for the implementation of the youth employment programmes was built on the data of the employment service register and on figures of the LFS on youth employment (economic sector, type of work, level of earnings). The indicators designed to measure the labour market outcomes of young participants were:

Employment at follow-up (by occupation, economic sector, type of contract and formal/informal work); and

Average monthly earnings (below/above the average wage).

**Source**: Krstic G., Corbanese V. 2009. In search of more and better jobs for young people in Serbia (Geneva, ILO), www.ilo.org/budapest/what-we-do/publications/WCMS\_168813/lang--en/index.htm

Baseline information needs to be collected and analysed for each of the indicators selected. For this reason, the more indicators there are, the more complex (and costly) the monitoring process can be. In addition, if an impact evaluation is planned, baseline data needs to be collected on a representative sample of the whole eligible population to allow for the random selection of participants and non-participants. This may considerably increase the cost of data collection. Performance indicators and data collection strategies need to be grounded in the type of data system in place and the figures that can be produced (sources, collection methods, frequency and costs).

#### 5.1.1. Sources of baseline data

The first challenge is to identify the data sources for the performance indicators of the youth employment programme. Data sources for the baseline can be primary (collected by the implementing agency specifically for programme purposes through the running of a baseline survey) or secondary (collected by other entities for other purposes), or (more often) a combination of both.

### Secondary data sources

The main sources of secondary data for building the baseline of youth employment programmes are national labour force surveys (or other household-based surveys that collect labour market data) and administrative figures on labour and education. A labour force survey (LFS) is a specialized sample survey of households and individuals conducted by national statistical offices on a regular basis to obtain data on the number of employed, unemployed and underemployed. Labour force surveys cover basic demographic characteristics (e.g. sex, age, educational attainment, marital status and so on) and core labour force variables (e.g. current activity status, hours of work, occupation and so on).

If the country runs labour force surveys, the key elements to verify are:

- 1) the sample size;
- 2) the design of the questionnaire to determine whether the outcomes of the youth employment programme are covered; and
- 3) the frequency of the survey.

12 Hussmanns, R., Mehran, F., Verma, V. 1990. Surveys of economically active population, unemployment and underemployment: An ILO manual on concepts and methods (Geneva, ILO).

### Monitoring of results: Data gathering and analysis

5

The main difficulty in using LFS data for the establishment of a baseline is the sample size for the population of interest. Since the LFS aims to identify the labour market status of household members aged 15 and over, the number of youth with the characteristics of the target group may be too low to be reliable. For example, if the programme targets early school leavers in poor rural areas, it is likely that the number of youth with this characteristics in the LFS is too low to form a reliable baseline. The frequency of the LFS is also important for the gathering of monitoring data. If the LFS is run on an annual basis only, it is not feasible as a source of monitoring data (which ideally should be available on a quarterly or semi-annual basis).

There are also other household-based surveys that can provide labour market information (Household Budget Surveys, Living Standards Measurement Survey, Income and Living Conditions Survey). However, all suffer from the same constraints (e.g. sample size of the population of interest, level of disaggregation and frequency).

Administrative data, generally derived as by-products of administrative procedures, can be a cheap and efficient source of information. Two sources of administrative data are most relevant for youth employment. The first is data from labour-related administrative records. The second is data from the national educational system. Labour-related administrative data can be obtained from social security organizations, public and private employment offices, unemployment insurance schemes, payroll tax registers and civil service administrations. Data from the national education system provide information on the quantity and skill level of new entrants into the labour force. These data record literacy rates of young people, educational attainment, school enrolment ratios and dropout rates.

Data derived from administrative records often suffer from a number of shortcomings, such as limited coverage of the total population, inflexibility of concepts, definitions tied to the administrative rules, and restricted access. For example, in countries with largely unorganized labour markets, administrative sources such as employment insurance records often do not exist. Similarly, the data of the public employment service record information on

13 Pember, B. 1997. Labour statistics based on administrative records: Guidelines on compilation and presentation, (Bangkok, ILO/EASTMAT).

those individuals that choose to register and are "unemployed" according to existing rules. This may exclude large segments of the target population. The main advantage of administrative data is that they follow individuals over time and can, therefore, be a useful source of flow statistics and other longitudinal data.

The government has decided to launch a second chance programme targeting youth aged 15—24 belonging to a national minority group. According to census data, this minority group accounts for 5 per cent of the 15–24 population (or 100,000 individuals). Approximately 50 per cent of this group fails to complete compulsory education (nine years of school). Educational attainment in the country is a good predictor of labour market outcomes. Individuals with primary education or less have unemployment rates that are twice those of individuals with lower secondary education (there is no disaggregation based on national origin). A recent poverty assessment also indicates that the lower the education of an individual, the higher the incidence of poverty. These figures are confirmed by the data of the Public Employment Service. These show that youth belonging to national minorities with low educational attainment experience longer unemployment spells compared to their peers in the national majority group. They are also more likely to be recipient of social assistance.

As human and financial resources are insufficient to cover all the eligible population (50,000 youth according to the census data) at the programme's start, the government decided to phase in the programme over five years, starting from eligible youth with longer unemployment spells. The programme has three outcome indicators: (i) compulsory education completion rate, (ii) employment rate, and (iii) earnings/poverty levels.

To set the baseline, available household-based survey data were examined and discarded. The census figures are over three years old. In order to measure results with the same data source the programme management needs to wait for annual population census updates to obtain the data of interest. The same applies to the poverty assessment, which is carried out every five years. The figures of the Labour Force Survey are made available once a year and contain data on educational attainment and employment rate, but not on wages and earnings, for which other sources

### Example <

Setting the baseline for youth employment programmes with secondary data

3

4

5

# Monitoring of results: Data gathering and analysis

### >>>> Example

Setting the baseline for youth employment programmes with secondary data

cont..

would have to be used. In addition, the number of minority youth who have been unemployed for one year or longer in the LFS is too low to be reliable.

The records of the Public Employment Service (PES) show that national minority youth registered as unemployed and with a low level of education account for 54,500 individuals. The records include information on household and individual earnings and highest level of education attained. They also track individuals over time (individuals have to re-register every three months). The PES also track individuals that exit the unemployment register through the data of the National Insurance Fund. This set of data provides information on formal (registered) employment.

The baseline for the second chance programme, therefore, was constructed on a mix of PES and National Insurance Fund data sources, as shown in the table below:

Indicators	Baseline
Percentage of eligible youth with uncompleted compulsory education	58 per cent of minority youth registered with the PES had uncompleted compulsory education
Percentage of eligible minority youth employed	45 per cent of minority youth (aged 15–24) was employed in the formal economy (35 per cent in wage employment and 10 per cent in self-employment)
Earnings/poverty level of eligible youth	90 per cent of minority youth registered as unemployed are living below the national poverty line.  Out of the 45 per cent minority youth employed, approximately 50 per cent live in households below the poverty line

### Primary data sources

When the existing statistical system is unable to provide the data necessary to set the baseline for the youth employment programme, it is necessary to collect figures from primary sources. Primary data can be collected in a number of ways (e.g. through focus group interviews or direct observation). The most common method is to conduct a specifically designed, one-to-one survey of the target group. An example of an individual survey to collect primary data on youth employment is the ILO school-to-work transition survey (see Box 5.1).

### Box 5.1. School-to-work transition surveys

The school-to-work transition survey (STWS) is a framework combining two surveys. The first addresses young people (labour supply and conditions of work) and the second targets employers (labour demand). The youth survey addresses individuals aged 15–29 with a suggested sample size in the range of 1,300 to 4,000 young people. The survey is designed to measure the transition paths from school to work of different categories of young people. A transition is defined as the passage from "the end of schooling to the first regular or satisfactory job". A regular job is defined in terms of the duration of the contract or the expected length of tenure. A satisfactory job is subjective, based on the self-assessment of the jobholder.

The school-to-work transition framework provides the basis for obtaining most of the indicators necessary to analyse the youth labour market. These include: labour force participation rate, employment-to-population ratio, employment status, employment by branch of economic activity, unemployment rate and inactivity rate.

The ILO has developed a methodological guide covering questionnaire design, sampling methodology and definitions of key youth labour market indicators. This helps users to design the survey and manage the various steps involved in data collection and analysis.

Sources: ILO. 2009. School-to-work transition survey: a Methodological guide (Geneva, ILO).

# Monitoring of results: Data gathering and analysis

5

If primary data need to be collected to build the baseline, the steps to be followed include:

- 1) selection of the sampling strategy,
- 2) design and pilot-testing of the questionnaire; and
- 3) data collection and cleaning.

The timing is also important. A baseline survey for a youth employment programmeshould be run before the programme actually starts. In this way it is also possible to plan the impact evaluation approach, if this is envisaged. Since the running of a good baseline survey is not an easy task, it is better to contract a data collection agency to undertake the work. The national statistical office is the most obvious choice, but research institutes, private think-tanks and the employment service are also possible options. It is important to ensure that the methods used for collecting and analysing the data comply with the criteria of transparency, reliability and integrity. Tool 1 offers an example of terms of reference for commissioning a baseline survey for monitoring a youth employment programme.

### 5.1.2. Sampling strategy

The selection of the sampling strategy requires:

- 1) the identification of the population of interest;
- 2) the selection of the sampling frame; and
- 3) the drawing of units from the sampling frame for data collection.

The *population of interest* needs to be clearly specified. This requires the definition of the unit for which outcomes will be measured, with clear specification of the precise geographic coverage, as well as any other features that characterize the population group. The population of interest for the collection of baseline data for a youth employment programme is the programme target group. For example, the population of interest may be teenagers (aged 15–19) or young adults (aged 20–24) in the entire country, those living in one or more particular regions, or teenagers/young adults belonging to a national minority and so on. If the programme targets young school leavers, the population of interest comprises all youth (the age parameters of "youth" need to be specified) who left school early (a definition of "early school leaving" is also necessary). <sup>14</sup> If the programme does not

14 For instance, EUROSTAT measures early school leaving as the share of youth aged 18–24 with at most lower secondary education and not in school or training.

target a specific group, the survey needs to gather data on the whole youth population.

Once the population of interest has been identified, it is necessary to establish a **sampling frame**. This is the most comprehensive list that can be found of individuals in the population of interest. Ideally, the sampling frame should coincide with the population of interest (i.e. the sampling should include all individuals with the characteristics of the population of interest). A recent census of the population of interest constitutes the ideal sampling frame. In practice, however, existing lists, such as population censuses or administrative listings are used as sampling frames. For example, the sampling frame for a programme targeting early school leavers may comprise the listing of students that left the education system after lower secondary education, as recorded by education authorities (administrative listings), or those recorded by the census as having lower secondary education only, or those detected as early school leavers by household-based surveys. A sampling frame that does not coincide with the population of interest creates a coverage bias. When this happens, the results of the survey do not have full external validity (i.e. the findings cannot be generalized to the whole population of interest, but apply only to those individuals included in the frame). For example, if a second chance programme targeting unemployed youth (aged 15–24), who did not completed nine years of basic education, uses the records of the Public Employment Service as the sampling frame, coverage would be biased because not all the unemployed actually register with the employment services. If the records of the PES are used, the sampling frame is valid only for those registered as unemployed, but not for the whole population of interest (e.g. youth aged 15—24 who did not complete basic education and are unemployed).

Once the population of interest and the sampling frame has been identified, it is necessary to choose a method for the *drawing of units* (the selection of individuals to be interviewed) from the sampling frame. This is done through *statistical power calculations*. These calculate the minimum sample size (minimum number of individuals to be interviewed) needed for the survey results to be representative of the outcome of interest. This is based on two parameters: *confidence level and confidence interval (or margin of error)*. Usually researchers use a 95 per cent confidence level, which means that the same result will be obtained in 95 per cent of cases. *Confidence intervals* are calculated based on the standard error of a measurement. For a 95 per cent confidence

Monitoring of results:

Data gathering and analysis

5

level, the interval is 1.96 per cent (even though a margin of 5 per cent is the usual choice as a tolerable margin). Web-based software tools can be used to calculate the required sample size for a research. For example, the sample size of individuals to be interviewed out of a population of interest of 150,000 persons – assuming a 95 per cent confidence level and a 1.96 per cent interval – is 2,459 young people. Usually 20 per cent is added to take into account non-responses. In the above example, this would bring the number of individuals to be interviewed to 2,950. If a 99 per cent confidence level is sought, the sample size in the example above would be 4,201 individuals with a margin of error of 2.5 per cent (5,041 to be interviewed).

Once the minimum sample size has been determined, the drawing of the final sample is done through probability sampling methods. The two most commonly used methods are random sampling (every individual in the population has exactly the same probability of being drawn) and *stratified random sampling*. Random sampling can be performed using an excel spreadsheet. The names of all individuals listed in the sampling frame are recorded on an excel sheet with a unit identification number (see Figure 5.2 below). The command "=RAND ()" will assign a casual number to each individual listed. As this casual number is volatile, it is copied through the command "Paste Special >Values" into column D (final random number). In column F, type the formula "=IF(C){row number}>0.5,1,0). This command will select all those individuals whose random number is between 0.5 and 1.

<sup>15</sup> For example, at www.surveysystem.com/sscalc.htm one can insert the total number of the population, the confidence level required and the confidence interval and obtain the sample. For confidence level 99 per cent, the interval is 2.576. Another useful site is www.raosoft.com/samplesize. html, where it is necessary to insert the acceptable margin of error, the confidence level required, the size of the population of interest and the response distribution (the site provides explanations of what each of these terms means).

		Table gr	aphs		
ID Individual number	Surname	Name	Random number*	Final Random number**	Assignement***
1112985925026	Surname 1	Name 1	0.898253348	0.165116351	False
1608984767031	Surname 2	Name 2	0.343764256	0.187827466	False
1805985950018	Surname 3	Name 3	0.840798476	0.671441855	1
2011986916015	Surname 4	Name 4	0.851585619	0.717576697	1
2602985733519	Surname 5	Name 5	0.764414624	0.905662483	1
2802982974700	Surname 6	Name 6	0.239033173	0.424156533	False
1407984954766	Surname 7	Name 7	0.750141661	0.004283028	False
1810983715309	Surname 8	Name 8	0.665626706	0.435148801	False
1908990964993	Surname 9	Name 9	0.902639681	0.920325221	1
2502977979938	Surname 10	Name 10	0.234769881	0.601913529	1
2707992964999	Surname 11	Name 11	0.728558572	0.540060186	1
0311984710068	Surname 12	Name 12	0.468721838	0.765058881	1
0401974715478	Surname 13	Name 13	0.387191166	0.845650001	1
0402971922029	Surname 14	Name 14	0.227581823	0.928160713	1
0408979777069	Surname 15	Name 15	0.500800898	0.403503014	False
7039825015179	Surname 16	Name 16	0.135493727	0.694628184	1
0704983777020	Surname 17	Name 17	0.478342617	0.479499244	False
0708974715387	Surname 18	Name 18	0.352074017	0.785225766	1
0710976805019	Surname 19	Name 19	0.605905958	0.729053519	1
1112994710105	Surname 20	Name 20	0.651030692	0.966715986	1

<sup>\*</sup>Type in the formula bar "=RAND (cell numer)". This will assign a casual number to each individual listed.

<sup>\*\*</sup> Copy the list in column D through the command "Paste Special > Values" into column E (final random number).

<sup>\*\*\*</sup> In column F type the formula "= IF( C ) (row number)> 0.5, 1, 0).

This command will select all those individuals with a random number of between 0.5 and 1, while other individuals will be assigned "FALSE" or "o".

Monitoring of results: Data gathering and analysis

5

		Random ass	ignment		
ID Individual number	Surname	Name	Random number*	Final Random number**	Assignement***
1112985925026	Cerimi	Merihan	0.885932996	0.112496543	False
1608984767031	Ademi	Indira	0.532320521	0.6777098	1
1805985950018	Vejseulji	Ernez	0.246503954	0.307914123	False
2011986916015	Krasnici	Fetije	0.555138877	0.382976354	False
2602985733519	Azirovic	Ragip	0.877836058	0.796525987	1
2802982974700	Krueziu	Esat	0.741181934	0.455694378	False
1407984954766	Neziraj	Bekim	0.139555527	0.502855929	1
1810983715309	Uka	Arjeta	0.270447944	0.476159558	False
1908990964993	Sinanovic	Veton	0.915054909	0.892661832	1
2502977979938	Ismailii	Ibrima	0.114290271	0.097332377	False
2707992964999	Ismail	Gani	0.604188604	0.837632674	1
0311984710068	JEKIC	Slavko	0.206267095	0.128924817	False
0401974715478	Radomirovic	Sladana	0.903299218	0.922051553	1
0402971922029	Berisa	Avdus	0.052165541	0.971787754	1
0408979777069	Knezevic	Marija	0.745522766	0.094878698	False
7039825015179	Mursic	Natali	0.131746552	0.575828563	1
0704983777020	Marinkovic	Milena	0.502853515	0.841349088	1
0708974715387	Cerimi	Dalijeta	0.133470563	0.188407966	False
0710976805019	Matijevic	Danica	0.946520777	0.460976906	False
1112994710105	Simic	Milan	0.060615452	0.149534306	False

<sup>\*</sup>Type in the formula bar "=RAND (cell numer)". This will assign a casual number to each individual listed.

This command will select all those individuals with a random number of between 0.5 and 1, while other individuals will be assigned "FALSE" or "o".

<sup>\*\*</sup> Copy the list in column D through the command "Paste Special > Values" into column E (final random number).

<sup>\*\*\*</sup> In column F type the formula "= IF( C ) (row number)> 0.5, 1, 0).

In stratified random sampling, the population is divided into groups (strata) and random sampling is performed within each group. As a result, every unit in each strata has the same probability of being drawn. Provided that each group is large enough, stratified sampling makes it possible to draw inferences about outcomes not only at the level of the population, but also within each group. The most common strata used are age, sex, socioeconomic status, geographical location and educational attainment. With stratified sampling techniques, one may obtain a higher statistical precision compared to simple random sampling. This is because variations within subgroups are lower compared to those of the entire population.

A stratified random sampling by sex is required for a population of interest of 150,000 young people aged 15–24. There are 76,225 young men and 73,775 young women. The stratified sample, at a 95 per cent confidence level, comes to 383 for young men and 383 for young women. Adding 10 per cent each for non-responses produces a 422 sample for young men and 422 for young women (i.e. 844 individuals to be interviewed).

If the stratified sample is to be applied across different locations, the same approach is used (the sample may also be stratified by multiple strata, for example, location and sex), as shown in the table below (95 per cent confidence level with no addition for non-response).

Region	Men	Women	Sample M	Sample W	Total
North	19,865	18,567	377	377	754
South	18,231	17,251	377	376	753
East	17,654	16,997	376	376	752
West	20,475	20,960	378	378	756
Total	76,225	73,775	1,508	1,507	3,015

Example <<

Stratified random sampling by sex and geographical location

### 5.1.3. Designing and field-testing the questionnaire

The research instrument (questionnaire) of a baseline survey needs to focus on the type of information needed for the measurement of indicators. Its design should follow the frame and format usually suggested for individual and household surveys. 16

### There are some general rules to design survey questionnaires:

- Each question should be written out in full, so that the interviewer can read each question word for word.
- The questionnaire should include precise definitions of all of the key concepts used in the survey, so that the interviewer can refer to them, if needed.
- Each question should be as short and simple as possible and should use common, everyday language.
- The questionnaires should be designed so that the answers to most questions are pre-coded (e.g. each question has a number of possible answers, to each of which is assigned a number).
- The survey should include skip codes that indicate which questions are not to be asked based on the answers given to the previous questions.
- The coding scheme for answers should be consistent across all questions and also include codes for missing data (such as "respondent did not remember", "respondent did not know", "no available data", "processing error").

Since in a youth employment programme the indicators usually concern labour market status (activity, inactivity, employment, unemployment) and earnings (level of individual and/or household income) the sequence and format of the survey questions are generally similar to those of the national labour force survey, where available.<sup>17</sup> If the country has no labour force survey, the sample questionnaire provided in the *ILO School-to-Work Methodological Guide* can be used to draft a specific questionnaire for gathering baseline data. The questionnaire should be field-tested with a sample of respondents

<sup>16</sup> Examples of national labour force surveys can be downloaded from the Internet. For example, the following site contains most of the core questionnaires used by EU countries: http://circa.europa.eu/irc/dsis/employment/info/data/eu\_lfs/lfs\_core\_national\_questionnaires/index\_lfs\_core\_questionnaires\_by\_years.htm,

<sup>17</sup> This is because labour force survey questionnaires have a specific sequence designed to distinguish among the various labour market statuses.

prior to use. Field-testing the entire questionnaire in real-life conditions will allow to check its length (it should take no more than 45 minutes to administer) and to verify that the questions are adequate to produce precise measures of all indicators.

After field-testing, and once the questionnaire is final, survey supervisors and enumerators need to be trained on the questionnaire scope, coding system, skipping patterns and interview techniques. <sup>18</sup>

### 5.1.4. Data collection, entry and cleaning

Even when data collection is contracted to a specialized data collection agency, it is necessary to have a clear understanding of the data gathering, entry and cleaning processes to ensure that quality control mechanisms are in place. Data collection requires extensive coordination among interviewers, supervisors and data entry operators.<sup>19</sup>

Interviewers (also called enumerators) need to be trained prior to data collection, even though they may have prior experience in conducting household-based surveys. To this end, a reference manual needs to be developed outlining survey procedures to minimize non-response rates, replace units in the sample, revisit units that have provided no information or incomplete information, and so on.<sup>20</sup> Interviewers also need to be trained on how to use data collection tools. Nowadays, collection of individual and household data is carried out one-to-one, either face-to-face or via the telephone. When the telephone is used to interview individuals, a web-based software is used to allow interviewers to enter responses directly into the programme and automatically send the data to a storage facility.

Supervisors are responsible for the completion of the assigned workload and the maintenance of data quality. Specific responsibilities

<sup>18</sup> For a complete reference regarding household based survey see: UN Department of Economic and Social Affairs, Statistics Division, 2005. Household Sample Surveys in Developing and Transition Countries United Nations (New York, UNDESA).

<sup>19</sup> The World Bank has developed a web-based impact evaluation toolkit for health systems. This toolkit contains examples of terms of reference, research protocols and data analysis that are useful for understanding the research process. It is available at: web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/EXTHSD/EXTIMPEVALTK/o,,contentMDK:23262154-pagePK:64168427-piPK:64168435-theSitePK:8811876,00. html

<sup>20</sup> A response rate of 80 per cent to a survey would avoid the need to conduct an analysis of non-response bias (e.g. identification of whether non-respondents would have answered differently).

include making the necessary preparations for the fieldwork, organizing and directing the work of interviewers, and spot-checking the data collected through the household questionnaire.

After enumeration is completed, the data are entered into a database using specific software packages.<sup>21</sup> Rough data are then "cleaned", that is, screened to detect incomplete, incorrect or corrupt records. Examples might include answers that were not recorded or questions that were mistakenly skipped, typos or answers that indicate that questions were misunderstood, or answers that provide values beyond the typical range. The survey instructions (termed "protocol") should indicate which action to take when such situations occur (leave the data unchanged, correct it or delete it).<sup>22</sup>

The data are then processed to generate statistical tables according to a pre-determined tabulation plan (based on the variables of interest, such as the indicators of the youth employment programme). For example, the most common tables are those that show respondents' labour market status, highest educational attainment and level of earnings, and provide additional information on the specific features of employment, unemployment, activity and inactivity.

The primary data thus generated will form the baseline for the calculation of the performance indicators of the youth employment programme. If the baseline survey is conducted before programme implementation, participants should be selected (randomly) from the pool of survey respondents. In this way, a large part of the individual informationneededformonitoring the performance of the programme is already recorded. If this is not possible, then individual information on programme entrants has to be re-recorded at programme's start and maintained throughout the duration of the programme. A simple database software, or even Excel, can be used to record the information on individuals prior to their participation in the programme.

An example of how participants' records are complied prior and during a youth employment programme is shown in Table 5.1. This method of recording is useful to measure process and outcome indicators, as columns can be added to the right (to record completion of the programme or dropout, and amount disbursed, as well as employment at follow-up and earnings).

- 21 The Statistical Package for Social Science (SPSS) and Stata are the most common ones, but there are also many free data entry and tabulation software available for data processing (e.g. OpenStat, SOFA, PSPP and MicrOsiris).
- 22 The most common course of action is to attempt to correct the data by returning to the respondents and re-asking the question.

Table 5.1 Records of participants to a youth employment programme

Programme	le/						/ Telephone	Level of	Unem	Work			_	Refu. Amount	Programme	Programme
Code	/ ID No	Surname	Name	Sex /	Age	Address	nomber	(code)	spell (months)	experience PWD (months)		Roma/II	. do	gee disbursed	start date	end date
2.2.4.1	2106986710264	Surname 1	Name 1	Σ	23	Beuville	288-2896	10	11	4	> z		z	N 28,674.00	18/01/2010	05/02/2010
2.2.4.1	2511981715162	Surname 2	Name 2	ட	28	Beuville	2693-676	10	11	78	z	z	z	V 28,674.00	18/01/2010	05/02/2010
2.2.4.1	0710982710054	Surname 3	Name 3	Σ	27	Beuville	227-2184	10	16	0	∠ ≻	_	z	00.000.00	11/01/2010	22/02/2010
2.2.4.1	0503988715338	Surname 4	Name 4	ш	21	Vormak	2472-982	10	43	0	∠ ≻	z	z	14,700.00	16/02/2010	18/03/2010
2.2.4.1	2110985710004	Surname 5	Name 5	Σ	24	Vormak	394-27 34	10	19	0	∠ ≻	_	z	14,700.00	16/02/2010	18/03/2010
2.2.4.1	0709985804507	Surname 6	Name 6	Σ	24	New Ville	416-516	10	11	0	z	z	z	J 23,482.00	11/01/2010	19/03/2010
2.2.4.1	2505983800027	Surname 7	Name 7	Σ	56	New Ville	774493	10	21	0	z	z	z	J 23,482.00	11/01/2010	19/03/2010
2.2.4.1	2609981804513	Surname 8	Name 8	Σ	28	Vormak	872-586	10	16	3	z	7	z	V 23,482.00	11/01/2010	25/03/2010
2.2.4.1	2704981800006	Surname 9	Name 9	Σ	28	Vormak	822-192	10	16	0	z	z	z	N 23,482.00	11/01/2010	25/03/2010
2.2.4.1	2401993755011	Surname 10	Name 10	ш	17	Beuville	214-153	10	7	0	z		z	N 38,500.00	08/03/2010	26/03/2010
2.2.4.1	2602983738548	Surname 11	Name 11	ш	27	Vormak	362-729	10	36	0	z	<b>&gt;</b>	z	V 46,000.00	08/03/2010	06/04/2010
2.2.4.1	1704991744119	Surname 12	Name 12	Σ	18	Vormak	17471238	10	4	0	z		z	N 28,320.00	08/02/2010	08/04/2010
2.2.4.1	1704986800027	Surname 13	Name 13	Σ	23	New Ville	423-593	10	49	0	z	z	z	N 30,500.00	11/01/2010	09/04/2010
2.2.4.1	0907989710228	Surname 14	Name 14	Σ	20	Monrose	4244724	10	2	0	z	z	z	V 29,700.00	11/01/2010	16/04/2010
2.2.4.1	1610985722214	Surname 15	Name 15	Σ	24	Monrose	66-11-85	10	39	0	z		z	00.007,62 N	11/01/2010	16/04/2010
2.2.4.1	1702988730082	Surname 16	Name 16	Σ	22	Freelake	4566-464	10	18	5	z	z	z	N 34,000.00	01/03/2010	23/04/2010
2.2.4.1	2806980730060	Surname 17	Name 17	Σ	59	Freelake	4211-891	10	2	0	z		z	N 34,000.00	01/03/2010	23/04/2010
2.2.4.1	2205984710281	Surname 18	Name 18	Σ	25	Vormak	826-8306	10	9	24	<i>-</i> ≻	z	z	00.000,07 N	18/01/2010	28/04/2010
2.2.4.1	1312991722225	Surname 19	Name 19	Σ	18	Beuville	254-456	10	24	0	z	7	z	N 27,230.00	12/02/2010	30/04/2010
													-			

3 4

5

# Monitoring of results: Data gathering and analysis

Measuring the results of a youth employment programme To operationalize a performance monitoring system, records of programme entrants (including information on characteristics such as age, sex, education level and other labour market barriers such as early school leaving, long-term unemployed or degree of disability) need to be combined with evidence of individual outcomes at follow-up. The latter should be gathered either through administrative data or follow-up surveys.

To measure employment at follow-up through administrative data, information on programme participants are matched with information of social security contributions, national insurance or payroll tax agencies after a minimum period of time has elapsed (usually six months). The key used to match programme participants to administrative records is their unique identification number (e.g. identity number or social security contribution number).

There are a number of problems with using administrative records to measure employment at follow-up. First, access to personal data may be restricted by privacy protection legislation. Second, administrative records may not be completely accurate or reliable. They may not provide information on earnings, or there may be delays in the updating of records or the cleaning of corrupt, incomplete or inaccurate figures. Box 5.2 provides an example of how a number of administrative data sources are pooled together to measure (re) employment after participation in a programme.

### Box 5.2. Data warehousing in Austria's public employment service (PES)

The Employment Service Data Warehouse (DWH) is a central Management Information System of Austrian labour market policy. It is a monitoring and planning tool designed to perform both strategic planning (management by objectives) and employment statistics functions. The system pools information from various sources in one database. The different data sources are:

- the central register of unemployed individuals, including individual information on registration and benefit recipient status (e.g. unemployment benefit, unemployment assistance, social assistance);
- applications for labour market support schemes, including individual information on type and duration of support; and information on programme service providers (e.g. training providers, partner enterprises and so on); and
- the register of enterprises and vacancies, including applications to partner into programmes.

For the purpose of programme monitoring, the employment service relies on information provided by all the above data sources. The monitoring system is divided into three modules:

- *Follow-up monitoring* of individuals who participated in employment programmes. The system evaluates the situation before and after programme participation according to defined criteria to determine the effectiveness of the measure.
- Career monitoring. This analyses the flows underlying the observed changes, including how many new employment relationships are initiated, by whom and in which sectors. It also analyses how many existing relationships are terminated, where unemployed individuals come from (employment, unemployment, inactivity), where they go and so on.
- Enterprise monitoring. This deals with the human resources levels in companies and whole sectors, including staff increases and cuts. It also analyses whether enterprises with subsidized jobs differ from non-subsidized enterprises and so on. Enterprise monitoring covers registered employers. Therefore, it captures only individuals who are linked to an employer through the insurance system.

The DWH uses the following indicators: (i) personal characteristics (e.g. sex, age, nationality, educational attainment, disability, occupation); (ii) programme information (e.g. region, type of measure, period of participation, code of termination); and (iii) career information (before and after, including duration of unemployment prior to measurement, labour market position, status – employment, unemployment, out of labour force – economic activity and size of employer).

**Source:** Buzek, A. et al. 2004. Labour market monitoring based on the Data Warehouse of the Public Employment Service (Vienna Bundesministerium fur Wirtschaft und Arbeit Bundesminister).

#### Monitoring of results: Data gathering and analysis

5

The main flaw in computing employment at follow-up through administrative data, however, is the failure to capture participants that are (self)-employed in the informal economy. If employment in the informal economy is common in the country or geographical area where the programme is implemented, the measurement of employment at follow-up through official records is likely to be underestimated.

If data on the employment of participants after the programme cannot be derived with certainty from existing data sources, it may be necessary to conduct a follow-up (or tracer) survey on programme participants to measure their (re)employment rate.<sup>23</sup> The main distinction between a follow-up and tracer survey is that the latter may not be fully representative. Much depends on the number of beneficiaries that can be "traced" and interviewed many months (if not years) after their participation in the programme.

Six months after the programme's end, a follow-up survey is conducted with participants to verify their labour market status and level of earnings. Ideally, all participants need to be interviewed either through a one-to-one interview, face to face or by telephone. If the number of participants is large, a representative sample needs to be drawn to ensure the validity of the data collected.

The design and implementation of the follow-up survey follows the same steps outlined for the collection of primary baseline data. The only difference is that the survey instruments need to be designed to verify both quantity and quality of employment and measure the relevance of service provision to employment outcomes. An example of a questionnaire for follow-up surveys is presented as **Tool 3** of this learning package.

<sup>23</sup> Tracer studies obtain their name from the primary activity involved, which is to trace, find or locate a group of individuals. With tracer studies, significant effort is expended in tracing the interviewee. Tracer studies are sometimes referred to as follow-up studies because they trace individuals some time after their participation in a programme, and follow-up on what has happened in their lives subsequently. For a complete guide to tracer studies see International Programme on the Elimination of Child Labour (IPEC), 2011. Child Labour Impact Assessment Toolkit: Tracer Study Manual (Geneva, ILO).

A technical cooperation project with the Public Employment Service (PES) implemented a number of active labour market programmes targeting young people aged 15–29. These programmes were offered over two years (2011–12).

#### 5.3.1. Identification of the target group

The youth employment policy and action plan of the country identified low educational attainment as the most important determinant of poor labour market outcomes for youth. Low educational attainment was particularly common among national minorities, refugees and young people with disabilities. Youth with low educational attainment were more likely to be discouraged workers, to face long unemployment spells, to be working in the informal economy and to live below the national poverty line.

#### 5.3.2. Design of youth employment programmes

Active labour market programmes were designed to tackle the specific barriers faced by youth with low educational attainment (e.g. lack of skills, employers' concerns about the ability of these young workers to perform the task of the job, stereotyping attitudes in recruitment and so on). Table 5.2 summarizes the key features of programmes (type, length and compensation levels).

5-3
Performance
measurement
of youth
employment
programmes:
a practical
example

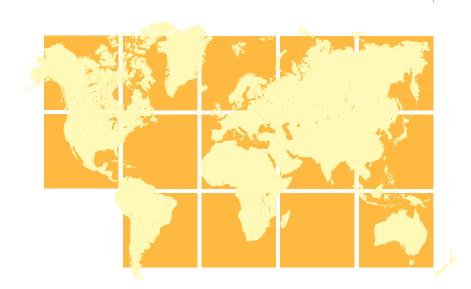


Table 5.2. Type of active labour market programmes offered		
Institution-based training	Competency-based training organized by a private training provider (minimum one month, maximum six months).	
On-the-job training	Competency-based training organized in a private enterprise (minimum one month, maximum six months) No obligation on the enterprise to retain the trainee.	
Work-training programme	Enterprises employing targeted individuals with a definite or indefinite employment contract receive a subsidy for training, providing that the job lasts twice the duration of the programme. Training is to be equal to at least 20 per cent of work time for a period of three months.	
Employment subsidies	Reimbursement (100 per cent) of the employer's share of social security contributions for a period ranging from one to three months, with an obligation to retain workers for a minimum additional period equal to the duration of the subsidy.	
Work trial programme	Subsidized trial period of maximum 30 working days for jobs of minimum three months duration (or six months if the job is part-time). No obligation to retain the beneficiary, but compulsory notification to the employment service about the grounds for dismissal.	
Self-employment programme	Self-employment services and lump sum grant of \$1,600 (12-month duration).	
Programme for persons with disabilities	Institution-based and/or on-the-job training followed by subsidized employment. For the recruitment of a young person with a disability (PWD) the enterprise can receive a monthly subsidy of \$2,500 for PWD with primary education level, and \$3,250 for PWD with secondary educational attainment or over. Duration is from one to six months with an obligation on the employer to retain the workers for a minimum additional period equal to the duration of the subsidy. Employers could also receive a grant for the adaptation of enterprise premises (\$1,800 payable only once); and a grant of \$1,000 to adapt work stations. The overall length of the PWD programme could not exceed 12 months.	

#### 5.3.3. Baseline and eligibility rules

The baseline was built on the administrative data of the PES. Individual records on registered unemployed were maintained on an IT-based platform. These contain all the information needed to compute process indicators, namely age, sex, level of education, length of unemployment spell, prior work experience and other individual characteristics. Use of this secondary data source caused a coverage bias (e.g. the data did not have full external validity for the entire population of interest, but only for the population included in the sampling frame). However, since the objective of the technical cooperation project was to build the capacity of the PES to provide client-oriented services and targeted active labour market programmes, it was decided that such bias was acceptable.

The data of the PES register also served to develop the eligibility criteria for individual participants. These were established as follows:

- Age: individuals aged 15–29, registered in five PES local offices (see also geographical location below);
- Level of education: youth with primary and less than primary education only For self-employment schemes and programmes targeting persons with disabilities, the criteria were relaxed to include also youth with lower and upper secondary education, provided that they faced additional barriers to labour market entry (e.g. national minority or refugee groups);
- Length of unemployment spell: minimum three months, with priority access granted to individuals with longer unemployment spells; and
- Geographical location: young people living in the North, South and South-West of the country, where the share of unemployed youth with low educational attainment was higher than the national average.

#### 5.3.4. Performance indicators

Process and outcome indicators were designed for all the programmes listed above.

Monitoring of results: Data gathering and analysis 5

Table 5.3. YEM Process and outcome indicators			
Programme	Process indicators	Outcome indicators	
Institution-based and on-the-job training	<ul> <li>Number of entrants by individual characteristics (age group, sex, educational level, unemployment spell, prior work experience, other labour market barrier) and geographical areas</li> </ul>	•	
		<ul> <li>Share of beneficiaries employed at follow-up earning above minimum wage</li> </ul>	
	<ul> <li>Share of entrants who successfully completed the training programme</li> </ul>	- Share of beneficiaries employed at follow-up in occupation of training	
	– Share of dropouts	<ul> <li>Share of beneficiaries employed at follow-up using skills gained in present job</li> </ul>	
		– Average cost per programme entrant	
		– Average cost per participant employed at follow-up	
Work placement programmes	<ul> <li>Number of entrants by individual characteristics (age group, sex, educational level, unemployment spell, prior work experience, other labour market barrier) and geographical areas</li> </ul>		
		<ul> <li>Share of beneficiaries employed at follow-up earning above minimum wage</li> </ul>	
	<ul> <li>Share of entrants by economic sector and size of partner enterprise</li> </ul>	– Average cost per programme entrant	
	– Share of entrants by occupational category	– Average cost per participant employed at follow-up	
Self-employment programme	Number of entrants by individual characteristics (age group, sex, educational level, unemployment spell, prior work		
	experience, other labour market barrier) and geographical areas	-Average earnings per individuals self- employed at follow-up	
	-Share of entrants by economic sector of business	-Additional jobs created by self- employed still in self-employment at follow-up	
		-Average cost per programme entrant	
		-Average cost per participant self-employed at follow-up	

#### 5.3.5. Benchmarking

To benchmark performance, process and outcome indicators were compared with those of the regular active labour market programmes implemented by the PES for the same period and geographical areas. The PES programmes selected for comparison had similar a design and objectives. As the PES programmes did not specifically target young people, only participants that complied with the basic eligibility criteria (age and educational level) set by the programme being monitored were extracted from the list of participants. The data of individual participants in the standard PES interventions were transferred into Excel sheets by type of programme. Out of the 12,000 individuals that participated in the regular PES programmes, only 2,000 (or 16.6 per cent) were youth with a low level of education.

Table 5.4. Features of regular PES active labour market programmes under review		
Regular PES active labour market programmes (used for benchmarking)		
Functional elementary education	Education courses provided to adults aged 15+ to obtain elementary qualifications (15-month duration); grant of \$700 to education providers per individual trained.	
Labour market training	Generic training courses (mainly languages and computer literacy) with a training provider (3-month duration); grant of \$400 to providers per person trained.	
Job-specific trainig	Training programme (4-month duration); grant of \$400 provided to enterprises per individual trained for skills acquisition in a specific occupation.	
Training in entrepreneurship	Self-employment training programme (3-day duration) provided by PES Business Centres.	
Self-employment subsidies	Grant of \$1,600 provided on a competitive basis to individuals which attended self-employment training (12-month duration; i.e. the individual has to maintain the business open for minimum one year or repay the grant).	
Subsidies for persons < 30 years of age	Employment subsidy (100 per cent reimbursement of employer's share of social security contributions) for 2 years with an obligation on employers to retain the subsidized workers for an additional 2 years.	
Subsidies for beginners < 30 years of age	As above, but targeting individuals less than 30 with no prior work experience for a period of 3 years, plus 3 years additional employment obligation on the employer.	
Subsidies for persons with disabilities	Employment subsidy (100 per cent reimbursement of employer's share of social security contributions) for 3 years; no further obligation on employers.	

#### Monitoring of results: Data gathering and analysis

#### 5.3.6. Measurement of process indicators

Process and outcome indicators were computed in early 2012. Process indicators were measured using the records stored in the PES IT-based client registration system. Aside from storing individual client characteristics, this system records the type of programme the individual was referred to, the total amount disbursed, the start and end dates of the programme, and whether the individual completed the programme successfully or unsuccessfully, or interrupted participation for justified or unjustified reasons.

Process indicators allowed measurement of the degree of creaming in programme administration (e.g. selection from the pool of potential beneficiaries of individuals with the best chance to succeed), and comparison of performance across local employment offices in terms of compliance with established eligibility criteria. To give an example, the level of creaming in the programmes supported by the technical cooperation project was found to be approximately 5 per cent of total beneficiaries. For standard PES programmes, only two out of five local offices included in the monitoring exercise were targeting youth with low educational attainment and minority population groups. Creaming among PES programmes exceeded 15 per cent. Creaming was calculated as the share of young participants in the programme under review, who did not comply with the entry criteria established by the programme.

#### 5.3.7. Measurement of outcome indicators at follow-up

Outcome indicators were measured only for participants who had completed one of the programmes at least six months prior to the date of monitoring. Individuals who had attended at least two months of training were also included in the follow-up survey. Due to the length of some of the PES regular programmes monitored and the fact that over 60 per cent of young participants had higher educational attainment, less than 500 young individuals (or 25 per cent of young entrants) were included in the measurement at follow-up. Conversely, over 68 per cent of young participants in the programmes supported by the technical cooperation project complied with the measurement criteria.24

24 The remaining 32 per cent were still attending programmes or had only recently completed them.

Outcomes were measured both through administrative records and via a follow-up survey, as employment in the informal economy in the country accounted for approximately 20 per cent of overall employment. The data of the National Insurance and Pension Fund – recording workers for whom social security contribution are paid – were queried to check whether individuals had a job at any time in the reference period.

The follow-up survey involved 1,328 individuals, all of whom were contacted by phone in the span of one week.<sup>25</sup> The non-response rate was approximately 30 per cent due to mistakes in the recording of contact telephone numbers. A non-response analysis was conducted to identify whether non-response was a random occurrence or showed a pattern. The analysis revealed that non-response was random and data was cleaned and aggregated according to a pre-established tabulation plan.

The questionnaire used followed the structure of the labour force questionnaire. Individuals were asked sequenced questions to identify:

- 1) whether they had a job at any time after the programme, and
- 2) their current labour market status (inactive, in-school/training, unemployed, wage employed or self-employed).

Young participants found to be unemployed were asked about their job search activity and the barriers they faced in getting a job. Young participants found to be working were asked about their job/self-employment activity (occupation, economic sector, formal/informal employment relations), their current level of earnings and their level of satisfaction with the services and programmes provided by the PES.

#### 5.3.8. Key findings of performance monitoring

The programmes supported by the technical cooperation project succeeded in targeting the most disadvantaged youth (e.g. low educated youth, with unemployment spells exceeding, on average, two years and no prior work experience). The level of cream-skimming was only 5 per cent and targeting rules were largely complied with. <sup>26</sup> Conversely, PES regular programmes were overly biased in favour of youth with higher levels of education and lower unemployment spells, bringing cream-skimming to approximately 15 per cent. Average cost per entrant was \$1,829 for the supported programmes and \$817 for the regular PES programmes. As employment and earning outcomes were roughly similar, PES programmes were found to be more cost-effective in the short run.

The outcome data of the monitoring exercise revealed that a good share of participants to both programmes were able to obtain a job after completion, but this was usually short-term and often part-

<sup>25</sup> The drawing of a sample was not necessary as the number of beneficiaries was relatively small

<sup>26</sup> This means that only 5 per cent of total participants did not comply with the eligibility criteria set by the programme they attended.

### Monitoring of results: Data gathering and analysis

time. Employment was also highly volatile with job losses over time. However, the data disaggregated by geographical site revealed that the supported programmes were successful in increasing employment opportunities in less dynamic labour markets.

The main differences between the programmes concerned the type of jobs participants gained and the quality of the training received. In particular: (i) over a third of participants of the supported programmes worked in the manufacturing sectors; (ii) over two-thirds worked in the occupation for which they received training; (iii) over 60 per cent worked for the same enterprise that partnered in the programme; and (iv) 70 per cent used the skills learned in their current position. Conversely, less than 23 per cent of participants of the PES regular programmes worked in the occupation for which they received training, and just over 40 per cent used the skills learned in their current job. Over two-thirds, in addition, worked in the service sector characterized by higher earnings, but also higher staff turnover.

A larger share of PES participants earned between \$200 and \$300 per month, while participants of supported programmes were in the lower wage bracket. This was partly due to the fact that these latter participants had been working for shorter periods of time compared to those of PES programmes and mainly in the manufacturing sector, where wages for entry positions were lower.

The dispersion of employment in the informal economy was high for both programmes (18.9 per cent for the supported programmes and 13.6 per cent for the PES programmes), but lower than the informality rate among low educated young workers detected by the Labour Force Survey (28.5 per cent of all young workers).

The figures disaggregated by type of programmes reveal that, with regard to employment at any time, the self-employment programme supported by the technical cooperation project was the measure yielding the higher results (76 per cent of participants self-employed at follow-up). Institution-based training was the least successful (10 per cent). On-the-job training scored somewhere in the middle range (35.5 per cent). In terms of current labour market status, however, institution-based training improved with time, with an employment rate of 20 per cent. On-the-job training gains decreased with time to 24 per cent. Self-employment remained

relatively high, with 40 per cent self-employed at follow-up and 8 per cent in wage employment.

For the PES regular programmes, job-specific training had a short-term employment return of 75 per cent, while labour market training had a much lower return (41.2 per cent). Functional education provided no labour market gain. At follow-up, employment of participants with job-specific training decreased to 50 per cent and that of labour market training to 29.4 per cent. This means that onthe-job training participants were better able than PES participants to hold on to their job. This was mostly due to the fact that most on-the-job training participants worked for the same enterprise that provided the training and in the same occupation. This supports the idea that allowing enterprises to screen future employees through training has merit.

Finally, the programmes promoted by the technical cooperation programme succeeded in increasing the labour market attachment of the target group – measured in terms of individual progression towards labour market activity, rather than employment. The fact that individuals who were unemployed after participating in a programme were still actively searching for a job, in addition to their high level of satisfaction with PES services, was found to be a good indication of this finding.

The Ministry of Labour used the above-mentioned findings to redesign targeting rules and implementation modalities of programmes targeting youth.

27 This was calculated on the share of participants who had a job at any time after the programme.



### Monitoring of results: Data gathering and analysis



#### Are primary data sources better than secondary ones for building a baseline?

Secondary data sources are cheaper to collect, but may not have full coverage of the population of interest. Specification of data may also pose problems. For example, the definition of "unemployed" use by employment services is dictated by administrative rules and it is often not coherent with the international statistical definition.

Primary data sources allow managers to calculate precisely all the indicators specified in a youth employment programme, but collection may take time and be costly. However, one of the main advantages of conducting a baseline survey is that it can be used for the random selection of participants for the youth employment programme. This, in turn, allows the implementation of an experimental impact evaluation (see Part 2 of this learning package).

#### When can stratified sampling methods be used?

Stratified sampling is used when inferences about different groups targeted by the programme need to be made. Examples include young men vs. young women, or teen-agers vs. young adults. Stratified sampling is most commonly used to draw units for interviews in different geographical areas.

#### Is a 95 per cent confidence level always required for a survey?

A 95 per cent confidence level is the parameter commonly used in sample surveys. However, some researchers also use a 99 per cent confidence level (with a confidence interval of 2.576).

#### When do follow-up surveys need to be run?

Follow-up surveys are usually conducted six months after programme completion. However, as the link between short-term and long-term programme outcomes is usually weak, it is better to undertake additional surveys at 12 and 18 months after the end of the programme, if feasible (in such cases a survey may become a tracer study). For example, various studies have found that the full labour market effect of training programmes tends to become material two years after the end of the programme.

#### What is the difference between a follow-up survey and a tracer survey?

The main difference relates to the ease or difficulty encountered in contacting programme participants for the purposes of the survey. Follow-up surveys are usually conducted 6 to 12 months after the end of the programme, thus participants are usually contacted with relative ease. Conversely, tracer studies employ methods to locate beneficiaries some years after the end of a programme. The choice of survey instrument may also depend on whether researchers wish to map the subsequent activities of beneficiaries in detail.



FAQ

cont.

#### Additional reading

**ILO.** 2009. *School-to-work transition survey: a Methodological guide* (ILO, Geneva), www.ilo.org/employment/areas/WCMS\_140862/lang-en/index.htm

International Programme on the Elimination of Child Labour (IPEC). 2011. Child Labour Impact Assessment Toolkit: Tracer Study Manual (International Labour Office, Geneva).

**Gertel, P.J. Martinez, S., et al.** 2011. **Impact evaluation in practice** (Washington, DC, World Bank).

United Nations, Department of Economic and Social Affairs, Statistics Division. 2005. Household Sample Surveys in Developing and Transition Countries (New York, UNDESA), http://unstats.un.org/unsd/hhsurveys/pdf/Household\_surveys.pdf

**World Bank.** 2010. *Impact Evaluation Toolkit* (Washington, DC, World Bank), http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/EXTHSD/EXTI/2154~pagePK:64168427~piPK:64168435~theSitePK:8811876,00. html

**Zall Kusek, J., Rist, R.C.** 2004. *Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners* (Washington, DC, World Bank).

Monitoring of results: Data gathering and analysis

#### >>>>> Glossary

#### Baseline survey

A baseline survey aims to provide the values of indicators before programme implementation. Baseline values can be measured against the values of the same indicators during or at the end of the programme to assess progress (or lack of it).

#### Primary data sources

Primary data sources comprise information collected by the programme implementation agency/organization specifically for programme purposes. Collection is done via a specifically designed questionnaire administered to a sample of individuals representative of the population of interest.

#### Secondary data sources

Secondary data sources comprise information collected by entities other than the programme implementation agency/organization for their own purposes. This means that the data may not be as complete as needed.

#### Population of interest

The population of interest defines the features that characterize the units (individuals to be interviewed) of a survey and its geographic coverage. For example, the population of interest may be college students, youth with primary education or less, or all young people aged 15–24, and so on.

#### **External validity**

For a survey to be externally valid, it is necessary that the sample is representative of the whole population of interest and that the nonresponse rate does not endanger the requirements of statistical power calculation. For this reason sample units for interviews are usually increased by 20 per cent.

#### Statistical power calculation

Statistical power calculations are a tool to calculate the minimum sample size (minimum number of observations) needed for the survey results to be valid.

#### Random sampling

Random sampling is a method to draw units (individuals to be interviewed) from a sampling frame, where every individual listed in the frame has exactly the same probability of being selected for interview.

#### Glossary < <<<<

#### Stratified random sampling

In stratified random sampling the population is divided into groups (e.g. men and women, or individuals at different locations) and random sampling is performed within each group/location. As a result, every unit in each group/location (or stratum) has the same probability of being drawn.

#### Follow-up survey/tracer study

Intracer studies, young people who have participated in a programme are located (traced) and questioned on their (detailed) employment history after they left the programme. The main objective of a tracer study is to design strategies that will allow participants to be traced several years after they have left the programme. The main objective of a follow-up survey, on the other hand, is to identify how many participants have obtained a job after the programme (and which type of job). Since these surveys are run early on after a programme's end, participants are not difficult to contact and no specific strategy is used to locate them.

#### Non-response rate

In sample surveys, the non-response rate describes the proportion of individuals in the sample from whom no information could be obtained for any reason (death, absence or refusal to reply). Attrition rate is a form of non-response and occurs when data are incomplete for one or more unit of the sample.

cont.

Monitoring of results: Data gathering and analysis 5



6

- 6.1. Treatment an control groups
- 6.2. Approaches to build valid control groups
- 6.3. Gross and net estimates of a youth employment programme
- 6.4. Choosing among different impact evaluation approaches

Learning objectives

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

Outline the main approaches used to estimate the impact of a youth employment programme;

Distinguish between "gross" (performance monitoring) and "net" estimates (impact evaluation); and

Decide which evaluation design to use for a specific youth employment programme.

Learning exercises

Which evaluation method is needed?

The goal of an impact evaluation is to measure the causal effect of a youth employment programme on the outcomes of interest (usually employment and earnings) at some point after the end of the programme<sup>28</sup>. An impact evaluation establishes causality by estimating to what extent a programme contributed to the change in the outcome, thus ruling out the possibility that any other factors – aside from the programme – explain the effect. The term "impact", in this context, represents the change in outcome that is directly attributable to a programme. The basic impact evaluation formula is:

#### $\alpha = (Y \mid P = 1) - (Y \mid P = 0)$

This indicates that the causal impact ( $\alpha$ ) of a programme (P) on the outcome (Y) is the difference between the outcome (Y) in the presence of the programme (when P = 1) and the same outcome (Y) without the programme (when P = 0). This formula is valid for an individual, a household, a community or any other type of beneficiary (unit of observation) that receives the programme.

The calculation of the first part of the formula is easy, as it measures the outcome of interest for programme participants (e.g. how many are employed and their average earnings). To calculate this part of the formula, it is necessary to carry out a follow-up survey at a certain point after the programme (as explained in Module 5).

### **Evaluation** goals

The term "programme" refers to youth employment interventions as detailed in the Introduction (namely employment services, vocational training, wage subsidies, public works and self-employment measures).

The second part of the formula, however, cannot be directly measured, since individuals cannot simultaneously participate and not participate in the programme being measured (this is called "the counterfactual problem"). To estimate the counterfactual (e.g. what would have happened without the programme), comparison (or "control") groups are used. This means that the formula cited above becomes:

$$\alpha = (Yp) - (Yc)$$

where the causal impact  $(\alpha)$  is equal to the difference between the outcome Y found for participants (p) and outcome Y found for the control group (c). For example, if the outcome of interest is employment, the causal impact is equal to the employment rate found among participants to a programme minus the employment rate found among the members of the control group. Provided that the control group is indeed a valid comparison group, the estimation of impact is straightforward.

# 6.1 Treatment and Control Groups

The key task in an impact evaluation is to identify a control group that is in all ways similar to the treatment group (i.e. the participants to the youth employment programme), with the only difference being programme participation. Treatment and control groups should share the same characteristics in at least three ways:

- 1. The two groups should be identical in terms of observable characteristics (sex, age group, level of education, labour market status and household income) and unobservable characteristics (preferences, motivation and interests). Although not every young participant will have an exact twin in the control group, the important thing is that, on average, the two groups are the same.
- 2. Treatment and control groups are expected to react to the programme in the same way. This means that the outcome of interest is likely to increase (or decrease) for both participants and control group members. For example, if a youth employment programme has the objective to increase the skills levels of participants, this should happen also for the control group, were they to participate in the programme.

- 3. Treatment and control groups are exposed in the same way to any other factor that is external to the programme to be evaluated. This also includes access to services that are not under evaluation. For example, if the intervention consists of a wage subsidy targeting youth registered with the employment agency, and the agency offers job search assistance to all its clients, these services need to be available to both the treatment and the control group.
- If the members of the treatment and control group are different in some substantial way, there is a "selection bias". This means that the different outcomes one may find after the programme can be due to these differences rather than to programme participation.

Impact evaluations need to be internally and externally valid. Internal validity means that the evaluation is able to show causality between the change in outcomes and the youth employment programme. In other words, individuals experience different outcomes according to whether they participated in the programme or not. External validly means that the findings can be generalized, that is, the same results can be expected when the programme is provided to larger groups of the same population. In order to ensure external validity, a suitable sampling strategy is used to select the individuals whose outcomes will be measured.

There are two approaches to select a valid control group for impact evaluations of youth employment programmes: experimental and quasi-experimental approaches.

#### 6.2.1. Experimental approaches

In experimental approaches (randomized experimental design or randomized controlled trial) a group of individuals who satisfy the eligibility requirements of the programme are randomly divided into two groups. The treatment group is assigned to receive the programme and the control groups is assigned not to receive it. Because assignment is random, the treatment and control groups would be expected to have similar experiences in

6.2
Approaches
to build
valid control
groups

the absence of the programme. Randomization, therefore, provides a simple method for building a counterfactual for the treatment group by using the outcomes of the control group. The impact of a programme in a randomized evaluation is the difference in post-programme outcomes between the treatment and the control group.

The key feature of this type of impact evaluation is random assignment. Other methods of selecting participants run the risk of selection bias (i.e. that different outcomes after the programme would have occurred even in the absence of the programme). Selection biases are avoided in a randomized design because participants and control group are selected randomly and differ (on average) only in terms of their programme participation.

Randomized control trial approaches require a simpler data analysis than the alternative options. Although a randomized selection design offers the highest internal validity to evaluate a youth employment programme, this approach is often unfeasible or impractical. For example, randomized design cannot be used: (i) when the programme has already started and beneficiaries have already been selected with other methods; (ii) when there are sufficient resources to cover all eligible individuals and it would be unethical to withdraw services just for evaluation purposes; (iii) when the intervention targets a limited number of individuals with unique characteristics; or (iv) when policy-makers oppose the provision of a programme to one group and not to another. When randomized designs are unfeasible, the evaluation has to revert to quasi-experimental approaches (see below).

#### >>>> Example

Impact evaluation: randomized experimental design Since 1992, the Inter-American Development Bank has financed training programmes in Latin America targeting less educated youth. These programmes combine classroom training with a period of onthe-job experience. Of these programmes, the Juventud y Empleo (JE) programme in the Dominican Republic was the first to include a randomized evaluation design.

Random assignment was performed on a group of potential participants identified by Centros Operadores del Sistema (COS) to which the Ministry of Labour outsourced the provision of training services. Each COS had to select 35 potential participants complying with a number of entry requirements (aged 16–29; belonging to the poorest households of the country; currently not attending school; with incomplete high school education or less; currently unemployed, under-employed or inactive). The baseline data were collected at

## Introduction to impact

evaluations

registration by each COS for all those eligible and interested in participating in the programme (a total of 10,309 individuals).

Once the list of 35 potential participants was completed, the names and identification numbers were sent to the Programme's Coordinating Unit (PCU), which randomly selected participants for the training course. From each group of 35 potential participants, 20 were randomly assigned to the treatment group and were notified by phone of the date and conditions of the training programme. The other 15 were assigned to the control group, but five were randomly placed in a waiting list as substitutes in the event that a member of the treatment group failed to attend.

Of all 2008 applicants, approximately 2,564 youth were randomly assigned to the control group and 5,801 to the treatment group. Of the original treatment group, 1,011 were "no-shows", while 4,791 attended the training. To fill the places of the no-shows, 941 members of the original control group were re-assigned to the treatment group, leading to a "realized treatment group" of 5,723 and a "realized control group" of 1,623.

A follow-up survey was administered 18 to 24 months after trainees had completed their course on a sample of 4,039 young people.

#### Example < < < <

Impact evaluation: randomized experimental design

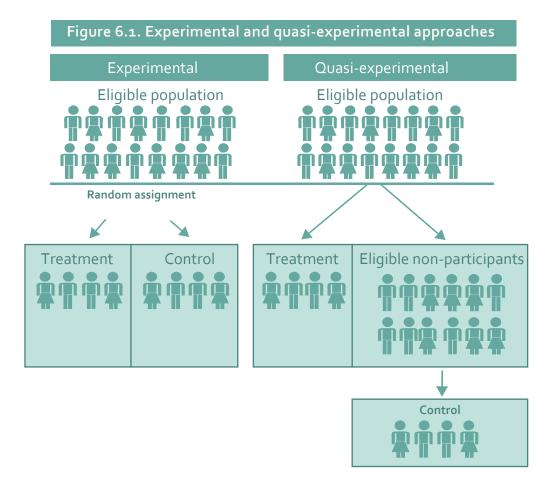
cont.

#### Source

Ibarrarán, P., Ripani, L., Villa-Lora:, J.M. 2011. Youth training in the Dominican Republic: New evidence from a randomized evaluation design (preliminary version of a paper prepared for the 6th IZA Conference).

#### 6.2.2 Quasi-experimental approaches

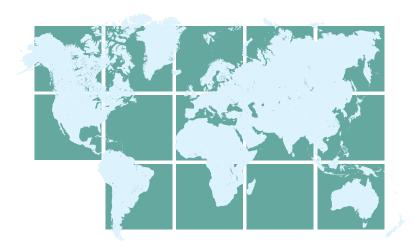
At the core of quasi-experimental approaches are the different statistical methods used to build a valid control group when experimental approaches are not the best choice. Figure 6.1 illustrates the differences between the two evaluation methods.



The most common quasi-experimental approaches used in the evaluation of youth employment programmes are regression discontinuity design, matching and difference-in-differences.

The regression discontinuity design (RDD) is a method that can be used for programmes which have entry rules with a clear cut-off point to determine who is eligible and who is not. Two main conditions are needed. The first is a continuous eligibility index, namely a scale on which the population of interest is ranked, such as length of unemployment spell for registered unemployed, or a test score for a training programme, or simply the age of potential beneficiaries. The second condition is a clearly defined cut-off score, namely a point on the scale above or below which the population is classified as eligible/ineligible for the programme. For example, young individuals with a test score of minimum 40 out of 100 are eligible for an on-the-job training programme. The cut-off point in this example is 40, as anybody below 40 is ineligible.

The regression discontinuity design calculates the difference in post-programme outcomes between individuals near the eligibility cut-off point. The individuals that scored just below 40 in the test become the control group and are used to estimate the counterfactual for those that scored just above 40 in the test.



#### >>>> Example

Impact evaluation: regression discontinuity design In 2005, Finland introduced a legislation sanctioning that young people had to be offered training or work practice within the first three months of unemployment. Labour market training consists of preparatory and vocational labour market training. The employment office managed applications. Typically there were more applicants than training places and the average rejection rate was around 50 per cent. The average duration of training was five months, while preparatory training was much shorter.

The legislation stated that labour market training was to be offered to persons aged over 20 either unemployed or at risk of unemployment. This resulted in a discontinuity between age and participation, which was used to identify the casual effect of programme participation on employment.

The evaluation was based on a dataset built on the annual records of Employment Statistics combined with information from over 20 different registers (tax, pensions, unemployed jobseekers). The dataset contained a wide range of information on individual demographic and socio-economic characteristics, as well as information on unemployment spells and participation in labour market programmes.

A series of sample selection rules were used to build the final sample. First, the sample was restricted to young people in the 16–25 age group who became unemployed in 1998. This made it possible to focus on young people on both sides of the legislative cut-off point (20 years of age). The sample was further restricted to individuals who had not participated in a programme (to address the problem raised by a multiplicity of labour market programmes). Finally, individuals who had completed higher education were excluded to improve comparability of age groups on both sides of the age limit.

As one of the aims of the evaluation was to explore the optimal timing for training, the sample was stratified by monthly duration of unemployment, starting when an individual registered as unemployed. During the second month of unemployment some participated in training (treatment group) while the control group comprised youth who were still unemployed and who did not participate at this point. The following month the treatment consisted of individuals who participated in training during their third month of unemployment. The control group consisted of individuals who, in principle, could have participated, but did not. The same set-up was applied to all durations of unemployment up to the 12th month. A genetic matching method was used to balance out the differences between participants and the control group. The effects of vocational training were found to be negative at the beginning of the follow-up period (locking-in effect).

After that, positive employment effects started to occur, and decreased at the end of the follow-up period. On average, the effect of the programme was found to be below 3 percentage points.

#### Example

<<<<<

Impact evaluation: regression discontinuity design

#### Source:

Hämäläinen and K.Tuomala, J.2007. Vocational labour market training in promoting youth employment (Helsinki, Government Institute for Economic Research, 2007), www.vatt.fi/en/publications/ latestPublications/publication/ Publication\_1345\_id/762



Matching methods can be applied in the context of almost any programme assignment rules, so long as a group exists that has not participated in the programme under evaluation. Matching methods typically rely on observed characteristics (sex, age group, educational level, unemployment spell, national origin, medical condition and so on) to construct a control group. Such methods, therefore, require the strong assumption that no unobserved differences exist between participants and control group, which are associated with the outcomes of interest. Because of this strong assumption, matching methods are normally used in combination with other quasi-experimental methods. Matching uses statistical techniques to build an artificial control group by identifying for every possible individual treated a non-treated individual who possesses similar characteristics.

As an example, take a wage subsidy for young unemployed registered with the employment service. The employment service dataset contains both youth that enrolled in the programme and youth that did not. The programme does not have any clear assignment rules that explain why some young people enrolled in the programme and others did not. In such a context, matching methods allow to identify those non-enrolled individuals that look most similar to participants based on the characteristics available in the dataset (the more individual characteristics available, the better). These "matched" non-enrolled individuals then become the control group used to estimate the counterfactual.

#### >>>> Example

Impact evaluation: matching methods

As a response to rising unemployment figures, the Swedish government increased spending on active labour market programmes in the early 1990s for re-training young unemployed. In 1992, a new large-scale programme targeting unemployed youth (Youth Practice) was introduced.

Propensity score matching was used to estimate the average treatment effect on annual earnings, re-employment probability and probability of regular education. The data used were a random sample of approximately 200,000 individuals collected from databases kept by the Swedish national labour market board (AMS) and Statistics Sweden (SCB).

The database at AMS included records of all individuals registered with the Employment Service, whereas SCB kept records of annual earnings for all persons registered in Sweden.

7 8 9 10

### Introduction to impact evaluations

For each individual in the study, registration dates, labour market status and individual characteristics between August 1991 and March 1997 were combined with information on annual earnings for the years 1985 to 1995.

From the database, all individuals aged 20–24 who registered with the Employment Service during 1992 and 1993 were extracted (10,579 individuals). From this group, all those enrolled in Youth Practice and labour market training were singled out. As a result, the treatment group consisted of 2,263 young people (1,657 youth practice participants and 606 labour market training participants).

The control group consisted of young individuals who registered as unemployed in the same period (1992-93), but never participated in any of the programmes (approximately 5,000 individuals). In principle, all of them could be used as control group. However, the length of the unemployment period before the programme played a significant role in explaining participation. Therefore, to use this information in the estimation of the propensities, a hypothetical programme start date for non-participants was established.

First, both participants and non-participants were divided into subgroups by month of registration with the Employment Service. Second, each non-participant in a sub-group was randomly assigned to a sub-group of participants according to "length of pre-programme unemployment". In cases where the non-participant's actual unemployment period was shorter than the assigned pre-programme unemployment period, the individual was removed from the sample.

This procedure deleted approximately 60 per cent of the sample leaving slightly more than 2,000 non-participants. There were two important issues for this remaining group of non-participants. First, it differed from the original, larger group of 5,000 individuals, due to the shorter pre-programme unemployment spell. Second, unemployed who did not succeed in finding a job (or for some other reason de-registered from the Employment Service) were offered new possibilities for participation in programmes. The remaining group, therefore, was verified to ensure that they had never participated in either of the programmes under evaluation. Of these, 2,000 persons had never participated and formed a valid control group.

#### Example

Impact evaluation: matching methods

cont.

#### Source:

Larsson L. 2000. Evaluation of Swedish youth labour market programmes (Uppsala University), http://qvar.nek. uu.se/Pdf/oowp6.pdf The difference-in-differences (DD) technique is commonly used in impact evaluations as a good alternative to randomized design. The DD design uses information from a non-randomly selected control group to construct the counterfactual. In general, the difference in the post-programme outcomes of participant and control group will include both the "true" treatment effect and the "selection bias" caused by the differences between the two groups.

The DD design compares the difference in outcomes between the participant and the comparison group after programme completion with the difference that existed before the programme. Provided that the selection bias is constant over time, the change in the difference between the participant and control groups from before the programme to after (i.e. the difference-in-differences) constitutes its effect.

#### >>>>> Example

Impact evaluation:
Difference-in-differences

Between 1973 and 1978, the Indonesian Government constructed over 61,000 primary schools throughout the country (INPRES programme). The effect of this programme on education and wages was evaluated by combining differences across regions in the number of schools constructed with differences across cohorts due to the timing of the programme.

The evaluation research used the 1995 inter-census survey of Indonesia (SUPAS), which used a sample of over 200,000 households. From this, a sample of 152,989 individuals born between 1950 and 1972 was extracted (e.g. individuals who had completed their education). The SUPAS collects data on the previous month's wage for people working for pay. From this, an hourly wage was calculated by dividing the monthly wage by the number of hours worked during the month. The effects of the programme on education were calculated using the complete sample, but wages were estimated using the sample of wage-employed individuals (e.g. 60,633 persons).

The identification strategy used the fact that exposure to the school construction programme varied by region of birth and date of birth. Substantial differences existed in programme intensity across regions, due to the government's efforts to allocate more schools to regions where initial enrolment was low was launched should have been higher than the education of older individuals in

all regions, but the difference should have been larger in regions that received more schools.

A difference-in-differences method was used to control for variations in education both across regions and across cohorts. Among early cohorts – who did not benefit from the programme because they were too old to attend primary school when it started (individuals aged 12 or older in 1974) – the increase in educational attainment from one cohort to another was not correlated with the number of programme schools per capita built from 1973 to 1978. The same strategy was used to estimate the impact of this programme on 1995 wages.

The estimates suggest that the construction of primary schools led to an increase in education and earnings. Children aged 2 to 6 in 1974 received 0.12 to 0.19 more years of education for each school constructed per 1,000 children in their region of birth. Using the variations in schooling generated by this policy as an instrumental variable for the impact of education on wages, gives economic returns to education ranging from 6.8 percent to 10.6 percent.

#### Source:

Duflo, E.2001. Schooling and labor market consequenses of school construction in Indonesia: Evidence from an unusual policy experiment.-(Massachusetts Institute of Technology Working Paper, Boston).

When discussing performance monitoring, the term "gross employment at follow-up" was used to indicate the type of results a performance monitoring provides on youth employment programmes. This was done to distinguish performance-monitoring results from those of an impact evaluation (which are "net"). The two examples below illustrate two methods commonly used to benchmark the results of employment programmes, both falling short of measuring the "net" effects.

#### Example

Impact evaluation: Difference-indifferences

**6.3** *Gross* and *net* estimates of a youth employment programme

**1 2 3 4 5** 

#### >>>>> Example 1

# Before and after comparisons

In 2005 the government launched a programme aimed at easing the recruitment of young people aged less than 29 in private sector enterprises. The programme envisaged the full waive of social security contributions (employers' share) for 12 months, with an obligation to keep subsidized workers for an additional 12 months. Managed on a first-come, first-served basis, the programme had an annual allocation of \$50 million, sufficient to cover the whole population group. Just before the launch of the programme, the Statistical Bureau ran an ad hoc youth module attached to the Labour Force Survey to measure the transition of youth from school to work. This research provided the following baseline for the year 2005:

- The employment-to-population ratio of youth aged 15–29 was 28 per cent (100,000 youth); 80 per cent were employed as wage workers, 12 per cent were self-employed and 8 per cent were employed as contributing family members.
- The youth unemployment rate was 32.5 per cent, with young people living in rural areas more likely to be unemployed than urban youth.
- The transition to work was measured at over 30 months for youth with less than secondary education and over 12 months for youth with higher education degrees.

In 2009, the Statistical Bureau re-ran the research on the same sample of respondents and found that youth employment had increased to 31 per cent (110,700 individuals) and that wage employment accounted for 88 per cent of youth employed. The transition to work had decreased to 18 months for secondary school graduates and to six months for university graduates.

A before-and-after method would use the employment figure of 2005 as a counterfactual and would conclude that the youth programme increased employment for youth by 10,700 individuals, or 3 percentage points.

However, in 2007 the government liberalized the procedures regulating the entry of foreign investors into the country. As a result, the manufacturing sector, where most foreign firms invested, expanded its annual value added by more than 40 per cent annually.

The change in the number of young people employed (10,700) is shown at point A in Figure 6.2, but the expansion of manufacturing may be responsible for an increase in youth employment equal to the distance between C and A (or 5,700 individuals). Then, the real impact of the programme is only the distance between B and C (or 5,000 individuals), and not the distance between A and B. In other words, unless every other factor that can affect youth employment is accounted for, it is not possible to be sure that what the observed change is due to the programme and the programme alone.

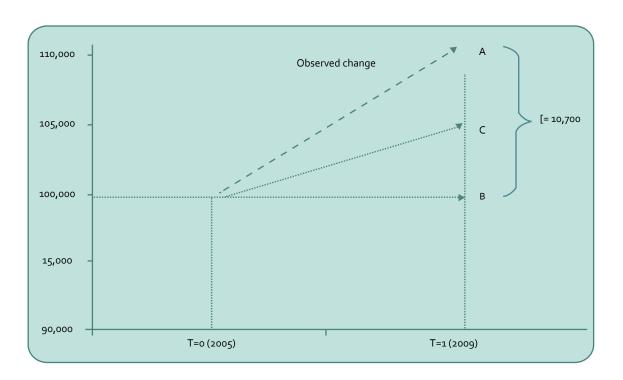
In Figure 6.2, the distance A to B is a "gross" estimate, while the distance B to C is the "net" impact.

#### Example 1 <<<<<

Before and after comparisons

cont.

Figure 6.2. Before-and-after estimates



#### >>>> Example 2

#### Enrolled/ not enrolled comparison

In 2009, the Employment Bureau launched a national vocational training programme for unemployed youth. The aim was to provide 40,000 young unemployed with the skills demanded by the labour market. As the programme could not involve allyouth registered as unemployed (230,000 individuals), it was decided that only those who applied could participate. After two years (2011), the Employment Bureau attempted to estimate the effect of the programme on earnings, by comparing the average earnings of participants who were employed at follow-up with the earnings of youth that did not apply to the programme.

The follow-up study involved 6,000 participants who had completed the programme and 5,000 youth registered as unemployed at the time enrolment into the national training programme started.

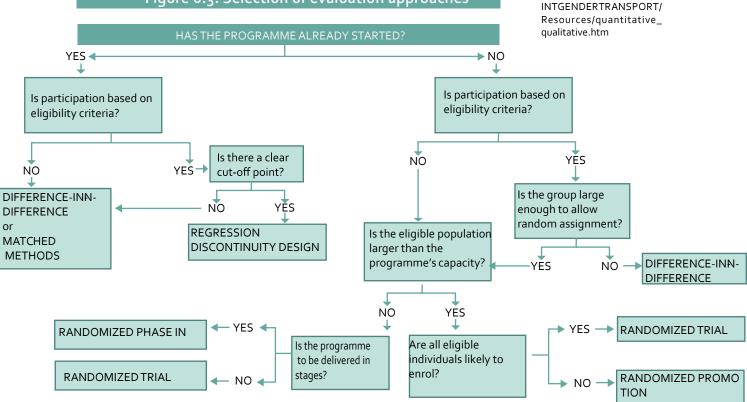
The earnings of participants were found to be on average 40 per cent higher than for youth who found employment but did not participate in the programme. Again, this result represents the "gross" outcome of the programme. This is because the characteristics of individuals who decided to enrol in a programme may be quite different from those who did not enrol. For example, those who enrolled may have had higher motivation, while those who did not may have been discouraged workers. It is likely that their performance in the labour market would have been different in any case.

Therefore, the group that chose not to enrol does not provide a good estimate of the counterfactual. If a difference in incomes is observed between the two groups, it is not possible to determine whether it comes from the training programme or from the underlying differences between the two groups. The fact that less motivated individuals chose not to enrol in the programme leads to a bias in the assessment of the programme's impact ("selection bias"). In this case, if the young people who enrolled have higher incomes even in the absence of the programme, the selection bias would be positive (e.g. the impact of the vocational training programme would be overestimated).

The selection of evaluation method is guided by the specific features of the youth employment programme to be appraised, namely its timing and coverage, the rules it established to enrol participants, and the resources available. Figure 6.3 offers a framework for selecting the most appropriate evaluation method, while Box 6 explains the steps leading to such selection.<sup>29</sup>

As mentioned in the Introduction, the objective of this learning package is to offer a guide for those responsible for youth employment programmes to manage monitoring and impact evaluation, rather than actually running performance monitoring and impact evaluation themselves. Impact evaluation, in particular, is a complex exercise that is best carried out by a team with specific expertise. **Tool 2** of this learning package offers an example of terms of reference for commissioning an impact evaluation. An important part of the terms of reference is the actual selection of the best method to measure the impact of a specific youth employment programme.

#### Figure 6.3. Selection of evaluation approaches



#### 6.4 Choosing among different impact evaluation approaches

29 This learning material discusses mainly quantitative impact methods. However, quantitative evaluations are often accompanied by qualitative assessments (mixed methods).

See, for example, the impact assessments of the World Bank on poverty programmes and projects at http://siteresources. w o r l d b a n k . o r g / INTGENDERTRANSPORT/ Resources/quantitative\_

#### Box 6.1. Guidelines to select an impact evaluation methodology

#### 1. The programme has already started

If the programme did not use randomization to assign participants to the programme, only quasi-experimental approaches can be used. The question of method will work best for a specific programme will depend on the rules established to enrol participants.

#### a. The programme is ongoing and participant selection is based on eligibility criteria

If the programme determines eligibility according to clear criteria with a cut-off point, then regression discontinuity design is an effective method for impact evaluation. Examples of cut-off points include: length of unemployment spell or age (e.g. only youth unemployed for six months or more only youth aged 20–24 and so on).

#### b. The programme is ongoing and participant selection is first-come, first-served (or other rules)

If the programme used methods such as first-come, first-served or caseworker referral for participant intake, a control group can be built using difference-in-differences or matched methods, providing that a group of non-participants exists to form the control group. If a programme targets the entire eligible population and is delivered all at once, it is not possible to construct a control group. If the programme is phased in, then the control group can be constructed using DD or matched methods with those individual who have not yet been enrolled.

#### 2. The programme has not started

If the assignment of participants is based on observed characteristics (e.g. young people with longer unemployment spells) or unobserved characteristics (e.g. letting individuals apply to a programme based on their own motivation), it is still possible to use a randomized evaluation design. However the sample of the sub-group of the eligible population (in the examples above, those with longer unemployment spells and those who apply to the programme) must be sufficiently large to allow for random assignment between participants and non-participants. If randomization is not feasible, difference-in-difference methods can be used. Matching methods may be problematic, as it may be difficult to match unobservable characteristics.

## Introduction to impact

evaluations

#### 3. Availability of resources

The last question concerns the resources the programme has at its disposal for impact evaluation. Different methods have different requirements in terms of sample size, data collection and complexity of statistical analysis. All of these will determine the overall cost of the exercise.

If when designing a programme, the eligible population is found to be larger than the number of participants the programme can accommodate at a given time (because of human and/or financial resources constraints), the programme can use a randomized controlled trail (also called randomized assignment) to decide who in the eligible population receives the programme and who does not. If the programme is delivered all at once, a lottery design will work well (see Module 7). If the programme is to be delivered in stages, a randomized phase-in design can be used. If randomization is not feasible (e.g. because of political opposition, ethical questions), the evaluation will have to use quasi-experimental approaches. A discontinuity design approach can be used if participant intake uses clear eligibility ranking with a cut-off point. Difference-in-differences or matched methods can be used if intake is based on specific characteristics, xx on first-come, first-served basis or caseworkers referral.

The following modules focus mainly on randomized evaluation designs (experimental) and difference-in-differences (DD) design (quasi-experimental). Module 9 briefly examines other quasi-experimental methods

#### >>>>> Additional reading

**Card, D., Ibarrarán, P. Villa, J.M. 2011.** Building in an evaluation component for active labor market programs: A practitioner's quide, (IZA working paper No 6085).

**Duflo, E. 2001.** Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. (Massachusetts Institute of Technology Working Paper Boston).

**Duflo, E., Glennersterz, R., Kremer, M. 2007.** *Using randomization* in development economics research: A toolkit (Centre for Economic Policy Research, Discussion Paper No. 6059, Boston), http://econwww.mit.edu/files/806

**Gertel, P.J., Martinez, S., et al. 2011.** *Impact evaluation in practice* (Washington, DC, World Bank).

Hämäläinen, K., Tuomala, J. 2007. Vocational labour market training in promoting youth employment (Government Institute for Economic Research, Helsinki).

**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).

Ibarrarán, P., Ripani, L., Villa-Lora, J.M. 2011. Youth training in the Dominican Republic: New evidence from a randomized evaluation design (Preliminary version of a paper prepared for the 6th IZA Conference)

**Larsson, L. 2000.** Evaluation of Swedish youth labour market programmes(UppsalaUniversity), http://qvar.nek.uu.se/Pdf/oowp6. pdf

Morra-Imas, L.G., Rist, R.C. 2009. The road to results: Designing and conducting effective development evaluations (Washington, DC, World Bank).

## Introduction to impact evaluations

#### Glossary

#### Impact (effect)

In the context of impact evaluations, this term denotes the intended or unintended changes in outcomes that can be attributed directly to a programme.

#### Impact evaluation

An impact evaluation is an evaluation that 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.

#### Treatment group

This is the group of participants to a youth employment programme.

#### Comparison (or control) group

A comparison (or control) group has the same characteristics as the treatment group, except that those in the comparison group do not benefit from the programme. Control (comparison) groups are used to estimate the counterfactual.

#### Counterfactual

Counterfactual is an estimate of what a certain outcome (Y) would have been for a programme participant in the absence of the programme (P). By definition, the counterfactual cannot be observed directly. Therefore, it must be estimated using control groups.

#### Experimental design (randomized assignment or randomized control designs)

Experimental design is considered the most robust method for estimating counterfactuals and is often referred to as the "gold standard" of impact evaluation. With this method, beneficiaries are randomly selected to receive a programme, and each has an equal chance of being extracted. Those who are not selected become the control group. With samples that are large enough, the process of random assignment ensures equivalence, in both observed and unobserved characteristics, between the treatment and control groups.

#### >>>>> Glossary

#### cont.

#### Quasi-experimental design

Quasi-experimental design approaches are used to construct a valid control group by using statistical means to control for differences between the participants of a programme and those that did not participate.

#### Gross and net impact

The gross effect of a youth employment programme is the change observed in one or more outcomes of interest for participants to a programme. Such changes, however, may be due to a number of factors, including the programme. Net impact denotes the effect that is directly attributable to the programme, and the programme only.

#### Selection bias

Selection bias occurs when the reasons for an individual's participation in a programme are correlated with outcomes. This bias occurs when the control group is ineligible or self-selects out of the programme.

#### Internal validity

To have internal validity an impact evaluation must have a control group that provides a valid estimate of the counterfactual.

#### **External validity**

In impact evaluation, external validity means that the causal impact observed can be generalized to all eligible units. For an evaluation to be externally valid, it is necessary that the evaluation sample is a representative sample of all eligible individuals.

#### Before-and-after comparision

Also known as "pre-post comparison", it is an approach that attempts to establish the effect of a programme by tracking changes in the outcomes of beneficiaries over time, using measurements before and after the programme is implemented. This method does not have internal or external validity, as it cannot determine the counterfactual and it is applied only to participants.

#### Introduction to impact evaluations

#### Glossary <<<<



cont.

#### Difference in differences

This method estimates the counterfactual for the treatment group by taking the change in outcome of the control group. It allows evaluators to take into account any differences between the treatment and control groups that are constant over time. The two differences thus occur before and after the programme, and between the treatment and control groups.

#### Matching

Matching is a quasi-experimental method that uses large datasets and statistical techniques to construct the best possible control group for a given treatment group.

#### Regression discontinuity design

Thismethodofguasi-experimental evaluation is useful for programmes that use a continuous index to rank potential beneficiaries and that have a cut-off point along a scale that determines programme eligibility (e.g. a dividing point between the treatment and comparison groups).

#### What is impact evaluation?

Impact evaluation is the measurement of the causal effect of a programme on the outcomes of interest (usually employment and earnings) at some point after programme completion.

#### What is the counterfactual?

The counterfactual is an estimate of what the outcome of interest would have been for a programme participant in the absence of the programme.

#### What is the difference between experimental and quasiexperimental evaluation?

Both approaches are aimed at measuring the outcome of interest for participants and non-participants to a programme. The difference lies in the way in which the non-participant (control) group is constructed. In experimental evaluation participants and the control group are assigned randomly from the pool of all eligible individuals prior to programme start. Quasi-experimental methods are used to construct a valid comparison group by using statistical means to control for



Questions and answers



observable and non-observable differences among individuals.

#### When does selection bias occur?

Selection bias occurs when participants and the control group differ in some substantial way likely to affect labour market outcomes. For example, if participants are long term unemployed and the comparison group is mostly composed of short-term unemployed, this is a substantial difference that constitutes a selection bias.

#### When can randomized design be used? When not?

Randomized design can be used when it is embedded in the design of the programme (e.g. the rules for programme participation include random assignment) and when the pool of eligible individuals is large enough to allow randomization. Randomization cannot be used when a programme is already ongoing and the rules for participation did not include random assignment. It is also impractical when the pool of eligible young people is too small to allow assignment into two groups (treatment and control).

#### What determines the impact evaluation to be used?

The choice of impact evaluation depends on the timing of the programme (whether it has already started or not), its coverage (all eligible population or not), the rules for participants' enrolment (voluntary, based on strict eligibility criteria, caseworker discretion and so on) and the resources needed to carry out the evaluation.

#### >>> Exercise

Which evaluation method is called for?

For the last few years, the government has been running a work experience programme targeting youth aged 20–24, graduating from technical high school and scientific university streams. After five years of implementation, the managers of the programme want to measure how many of the individuals that participated in the programme have improved their labour market status. A baseline survey was run before the launch of the programme. To measure improved labour market outcomes, managers plan to run a follow-up survey using a sample of individuals who graduated in the last five years from any high school or university in the country and to compare it with the results for participants.

Would this evaluation approach provide a valid measure of the programme impact? What would you propose as a suitable evaluation approach?

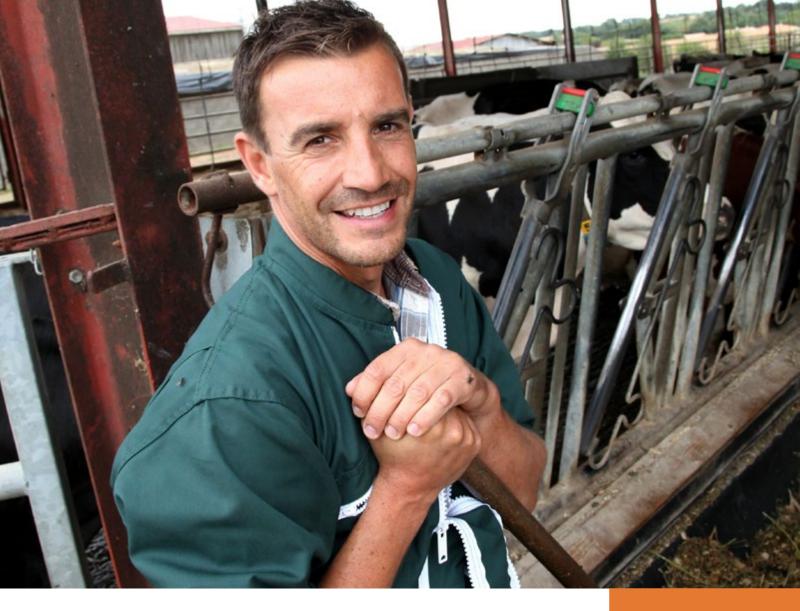
### Introduction to impact evaluations

#### Check your answer



The approach described is similar to a "before-after" measurement and is not suitable to estimate impact. There are also two additional issues. First, the programme targets young graduates from technical streams (for whom a baseline exists), while the follow-up survey is to be done with a sample of graduates from the last five years from all educational streams. This means that the two groups are substantially different. Second, since it is unlikely that the programme was offered to all graduates in the last five years, it is necessary to understand which rules applied to programme participation to decide on the best evaluation method. What is clear is that randomization was not used and that, therefore, a quasi-experimental approach is needed to build a valid control group.

If the programme did not reach all graduates of technical high school and scientific university streams, and the pool of graduates who participated and those who did not is large enough, a matched evaluation design could be used to build a control group. Difference-in-differences could also be used, if there were a sufficiently large dataset for participants and non-participants listing a number of individual characteristics. Also a regression discontinuity design could work, using the lower and upper age range of participants (namely, 20 and 24). Two comparison groups could be constructed (high school graduates that are just short of 20 and are not eligible and university graduates who are aged 25).



- 7.1. Introduction
- 7.2. Randomized assignment of treatment (lottery design)
- 7.3. Randomized phase-in design
- 7.4. Randomized promotion
- 7.5. No-shows, dropouts and crossover
- 7.6. Programme enrolment and impact in different locations

Learning objectives

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

List the main features of randomized evaluation techniques (randomized assignment, phase-in and randomized promotion); and

Estimate impact under randomized control trials.

Learning exercises

Lottery design for a youth employment programme which randomized control trials.

In randomized design (or randomized controlled trial) those who satisfy the eligibility requirements for programme participation are randomly divided into two groups: the treatment group assigned to receive the programme, and the control group. This method provides the best estimate of the counterfactual. Randomization, however, has to be planned from the onset of the programme and applied before the programme actually starts. There are three instances in which randomized assignment works very well as an impact evaluation method:

of programme places available. When the demand for a youth employment programme exceeds supply, a simple lottery can be used to select participants within the eligible population. The winners of the lottery participate in the programme and the others form the control group. As long as there are resource constraints that prevent the scaling-up of the programme to the entire population, the control group can be maintained to measure its short, medium- and long-term impact. In this context, no ethical dilemma arises from holding a control group indefinitely, since a proportion of the eligible population will always be left out of the programme.

When a programme is to be gradually phased in until it covers the whole eligible population. When a programme is expected to cover the whole eligible population, but is delivered in phases, randomization gives each eligible individual the same chance of participating in the first phase or the last phase of the programme. As long as the "last" group has not yet been phased in, it can serve as a valid control group to estimate the counterfactual for those who have already participated in the programme.

## **7.1** Introduction

3. When it is necessary to test a new or potentially costly youth employment programme. Before extending a new (or costly) youth employment programme across all the eligible population, programme administrators may want to verify its effect. In this context, randomized assignment is justified during a pilot evaluation period to test the impact of the programme prior to its wider roll out.

The remaining part of this module will examine three randomized methods for evaluating youth employment programmes: randomized assignment of treatment (lottery), randomized phase-in and randomized promotion.

**7.2**Randomized assignment of treatment (lottery design)

Randomized assignment uses a lottery to decide which individuals among the eligible population receive the programme and which do not. Every eligible unit (e.g. an individual, household, community, geographical district or other) has the same probability of selection. The process of randomly assigning individuals, households or communities to the treatment and control group produces two groups that have a high probability of being statistically identical, as long as the number of potential participants is large enough. If this is the case, randomized assignment will produce groups that have statistically equivalent averages for all their characteristics (observable and unobservable). Through randomization, the difference in outcomes between the treatment and control group at the end of the programme can be attributed to the programme itself because all other factors that could potentially influence the outcome are, on average, equal for both groups.

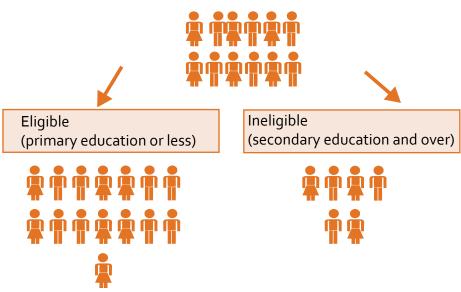
#### 7.2.1. Steps in randomized assignment

The first step in a lottery design is to define those that are eligible for the programme. In youth employment programmes the "unit" is typically a young person (aged 15–24 or 29 according to the national definition). Eligibility is determined by the characteristics of the group to be targeted by the programme. What is important is that the eligibility rules are certain (e.g. they will not change throughout the duration of the programme) and transparent (e.g. they are easily understood and can be clearly communicated to the general public).

As an example, suppose the government wants to introduce a vocational training programme for youth aged 15–24 with primary educational attainment or less, given that a low level of education is a strong predictor of poor labour market performance. The first step would be to extract from the total population of youth aged 15–24 only those that have primary education or less, as shown in Figure 7.1.

30 Module 2 of this learning package offers a sequence illustrating how a target group of young people is selected. Module 5 shows the relation the target group and eligibility criteria

Figure 7.1 Total population (young people aged 15-24)



To extract those young people eligible for the programme, it is necessary to have a complete and updated listing of the whole population aged 15-24 and their characteristics (e.g. sex, age-group, level of education, labour market status). This can be derived from census figures, household-based surveys or administrative listings. When a youth employment programme is managed by the public employment services, the unemployment register is commonly used as the framework for the eligible population of the programme. Administrative data may be used when one is reasonably confident that they cover most of the population (see Module 5 of this learning material). If this is not the case, household-based surveys (such as the Labour Force Survey or Household Budget Survey) provide more comprehensive data of the population of interest. In the example above, both the Labour Force Survey and the Household Budget Survey would provide, among others, the information on educational attainment of the population of interest. The ideal data source, however, would be a recent census of the population. If none of the above-mentioned data sources are accessible (or exist), it will be necessary to collect primary data.

#### >>>> Example

#### Step 1: Definition of eligible units

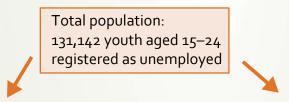
The figures of the Labour Force Survey indicate that youth (aged 15–24) with primary education only are more likely to be unemployed, underemployed or inactive than youth with secondary and tertiary educational attainment.

The figures show that youth represents 12 per cent of the total population of the country (840,000 individuals). Those with primary education are 250,000 individuals. Of these 80,000 are still in education or training (32 per cent), 90,000 are employed (36 per cent) and 80,000 are unemployed (32 per cent). The individual records of the National Employment Agency (NEA) show that youth (aged 15–24) registered as unemployed comprise 131,142 persons, 94,267 of which have received primary education or less and 36,875 have accessed higher education.

To address this problem, the government decided to provide students coming from poor households with scholarships to attend secondary and tertiary education. Concurrently, a large vocational training programme will be launched aimed at providing unemployed youth with low educational attainment with the skills most demanded by the labour market. The programme, to be administered by the NEA, envisages the training of young beneficiaries in public and private adult training centres in occupations most demanded by employers for a period ranging from three to six months. Resources are not an issue as the country is rich in natural resources. The Council of Minister decided that 5 per cent of all the revenues stemming from natural resources would be invested in the scholarship programme and 1 per cent in vocational training (approximately \$8 million per year).

As the records of the NEA are exhaustive and their staff will be entrusted with programme administration, NEA data sources are used for programme purposes.

The total population, therefore, comprises all individuals 15-24 years old registered as unemployed with the NEA (131,142 individuals). The eligible population is defined as youth with primary education only. This amounts to 94,267 individuals.



Eligible population: 94,267 youth with low education

Ineligible population: 36,875 youth with higher education

#### Example <<<<

Step 1: Definition of eligible units cont. The second step in randomized assignment is the recruitment of the eligible population for programme purposes. The programme is advertised as widely as possible to ensure that all eligible individuals are informed and given a chance to participate. Typically, participation in youth employment programmes is voluntary, in other words, those interested in the programme are asked to apply (or enrol). At this stage it will become apparent to programme managers whether the programme is over-subscribed and whether the number of applicants is sufficient to permit a lottery design.

#### >>>> Example

#### Step 2: Recruitment of the eligible population

The total number of training places that can be made available in private and public training centres in the country is 25,000 per year. To treat all the eligible population, the programme will have to be phased in over a number of years (see the Introduction to this section).

The NEA Headquarters instructs its 35 local offices to advertise the programme via the media (radio, television, Internet), advertising material (leaflets, posters), e-mail and text messages. Youth registered as unemployed are given two months (from 1 January to 28 February) to apply. At the end of the enrolment process 86,720 young unemployed with primary education have applied to participate in the vocational training programme.

As a result the programme is over-subscribed for the first year (as only 25,000 places are available) and the programme can be offered only to a fraction of the eligible population.

Eligible population: 94,267 youth with primary education

Applicants 86,720 youth in 35 district offices of the NEA

The **third step** in randomized assignment is the selection of the sample for evaluation. A representative group will be selected for evaluation purposes and baseline data collected for the individuals in the sample.

The drawing of units is determined by statistical power calculations, which compute the minimum sample size needed for the results to be representative of the outcome of interest. This is based on *confidence level and confidence interval* (see Module 5).

The control group needs to be approximately the same size as the treatment group. If the number of places available for a programme is limited compared to the eligible population (e.g. 600 places available over a population of 10,000), the sample that can be drawn comprises 1,200 individuals (although power calculations at a 95 per cent confidence level require a sample of 2,000 individuals). Provided that the two groups are selected randomly, even a smaller number of observations will ensure that findings can be generalized.

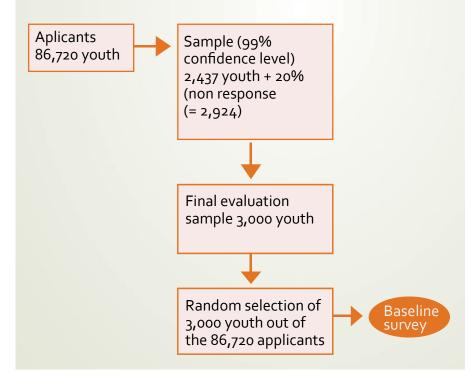
Once the overall sample for evaluation has been drawn (and before proceeding to assignment to treatment and control) it is necessary to run a baseline survey to collect data for each of the indicators to be measured at follow-up (in this example, employment and earnings). Module 5 provides further details of this process.



#### >>>> Example

#### Step 3: Selection of the sample

Power calculations are used to determine the size of the sample necessary for evaluation. If the confidence level required is 99 per cent, the sample size out of an applicant group of 86,720 individuals will be 2,437. Another 20 per cent needs to be added to take into account non-response during surveys. The NEA decides to take a sample of 3,000 individuals for evaluation purposes. As a rule of thumb, the bigger the sample, the better.



To select 3,000 young people out of 86,720 for sampling purposes, it is possible to use either random sampling (see below) or stratified sampling methods. As the 86,720 individuals applied through 35 local employment offices, each office may randomly make a selection for evaluation purposes. For example, the local employment office of the capital city received 2,420 valid applications. Their share of the evaluation sample comprises approximately 86 individuals (43 to be assigned to treatment and 43 to control). The office could use Excel to assign random numbers to the list of applicants, then select only those who have random numbers between 0.8 and 1 ("=IF(C){row number}>0.8,1) until a final list of 85–90 individuals is drawn. Alternatively, the central office could perform a randomized assignment once it receives the applicant lists from the local offices.

Randomization at the level of the local employment office rather than at the individual level means that half the offices receiving applications for the programme will be omitted from the evaluation. Since little knowledge is available of local labour markets and the number of offices is limited, it would be better to avoid this.

Once the sample of 3,000 young people has been selected, it is necessary to gather baseline data. If the records of the employment services provide all necessary information on individuals, there is no need to run a survey. However, a baseline survey may be run in any case to collect additional information. For example, employment records may not record the presence of children and their ages. Such data may constitute an important variable especially when combined with the sex of the young individual.

#### Example <<<<

Step 3: Selection of the sample

cont.

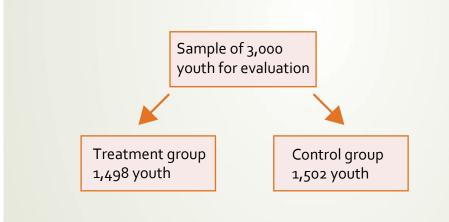
The **fourth step** is to assign young individuals in the sample to two groups of equal size. Random sampling can be performed using an Excel spreadsheet. The names of all individuals listed in the sampling frame are recorded on the spreadsheet with a unit identification number (see Figure 5.2 in Module 5). The command "=RAND ()" will assign a casual number to each individual listed. This number is copied through the command "Paste Special >Values" into column D (final random number). The formula "=IF(C){row number}>0.5,1,0) selects all those individuals whose random number is between 0.5 and 1.

#### >>>> Example

#### Step 4: Randomized assignment

A) The NEA decided to centralize random selection of the sample for evaluation purposes and selects 3,000 young people out of the total number of applicants (86,720 young people).

No baseline survey is necessary because the individual records collected by the NEA contain all the information required to measure the outcomes of interest for the programme (employment and earnings). A final randomization exercise (performed with Excel) assigned 1,498 youth to treatment and 1,502 to comparison.



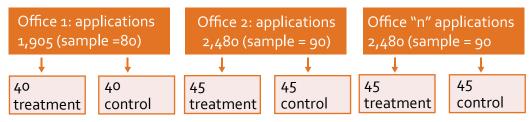
Since there are 25,000 training places for the current year, local employment offices can start organizing the intake into training programmes of 23,498 young applicants (i.e. 25,000, including the 1,498 of the treatment group, minus the 1,502 of the control group). When the 1,498 youth in the treatment group (among the first to be trained) have completed the programme and a sufficient span of time has elapsed, a follow-up survey can be run on both the treatment and control groups to compute their level of employment and earnings.

B) The NEA decided to decentralize random selection of the sample for evaluation purposes to local offices. Each office was asked to draw a number of young people randomly from their list of applicants (80 for small offices and 90 for large offices). Each office then randomly assigned half of the sample to treatment and the other half to control.

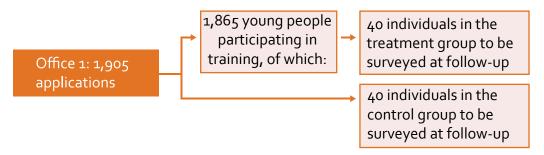
#### Example <<<

Step 4: Randomized assignment

cont.



Local offices were then instructed to begin enrolling applicants into training programmes, excluding those in the control group.



#### Advantages



#### Disadvantages



Lottery design is one of the most credible methods for building a counterfactual, as it leads to two groups that, on average, are equal in all their observed and unobserved characteristics.

It is the simplest method for evaluating youth employment programmes, as the impact equals the difference in outcomes between participants and individuals in the control group.

Lottery design requires that one group be excluded from the available services. This may cause ethical concerns and encounter the opposition of policy-makers and programme administrators.

Internal validity requires that randomization is carried out correctly and that it is maintained throughout the evaluation. The credibility of findings may be jeopardized if randomization is not done accurately or crossover occurs between the two groups (treatment and control).

#### >>>>> Exercise

Lottery design for a youth employment programme

The government is planning to launch a large wage subsidy programme targeting all youth aged 20–24 who have unemployed spells of more than six months. The subsidy will provide a full waiver of social security contributions (employer's share) for 12 months. The government would like to pilot-test the programme with a credible evaluation method before scaling-up. The labour source survey of the last quarter shows 120,000 young people aged 20–24 unemployed for at least six months.

You are asked to assist with the design of a randomized trial for the piloting of this new programme. Specifically you are asked to:

- 1. Choose the sample size needed for the evaluation;
- 2. Select a process for the enrolment of individuals; and
- 3. Draw a treatment and comparison group.

#### Check your answer



- 1. Since results have to be credible, a confidence level of 99 per cent should be suggested. This means that the sample should comprise at least 2,456 individuals.
- 2. As the labour force survey lists records both addresses and telephone numbers of households interviewed, it is theoretically possible to contact each young person out of the eligible population.

The easiest way to obtain a final sample of participants is to randomly select 40 per cent over the number needed for the sample out of the 120,000 eligible individuals. This would give around 3,400 names with another 100 in reserve, should they be needed. All 3,400 individuals could be contacted by phone and offered the chance to participate in the programme. Those that accept are then asked to enrol, while those that refuse are kept on a separate list.

Hopefully more than 2,500 individuals will be recruited. Once this is achieved, those who have accepted to enrol are randomly assigned to either the participant or the control group.



# 7-3 Randomized phase-in design

The main difficulty with a lottery design is that a programme aimed at improving the well-being of young people is denied to some for evaluation purposes. This becomes especially problematic when the programme remains available over a longer period. When a programme is rolled out over time, a more acceptable variation of the lottery design is phase-in randomization. This method exploits the operational inflow of participants over different periods of time to construct a treatment and a control group.

The main difference between lottery and phase-in design is that the latter method employs randomization to select the group that will receive treatment first, second, third and so on. Any group that is waiting to receive the programme also constitutes a valid comparison group. The steps to be followed for phase in-design are the same as for lottery design. These are definition of the eligible units, recruitment, selection of the sample, baseline survey, and random assignment to treatment and control (see the following example).

#### >>>> Example

#### Phase-in random assignment

The scenario is identical to that used for the lottery design example.

There are 94,267 young people with primary educational attainment eligible for a vocational training programme to be launched by the NEA. There are only 25,000 training slots per year available in private and public training centres. This indicates that the programme can be rolled out over approximately four years and that a phase-in evaluation design may work well.

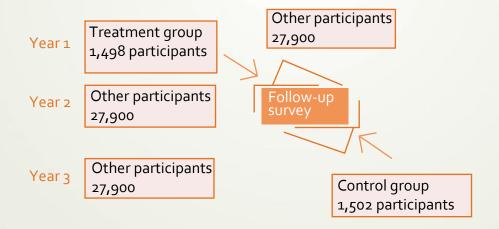
The vocational training programme is advertised in all 35 local offices. By the end of the enrolment process 86,720 young individuals have applied to participate.

The NEA knows that a sample of 3,000 individuals must be drawn in order to carry out the evaluation. Either random sampling or stratified sampling can be used for this purpose (see Step 3 under lottery design). The following two examples illustrate these options. Option A. The NEA decides to centralize the random selection of the sample for evaluation purposes. The final randomization exercise (performed using Excel) assigns 1,498 youth to treatment and 1,502 to the control group.

The treatment group is among the first to be trained, while the control group of 1,502 individuals is wait-listed for training in the third year.

Example <<<<

Phase-in random assignment



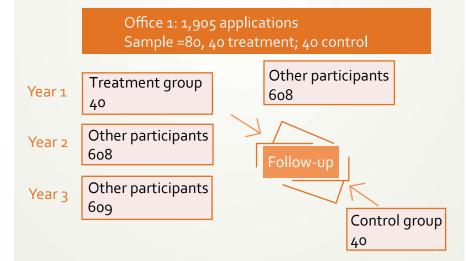
A follow-up survey is carried out at the end of the second year on both individuals in the treatment group (trained at the start of the first year) and the control group wait-listed for training in the third year.

#### >>>> Example

## Phase-in random assignment

cont.

Option B. The NEA decides to decentralize random selection of the sample for evaluation purposes to local offices. Each office is asked to draw randomly from their list of applicants a number of youth (80 for small offices and 90 for large offices). Each office then randomly assigns half to a treatment group and the other half to a control group.



Individuals in the treatment groups are trained in the first year. Individuals in the control group are wait-listed to be trained in the third year. The evaluation is carried out at the end of the second year on the 40 individuals in the treatment group and the 40 individuals in the control group.

#### **Advantages**



Phase-in design provides a credible method for building a counterfactual and is particularly suited for large programmes that need to be rolled out over time.

Differences between treatment and control groups are simple to compute. Since the approach does not exclude anybody from treatment, it is also easily accepted by policy-makers and programme managers.

#### Disadvantages



Like lottery design it may be difficult to ensure that randomization is maintained throughout the period of the evaluation. The credibility of the findings may be jeopardized if randomization is not done accurately or if there is crossover between the two groups (e.g. individuals assigned to the control group cross to the treatment group and receive the programme).

This method is not suitable to measure long-term impact as at a certain point the control group will receive the programme. The follow-up survey, necessarily, measure only short- to medium-term impact.

This approach requires a well-designed roll-out strategy with committed staff checking that individuals do not cross over from the control group. From an operational point of view it may be difficult to implement.

Randomized promotion is appropriate when it is not possible to exclude any potential beneficiary from the programme, either because participation is voluntary – and anyone who wishes may enrol – or because the programme has sufficient resources to serve all eligible individuals.

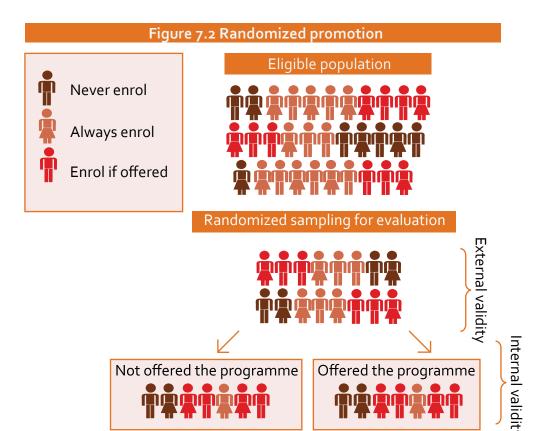
In real life, youth employment programmes only rarely achieve full compliance with assignment criteria (e.g. all those assigned to treatment participate in the programme and all those in the control group do not). Under these circumstances, a comparison of the group originally assigned to treatment with the group originally assigned to control will yield the intention-to-treat (ITT) estimate.

**7-4**Randomized promotion

To estimate the impact of the programme on those who actually take up treatment requires correcting for the fact that some of the individuals assigned to the treatment group do not receive it, or that some of the individuals assigned to the comparison group actually receive it. This is the treatment-on-the-treated (TOT) estimate.

Random promotion is based on the consideration that in many programmes the eligible population comprises three different types of potential beneficiaries. The first is those who never enrol, the second is those who always enrol, and the third is those who enrol only if they are encouraged to do so. The difference between randomized promotion and the other two methods discussed above is that random assignment is performed to select the group that will be offered the programme and the group that will not.

The first step in randomized promotion is to select from the eligible population the sample for the evaluation.



This is done through randomized sampling or stratified randomized sampling to ensure external validity.

The **second step** is to divide randomly the evaluation group into two groups of more or less equal size. This ensures internal validity (i.e. that the two groups have on average the same observed and unobserved characteristics).

The **third step** is to offer one group the chance to participate in the programme (promotion) and then enrol them. The final participants will include a good share of those who were offered the programme, but also some of those beneficiaries who always manage to enrol, even if not offered. Similarly, the control group will include a good share of those who would have enrolled if they were offered the programme and some of those who would never enrol even when offered.

The **fourth step** is to estimate the intention-to-treat (ITT) effect. This is the simple difference between those who were offered the treatment and those who were not (irrespective of participation).

Young vocational graduates encounter difficulties in obtaining their first job because the education they receive is largely theoretical and provides little hands-on experience. As a result, many vocational graduates experience long spells of unemployment, while those who are able to get a job usually earn wages at the statutory minimum. To remedy this situation, the government decided to finance a three-month internship scheme in private and public enterprises for all vocational students attending the third and fourth year of school. This programme is expected to provide them with the necessary work experience to obtain higher entry-level wages. Programme take up is voluntary, as it is organized outside of school during the summer months by the Youth Offices (Ministry of Youth and Sport). The programme will be evaluated experimentally using a randomized promotion approach.

#### Example <<<

Randomized promotion of an internship programme

#### >>> Example

# Randomized promotion of an internship programme

cont.

#### Step 1

The Ministry of Education provides the listing of eligible youth. This comprises all young people enrolled in all national vocational schools by year of attendance (80,000 students). The number attending the second and third year of vocational education comprises 36,000 students (58 per cent boys and 42 per cent girls). The Ministry of Youth and Sport decided to centralize the evaluation process. The sample (95 per cent confidence level) is selected through stratified sampling (boys and girls) from the eligible population of 36,000 students. The final sample required is 2,233 boys and 2,145 girls.

#### Step 2

The final sample for evaluation is 2,700 boys and 2,600 girls. Of these, 2,650 are randomly assigned to treatment, and the other 2,650 are assigned to the control group.

#### Step 3

The Youth Offices are instructed to organize "career days" for two classes (one in the third and one in the fourth year) for each of the country's 53 vocational schools. Each class has an average of 25 students. During these career days, students are informed about the internship programme and are offered the chance to enrol on the same day. At the end of the enrolment process, 90 per cent of students who were offered enrolment show up for the internship (2,385 students, 1,002 girls and 1,383 boys). The control group comprises students from the same schools attending classes that did not receive career days.

#### Step 4

After two years, a follow-up survey was performed targeting individuals originally assigned to the treatment group and those assigned to the control group. The survey found that the average earnings for the 2,650 individuals offered the internship programme were \$110/month, while average earnings for the individuals in the control group were \$70/month. The intention-to-treat estimate (ITT) is \$40 (a simple difference between the income level of the two groups).

The fifth step is to estimate the impact of programme on the treated (TOT) from the intention-to-treat estimate (e.g. \$40 difference in the above example). This difference is not a result of the "never" individuals, as they would not have participated regardless, and their assignment to treatment or control group makes no difference. The same reasoning applies to the "always" group, because they will always find a way to participate even though excluded from treatment. The \$40 difference, then, is due to those who enrol if offered. Once only those that "enrol if offered" are extracted from the treatment and the control groups, the difference between those in the treatment and the control groups will be the TOT estimate. But if 90 per cent of those who were offered treatment actually enrol, this means that the remaining 10 per cent belong to the "never" group. In the control group, the 10 per cent who crossed over to treatment can be easily identified (since they were listed in the control group, but they ended up participating in the programme).

This means that the sample comprises 80 per cent of individuals who "enrol if offered", 10 per cent of "always" and 10 per cent of "never". As the sample was drawn randomly from the eligible population, the two groups are on average the same (e.g. the same ratio 80-10-10 is true also for the control group). This implies that if the average \$40 difference is attributable to 80 per cent of individuals, the impact of the programme is equal to \$40/0.80 or \$50.

To compute the impact estimate on individuals who were treated, the Ministry of Youth and Sport first verify whether there was a crossover between treatment and control. This accounted for approximately 10 per cent of the sample (265 individuals). These are "always" types of individuals. It is likely that the missing 10 per cent from the sample who did not participate in the internship comprise the "never" type (265 individuals). Hence, the individuals who fall into the "enrol-if-offered" category comprise 80 per cent of the sample (or 2,120 individuals).

The impact on earnings on the treated is equal to the difference in earnings (ITT) of \$40 divided by 0.8 (the 80 per cent categorized as "enrol-if treated"). The impact of the treated is then equal to \$50. This estimate is not valid for the whole eligible population as it applies only to the treated.

#### Example <<<<

Randomized promotion of an internship programme

cont.

Inmostyouthemployment programmes, however, participation is voluntary and programme administrators cannot control which individuals participate and which do not. Other programmes are universal (available to all who comply with eligibility criteria, for example a waiver of social security contribution for youth aged less than 25).

In this type of programmes, the eligible population will include the "never" type of individuals, the "always" type of individuals, and those who would participate if they had sufficient information about the programme or were provided with additional incentives to participate (promoted group).

#### >>>>> Example

## Encouragement design

[The scenario is the same as for the previous example. The Ministry of Youth and Sport organizes an internship programme for vocational students during the third and fourth year of school. Steps 1 and 2 remain the same, with a final sample for evaluation of 5,300 individuals (2,650 for participants and 2,650 for non-participants).]

#### Step 3

Youth Offices are instructed to advertise the internship programme in their monthly bulletin. Interested students are asked to visit the local Youth Office and enrol. Youth offices are also asked to contact all those on the list for treatment by telephone, describe the programme, and offer to enlist respondents directly into the programme.

Following this campaign, the Ministry of Youth and Sport calculates that enrolment in the non-promoted group is 30 per cent (e.g. 795 individuals assigned to the non-promoted group actually applied to the programme, i.e. these are "always" types). Among the group exposed to promotion enrolment is 80 per cent (2,120 individuals). Since the two groups have been randomized and are expected to have the same type of behaviour, this means that 795 individuals among the promoted group are the "always" type and 1,325 are the "enrol-if-promoted" type).

#### Step 4

The follow-up survey finds that the average earnings among the whole non-promoted group are \$70 and that the average for all those in the promoted group is \$110. The simple difference between the two groups is again \$40 (ITT estimate).

#### Step 5

The \$40 difference stems from the "enrol-if-promoted" group, which represents 50 per cent of the sample (1,325 persons). This means that the "enrol-if-offered" group will also represent 50 per cent of the eligible population (because the sample for evaluation was drawn randomly). The impact, therefore, will be equal to the difference found (\$40) multiplied by the percentage of "enrol-if-promoted" in the sample (50 per cent), which is equal to \$80.

#### Example <<<<

Encouragement design cont.

#### **Advantages**



Randomized promotion design is a good method to evaluate voluntary programmes as well as those that have universal eligibility.

Eligible beneficiaries decide autonomously whether to participate or not, so nobody is excluded from benefitting from the programme.

Thecounterfactualconstructed with randomized promotion design is robust. Since random assignment is applied to both the drawing of the sample for evaluation (external validity) and to the groups who will be promoted and not promoted (internal validity), it is possible to generalize the impact to the whole eligible population.

#### Disadvantages



The results of this method are valid only for the ITT and TOT and cannot be generalized to the whole population.

This approach often requires larger samples than other randomization methods. This increases the costs of running the evaluation.

**1 2 3 4 5** 

#### >>>> Exercise

# Which randomization approach to use?

For the last five years the PES has been a running a wage subsidy programme for registered unemployed. There are indications of low retention rates among subsidized workers and low earnings, so the PES has decided to redesign the programme. The new version envisages:

- One month (mandatory) on-the-job training;
- Three months subsidized wages (equal to 100 per cent of the minimum wage); and
- An obligation on the employer to retain the subsidized worker for an additional six months after the subsidy.

The PES will manage the programme through 50 outreach offices, which cover the country. Funding is available for approximately 30,000 individuals, half of whom will be youth aged 23–29. There are 70,000 youth aged 23–29 listed on the PES register as unemployed. Each PES outreach office, however, has staff to treat only 200 individuals per year.

The Deputy Director of the PES has asked your assistance to design a robust evaluation for the youth component of the wage subsidy programme, as requested by the government. Your assignment is to help the PES: (i) select which approach to use in the evaluation, including the outcomes to be measured; (ii) draw the sample for the evaluation; (iii) organize enrolment of beneficiaries' in PES outreach offices; and (iv) set the guidelines for the follow-up survey.

#### >>>>>

#### Check your answer

- Given the limited capacity of the PES outreach offices (10,000 places per year to be equally divided between youth aged 23–29 and adults), the programme is suitable for phase-in randomization over three years. Given the information provided in the text, three outcomes should be measured: (i) employment at follow-up; (ii) retention of subsidized workers among partner enterprises; and (iii) level of earnings.
- The eligible population is young unemployed aged 23–29 registered with the PES. The sample to be drawn randomly for evaluation purposes (at 95 per cent confidence level) is 2,414. Adding 20 per cent for non-response at follow-up implies that the sample should comprise at least 2,900 individuals (1,450 for treatment and 1,450 for control).

#### Check your answer

• Outreach offices should be instructed to enrol all 15,000 potential beneficiaries aged 23–29 before the start of the programme. After enrolment is concluded, total applicants are randomly assigned to three groups (year 1, year 2 and year 3). For example, if the programme enrols 14,200 youth, 4,733 individuals will be randomly assigned to year 1, another 4,733 to year 2 and so on. The treatment and control groups are randomly drawn from enrolled youth. The group assigned to treatment will enter the programme in the first year (1,450 youth), while the control group will be shifted to the third year (towards the end). A baseline survey should be carried out on the sample, if the information provided in the PES records is insufficient. Baseline data for the outcomes of interest (employment, retention and earnings) should be zero for both groups, since the eligible population is "unemployed youth aged 23–29".

cont.

The follow-up survey should be run at the end of the second year or the beginning of the third year, before the individuals who are in the control group participate in the programme. If all those in the treatment group participate in the programme first (one month training, three months subsidized work and an obligation to retain for other six months), all participants will complete the envisaged programme and enter the open labour market approximately one year before measurement. For example, if enrolment is done in November 2010, randomized assignment into three groups will occur the following month, with the treatment group receiving onthe-job training in January 2011 (first year). In April 2011, the period of three-months' subsidized employment ends and in October 2011 the obligation on employers ceases. By October 2012 (second year) the treatment group has been in the labour market for 12 months and can now be measured. The follow-up survey should ascertain: (i) how many individuals in the treatment and control group are employed; (ii) the level of average earnings for those employed; and (iii) among programme participants, how many remain employed in enterprises that received the subsidy.

**1 2 3 4 5** 

# 7-5 No-shows, dropouts and crossover

A number of things can go wrong during implementation of a randomized trial design. When this happens, the evaluation design is compromised and statistical techniques need to be used to attempt to fix the problem.

In many voluntary youth employment programmes a fraction of the people assigned to the programme either fail to participate (noshows) or drop out prior to completion. Indeed, completion rates over 80 percent are rare. Failure to anticipate problems caused by no-shows and dropouts constitutes the most common cause of compromised design in evaluations of youth employment programmes.

Another potential problem arises when people assigned to the control group manage to enrol in similar programmes (or in the same programme in different geographical locations). While non-compliance (or "crossover") by members of either group does not invalidate an evaluation design, it does complicate interpretation of the results as it means that the evaluation has to collect data on the actual programme participation rates of both groups.

The validity of a randomized trial design relies on the average equivalence between the treatment and control groups. In most cases this equivalence is compromised when members from one group or the other drop out or are otherwise lost. For this reason, any analysis of a randomized design should be based on comparison of those initially assigned to treatment and control groups, using data on everyone initially assigned to these groups (ITT). The ITT estimate can usually be adjusted after the fact for non-compliance or crossover behaviour of the groups.

No-shows and dropouts have to be included as part of the treatment group. A valid design requires post-programme outcome data for everyone in the group, regardless of whether they actually completed (or even attended) the programme.

Similarly, everyone assigned to the control group must be followed and incorporated into the analysis, including those who participated in the programme at a later date (or participated in similar programmes). This requirement means that the information required to track people for follow-up surveys must be collected at the time of random assignment, since after that point some people inevitably disappear. In short, including no-shows and dropouts as part of the treatment group means that analysis will centre on the intention-to-treat (ITT).<sup>31</sup>

<sup>31</sup> Treatment of the treated (TOT) can be estimated using the instrumental variable approach. For a detailed discussion of this approach, see: Gertel, P.J., Martinez, S., et al. 2011: *Impact evaluation in practice* (Washington, DC, World Bank, 2011).

Randomized control trials

Crossover (from the control to the treatment group) is also common. In such instances, it is important to ensure that enough units not affected by crossover remain, so as to estimate the outcomes of interest. The evaluation, however, will not be able to generalize the findings to the entire population.

Most youth employment programmes are offered in a variety of geographical locations, with variations in the characteristics of potential programme beneficiaries and the quality of services provided. In such cases, programme design needs to allocate adequate resources to monitor implementation at local level. Monitoring is particularly important in a randomized design because the validity of the design depends on proper implementation in each location. In particular, a randomized evaluation has to maintain a balance across sites between the treatment and control groups (i.e. equal fractions of the two groups at each site). This can be done by using cluster-sampling procedures. Cluster-sampling works by using a sample obtained by drawing a random sample of clusters (a cluster is a group of units that are similar), after which all units in the selected cluster or a number of units within each drawn cluster are randomly drawn. In this way, each cluster has the same probability of being drawn and the same applies to units within the selected cluster.

Checks should be built into the implementation process to ensure that managers at different sites actually enrol all those who are assigned to receive treatment at local level and to prevent people assigned to the control group from crossover. It is also important to keep a record of local information for both the treatment and control groups to allow an ex post analysis of the assignment process.

Finally, site variations are important to understand differences in measured programme effects. For example, if an evaluation compares two types of programmes, and one is mainly offered in certain locations, then site information is needed to separate of site-specific effects from any difference of impact of the two programmes. An impact evaluation is therefore necessary for each of the programmes in each site, each with a treatment and a control group.

7.6
Programme enrolment and impact in different locations

## >>>>> Additional reading

Card, D., Ibarrarán, P. Villa, J.M. 2011. Building in an evaluation component for active labor market programs: A practitioner's guide, (IZA working paper No 6085).

**Duflo, E., Glennersterz, R., Kremer, M. 2007.** *Using randomization* in development economics research: A toolkit (Centre for Economic Policy Research, Discussion Paper No. 6059), http://econ-www.mit. edu/files/806

Gertel, P.J., Martinez, S., et al. 2011. Impact evaluation in practice (Washington, DC, World Bank).

Hempel, K., Fiala, N. 2011. Measuring success of youth livelihood interventions: A practical quide to monitoring and evaluation (Washington, DC, Global Partnership for youth employment).

Hussmanns, R., 2004. Measuring the informal economy: From employment in the informal sector to informal employment, ILO Bureau of Statistics, Working Paper No. 53 (Geneva, ILO).

Ibarrarán, I. P. Ripani, L., Villa-Lora, J.M. 2011. Youth training in the Dominican Republic: New evidence from a randomized evaluation design (Preliminary version of a paper prepared for the 6th IZA Conference).

## >>>>> Glossary

## Randomized assignment (lottery) design

With this method, potential programme beneficiaries are randomly assigned to treatment and control groups, with each person having the same chance of being assigned to the programme.

## Randomized phase-in design

In this method, randomization is used to select which group will receive treatment first, then second, third and so on. Any group that is waiting to receive the programme constitutes a valid control group.

### Randomized offering

Although programme managers can randomly select those to whom they offer the treatment, they cannot obtain perfect compliance (e.g. force individuals to participate or refuse participation if somebody insists). The randomized offering method is used, in this instance as variable for actual programme participation.

Randomized control trials

## Glossary < <<<<

cont.

#### Randomized promotion

Randomized promotion is a method similar to randomized offering. Instead of random selection of individuals to whom the treatment is offered, individuals are randomly selected for promotion of the treatment.

#### Intention-to-treat (ITT) estimate

The ITT estimate describes the difference in outcomes between the group to whom treatment is offered (including no-shows and dropouts) and the same indicator for the group to whom treatment was not offered (including those who crossed over to the treatment group).

#### Treatment-on-the-treated (TOT) estimate

The TOT estimate describes the impact of the treatment on those individuals that have actually benefited from treatment.

### Screening and enrolment

Public or private service providers at established locations offer many youth employment programmes. Typically, the evaluation will recruit participants from the regular flow of clients at each location. In some cases, all clients who appear within a given timeframe period will be included in the evaluation. In other cases, only a fraction will be included. The size of the evaluation sample, relative to the flow of new clients at the programme locations included in the evaluation, will determine the length of the recruitment phase. Because only a share of the people recruited to a randomized design is assigned to actually receive the programme, intake for an evaluation may disrupt the normal flow of clients into a programme. This does not pose a problem when there are more applicants than places. In these cases, randomization serves as a rationing device. In instances where the regular flow of recruits is needed to fill programme places, however, programme managers may object to having some of their potential clients allocated to the control group, and may try to override the assignment process. If this is likely to occur, planning for the evaluation should include resources for additional recruitment efforts to increase the flow of new clients and to closely monitor compliance with enrolment rules.





Good to know...

#### **Baseline survey**

Once participants for the evaluation have been screened, they have to undertake the baseline survey and be randomized into the treatment and control groups. Randomization after the baseline survey ensures that respondents are unaware of their programme status (i.e. treatment or control group). This eliminates any concern that this knowledge may influence their answers to the survey. Normally, the survey should collect basic information on individuals in the evaluation (e.g. age, sex, educational attainment, household composition, income and so on), as well as information on the outcomes to be measured in the evaluation (e.g. earnings and employment status during the period just prior to the baseline survey). Such information enables randomization to be verified to ensure that it was correctly implemented. It can also be used to interpret the outcome results.

#### Randomization

The randomization process itself should be designed to meet two key objectives. First, the process must be carefully controlled and documented. Usually the assignment process is conducted through a centralized office using a random number generator (or other randomizing device). Second, the process should be designed such that everyone in the baseline survey has the same probability of assignment to the treatment group. Stratified designs by site are possible, but are more complex to implement and analyse.

There are two simple checks for randomization. First, the baseline characteristics of the treatment and control groups should be compared. Small differences will emerge randomly, and one or two characteristics may appear to differ significantly even by chance. A second useful test, therefore is to fit a logistic (or probit) model for treatment group status using all available baseline characteristics. In youth employment programmes, randomization failures are often due to imperfect compliance with enrolment and/or follow-up protocols. Imperfect compliance will vary by site. Thus, when a failure is suspected, it is useful to examine the data site-by-site and look for patterns that indicate a problem.

#### Treatment and control groups

After randomization, individuals assigned to the treatment group enter the programme and receive the normal services offered to participants. In many cases only a fraction of those assigned to Randomized control trials

the treatment group actually complete the programme, although all individuals initially assigned to treatment should be included in the surveys. Ideally, individuals assigned to the control group are actively prevented from receiving programme services for a period of time (e.g. one or two years) after random assignment. In this instance, names and addresses of control group members are entered into an "embargo list" of people who cannot receive programme services. A complete embargo may be difficult to enforce where members of the control group can apply to other sites. For this reason, the design should take into account and plan for the possibility of "crossover".

#### Analysis of compliance and crossover

The issue of compliance can be resolved via a simple cross-tabulation of assignment status (treatment vs. control group) and actual programme participation. This requires information on the actual programme participation of both groups. With full compliance, assignment status is the same as programme participation. When dropouts, no-shows and crossover are included, however, the intention-to-treat (ITT) effect (the difference in outcomes between the originally assigned groups) needs to be translated into a treatment-on-the-treated (TOT). This is done by dividing the ITT by the difference in participation rates of the treatment and control groups. For example, if 75 per cent of the treatment group received the programme (i.e. 25 per cent dropped out or did not show up) and 25 per cent of the control group received the programme or another similar, then the difference in participation rates amounts to 50 per cent. An adjustment is then performed by multiplying the ITT by a factor of 2.

It is best to define "participation" at the programme design level in terms of completion of a minimum part of the programme (e.g. four weeks for a programme lasting two months). This facilitates the distinction between those who drop out from those who participated.

## Sample sizes

The necessary sample sizes for a randomized evaluation of a youth employment programme are based on a standard power calculation. The sample sizes indicated by a simple power calculation are based on a best-case scenario, and ignore the problems caused by non-response, non-completion by the programme group (i.e. no-shows and dropouts) and crossover. For example, if the expected rate of non-completion by the treatment group is 25 per cent and the estimated non-response to the follow-up survey is 20 per cent, then the sample sizes should





be increased by 25–30 per cent from best-the case calculations. In general, larger sample sizes facilitate subgroup analyses and are also more likely to produce statistically informative comparisons.

#### Filling open programme slots

No-shows and dropouts pose an operational problem to many evaluations because service providers want to keep their programmes full. In a programme that has more potential recruits than available places, programme managers often maintain a waiting list. Places left open by no-shows and dropouts are then assigned to people on the waiting list. If possible, similar procedures should be adopted in an evaluation, allowing openings in the treatment group to be filled from the waiting list. In cases where enrolment into the evaluation sample occurs over several months, the waiting list can comprise newly assigned members of the treatment group. However, this assumes that intake for random assignment can "stay ahead" of the of places avaliable for programme participants. In cases where it is not possible to maintain a list of newly assigned members of the programme group, it is preferable to fill open slots using the regular selection process, rather than taking members of the control group.

## Timing and content of baseline and follow-up surveys

The baseline survey in a randomized evaluation should be conducted just prior to random assignment. The timing for the follow-up survey (or surveys) is less straightforward. Many programme evaluations use a one-year follow-up survey. However, the existing literature on active labour market programmes suggests that the impact of training programmes (classroom training and on-the-job training) tends to emerge two or three years after participation. A similar conclusion has been reached for the long-term effects of programmes targeting welfare recipients. Based on these studies, a post-programme timing of at least two years is preferable for programmes of longer duration. If resources permit, a good option is to conduct a first follow-up survey approximately one year after random assignment, and a second survey two years after random assignment.

Except in very special cases, the programme effect at a single point in time is only a partial measure. A more complete analysis would examine the programme's impact over the entire post-random assignment period. Normally, such an analysis is based on retrospective questions contained in the follow-up survey asking about

## Randomized control trials

employment and earnings in each quarter since random assignment. The design of the retrospective calendar should take account of the fact that participants often enter and leave the programme at different times.

The content of the baseline and follow-up surveys should reflect the objectives of the programme being evaluated. Since most youth employment programmes are centred on improving participants' earnings, the baseline and follow-up surveys should collect information on total earnings per week/month at the time of the survey, along with information on hours of work, number of jobs held since leaving the programme and the characteristics of each job (e.g. earnings, hours, industry, occupation, formality/informality). When the programme is aimed at improving skills, information on the characteristics of the job is particularly important, since it determines whether participants were successful in obtaining jobs with higher skill requirements.

Another key element of analysis is whether the jobs yielded through a youth employment programme form part of the formal economy. Survey questions devised to elicit this information may usefully draw on guidelines published by the ILO on measuring informal employment.<sup>32</sup>

#### Response rates to follow-up surveys

In most cases, it is impossible to obtain complete data from follow-up surveys for all participants originally assigned to the treatment and control groups. If non-responses are entirely random, this does not pose a problem. If the non-response rates vary by group, however, there may be biases in the observed outcomes that differ by group. This bias can be addressed by a simple comparison of mean response rates of the two groups. Although this does not guarantee the lack of a differential response bias, conventional models of response bias show that two samples with the same distribution of characteristics and the same response rate have the same response bias. If this comparison reveals that response rates vary significantly by group, a secondary analysis is necessary to check the validity of the estimated programme impact.



32 Hussmanns, R., 2004. Measuring the informal economy: From employment in the informal sector to informal employment, ILO Bureau of Statistics, Working Paper No. 53 (Geneva, ILO).



# Difference-in-differences evaluation design

8

# Difference-in-differences evaluation design

Learning objectives

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

Identify the key features of difference-indifferences (DD) evaluation design for youth employment programmes; and

Estimate impact under DD designs

Learning exercises

Estimating the impact of career guidance services.

Difference-indifferences evaluation design

Randomized assignment, as explained in the previous module, produces credible estimates of the counterfactual through explicit programme assignment rules. When programme assignment rules are less clear and randomized methods are unfeasible, the difference-in-differences (DD) approach offers a good statistical tool to estimate the impact of a programme. This approach is quite common and is often used together with matching methods (see Module 9) to refine the assumptions that need to be made about the treatment and comparison groups.

The DD method compares the changes in outcomes over time between the treatment and the control group to obtain a better estimate of the counterfactual. As mentioned in Module 6, simply observing the before-and-after change in outcomes among participants to a programme, or comparing the outcomes of those who enrolled against the outcomes of those who did not enrol, will not yield casual impact. Many other factors – aside from the programme – may be responsible for the observed change, and the treatment and control group may be substantially different (selection bias). The DD approach combines these two methods by taking into account: (i) factors that are constant over time by comparing the before-and-after situation of the treatment group (the first difference); and (ii) factors that change over time by comparing the before-and-after change in outcomes for a group that did not enrol in the programme (the second difference).

The DD method measures outcomes in the group that receives the programme and the control group both before and after the programme. The method does not need the specification of the rules on the basis of which the treatment is assigned. Treatment and control groups do not necessarily need to have the same pre-intervention conditions. But for the DD design to be valid, the comparison group must accurately represent the change in outcomes that would have been experienced by participants in the programme.

8.1
Introduction

8.2
DD Evaluation designs practice

## >>>> Example

DD approach:
New Deal
for young
unemployed
in the United
Kingdoms

The "New Deal for young unemployed" in the United Kingdom was designed to shift young individuals aged 18–24, who had been unemployed for six months and were receiving the jobseeker allowance, into work and away from welfare. The programme combined initial job search assistance followed by various subsidized options including wage subsidies, temporary government jobs and full-time education and training. Prior to this programme, young people could, in principle, claim unemployment benefits indefinitely. With the New Deal programme, after six months of unemployment, young people enter the "Gateway", a period of mandatory intensive job search assistance.

The programme was launched across the whole country in April 1998 (the "National Roll-Out"). The programme was preceded, however, by a pilot phase (Pathfinder) implemented from January to March 1998, in 12 areas.

The evaluation focused on the change in transitions of the unemployed claimant count to jobs during the first four months of the "Gateway" period. Instead of comparing exposed and non-exposed areas during the pilot period, the researchers used the DD approach to compare eligible and non-eligible individuals following implementation of the programme in pilot areas. Pre-and post-outcomes of beneficiaries of the Gateway component of the New Deal were compared to those of ineligible individuals "similar" to the eligible ones, that is, those unemployed for over six months aged just over 24 living in treatment areas. The choice of this group as comparison made it most likely that their overall characteristics and behaviour matched that of the treatment group, in other words, that the growth rate of employment for the two groups would be similar in the absence of the programme.

A common trend assumption was applied to the two groups in the pilot areas, namely that a comparison in the growth rates between eligible and ineligible individuals would provide an estimate of the impact of the programme on those eligible. To strengthen this assumption, the researchers further refined the control group by matching the recent history of outflows from unemployment in the treatment and control groups, to make them as similar as possible.

Difference-indifferences evaluation design

The evaluation found that the outflow rate to jobs had risen by about 20 per cent for young men as a result of the New Deal. This amounted to 5 percentage points more men finding jobs in the first four months of the New Deal above a pre-programme level of 26 percentage points.

The first requirement for the DD approach to work is the availability of good quality baseline data for both the treatment and control group before the start of the programme. Without this, it is not possible to use DD approaches and other methods need to be explored. Typically, the DD approach works backward, namely it tries to identify a valid control group for participants already enrolled in a programme when there is good longitudinal data (many units of observation over multiple periods of time) for both groups.

The **first step** in DD approaches, therefore, is to verify the type of data available. For a credible DD evaluation, data should include at least one year (and preferably more) of pre-programme figures for the participants and comparable data for the control group. Usually, household-based surveys (like the labour force or household budget surveys) are used as data sources because they provide comparable figures over time. At least three sets of observations are needed: prior to the programme (one for participants and one for the control group), and one after the programme (for both participants and non-participants). Non-participants can be randomly sampled, when there is a full list of comparable units.

## Example <<<<

#### cont.

Source: Blundell, R., Costa-Dias, M., et al. 2004. "Evaluating the employment impact of a mandatory job search program", in Journal of the European Economic Association, pp. 569–606, http://discovery.ucl.ac.uk

## >>>> Example

evaluation design in practice: work-training contracts Since 2010, the Ministry of Workforce Development has been running a work-training programme to ease the transition of youth from school to work. The programme, administered by regional authorities, targets youth who have graduated by at least six months (4,000 individuals at the start of the programme) with the aim of increasing their employment and earnings levels. Participants are placed with private enterprises for a period of on-the-job training (three months) followed by subsidized employment (another three months). As the programme is based on the willingness of enterprises to recruit participants from the onset, two regions in the country decided to offer the programme at a later date. After two years, the Ministry decided to measure the impact of the programme with a DD evaluation design.

### Step 1: Checking the availability of data

For the last five years the country has been conducting the Labour Force Survey (LFS) twice a year (in April and October). The LFS collects information on 220,000 individuals aged 15–65. The LFS has a panel design whereby one group of individuals remains in the sample for four observation rounds before being dropped from the survey. The rounds of 2010 and 2011 included 3,640 observations for youth aged 18–24 who had graduated at least six months before. Of these, 40 per cent were unemployed (1,456 individuals) when the programme started (May 2010), 50 per cent were employed (1,820 persons) and 10 per cent (364 individuals) were inactive. As the LFS population framework is constructed randomly on updated census figures, the panel dataset is fully representative of the population of interest.

The records of the National Insurance Fund contain individual information for 7,000 employed youth. Of these, 10 per cent had graduated at least six months before in May 2010. However, since the rate of youth employment in the informal economy is estimated to be over 23 per cent, this dataset is not sufficiently reliable to provide the data needed for the evaluation.



The **second step** is to select a potential control group. Most evaluations begin with a potential control group, and then impose additional restrictions to select a final group that better matches the characteristics of programme participants. For example, a potential control group may include people interviewed by the labour force survey in a month close to the date that participants were recruited for the programme. This potential group may then be further restricted by imposing age, labour force status or location restrictions. There are four common methods to define a potential control group:

- **1. Location:** in this case the potential control group consists of people who are similar to the programme participants, but live in areas where the programme is unavailable.
- **2.** Time: the potential comparison group comprises similar people observed at different times either before or after the programme was implemented.
- 3. Programme eligibility: the potential control group is composed of people from the same geographic area and time period who are as similar as possible to the programme participants, but are not eligible to participate.
- 4. Non-participation: here, a potential control group consists of people from the same area and time period who could have participated, but were either unaware of the programme or selected not to participate.

In youth employment programmes, comparisons based on geographical location are often the most compelling choice, as labour market outcomes of similar groups are often found to be similar in different locations, once fixed differentials among sites are accounted for. For instance, the average earnings of young men who live in different regions or cities of a country tend to be quite similar, apart from geographic differences that are relatively stable over time. In addition, geographic mobility among disadvantaged individuals tends to be low. Few people will move to another city (or region) just to take advantage of a programme offered there. This means that the behaviour of a comparison group of non-participants from areas where a programme was not offered is likely to be free of selection biases associated with the decision not to participate in the programme. Such an "unselected" control group greatly simplifies a DD design.

In some situations eligibility for a programme is limited to people with certain individual characteristics (like age or household status). In these cases it may be possible to design a control group with people who are "almost eligible". For example, if a programme is restricted to people under the age of 24, people between the ages of 25 and 29 would be an obvious comparison group. To be valid, a comparison group of "ineligible" people has to be based on fixed and readily verifiable characteristics. Eligibility rules that are based on a maximum education or income level, for example, could be easily under-reported leading to a situation where some of the ineligible group actually participate in the programme. Lacking strong enforcement of such eligibility rules, these criteria for selecting the control group may be weak.

The weakest basis for a DD evaluation design is a comparison group made up of people who could have participated, but did not for reasons unknown. The problem is that both the participant and the control group are self-selected (i.e. one decided to participate, the other did not). This implies that there are substantial differences (unobservable characteristics) between the two groups leading to a selection bias.

## >>>> Example

#### Step 2: Selecting a potential comparison group

The fact that programme implementation was postponed by some regional authorities permits the selection of a potential comparison group. These are selected from among young individuals (aged 18–24), who graduated six months previously or more (eligibility rules), at the time the programme started (time) from regions where the programme was not offered (location).

Note: At this level of disaggregation, there is the possibility that the number of "real" observations of the LFS will become too small. However, labour force surveys usually have a system of weights that allow representation of the whole sub-population group under study. Regardless, if the number of LFS observations is too small, a baseline survey can be carried out at the time of the evaluation (based on the listing of graduates in March 2010) that asks retrospective questions (namely, what was the labour market status of the respondent in March 2010).



The **third step** is to refine the selection of the control group. At this stage statistical matching procedures are used to select members from the potential control group whose profile prior to the programme matches as closely as possible that of the programme participants.<sup>33</sup> Such refinement, however, is not always necessary. For example, the participants to a youth programme aimed at easing first entry into the labour market will not have significant earnings prior to the programme. A control group of similar young people drawn from other regions will also show no significant earnings.

## Step 3: Refining the comparison group

The programme does not distinguish between secondary and tertiary education graduates for participation. However, these two sets of young individuals may have quite different labour market experiences. To refine the control group, therefore, young people in the regions where the programme was not offered are divided into secondary graduates and tertiary graduates. The same is done for participants.

Every participant with secondary school graduation is then matched as closely as possible to a member of the secondary graduation control group, and the same is done with tertiary graduates. Criteria may include age and sex, length of unemployment spell (6–12 months and 12 months and over), prior work experience (yes/no and length of experience in months) and educational stream (vocational and scientific vs. general stream).

Once a valid control group has been established, the **fourth step** is to measure the outcomes of interests for participants and the control group before and after the programme. Finally, the difference found in the outcome for the control group is deducted from the difference found for the outcome of participants.

33 Instead of finding pairs in treatment and control groups when there are many observables to take into consideration, propensity score matching can be used to calculate the likelihood (propensity score) of each youth enrolling/not enrolling in the programme based on observable characteristics. In other words, propensity score matching uses the propensity score to provide a single measure of the set of characteristics that influence the probability of participating and employment.

## Example <<<<

## >>> Example

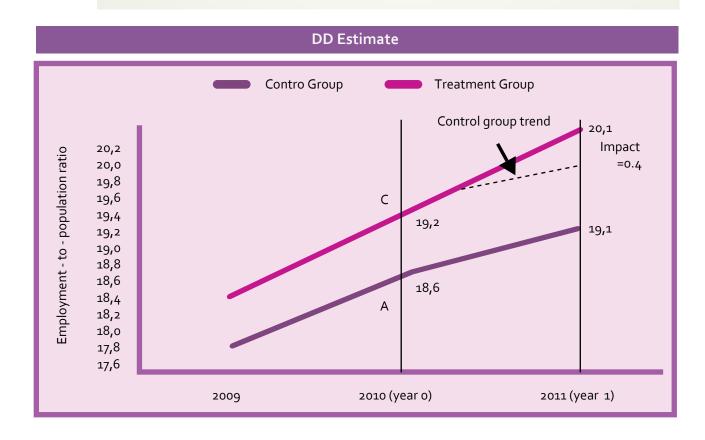
### Step 4: Measuring the impact of the work-training programme

The objective of the work-training programme is to increase employment and the earnings of participants. Therefore, these are the two outcomes that need to be measured for participants and the control group before and after the programme.

The April 2010 labour force survey data, disaggregated by region, indicate that the average employment-to-population ratio for youth aged 18–24, who graduated at least six months before in regions that did not offer the programmes, was 18.6 per cent (point A in the figure below). In regions where the programme was offered the ratio was 19.2 per cent (point C). In April 2011, a follow-up survey is run on participants and the control group and finds an employment-to-population ratio of 20.1 per cent (point D) and 19.1 per cent (point B), respectively.

The impact of the programme is calculated as:

$$=(B-A)-(D-C)=(19.1-18.6)-(20.1-19.2)=0.4$$





The underlying assumption in DD approaches is that, without treatment, the outcome of interest would increase at the same rate in both the treatment and control group (equal trends in the absence of the treatment). A good method for verifying this assumption is to compare the changes in outcomes for both groups at multiple times (at least twice) before the programme starts. This is why longitudinal data are required. If this is not possible, the DD estimate will be biased.

A second way to verify the validity of results is to estimate the difference-in-differences impact with a fake control group, namely one that is not affected by the programme. In the above example, one could verify whether the outcomes for graduates aged 25 (that are not affected by the programme) are similar to those measured with the real control group.

## **Advantages**



Difference-in-differences design provides a simple method of accounting for differences between participants and non-participants.

It does not require a prospective evaluation, if secondary data sources (either household-based or administrative data) are already available.

The counterfactual constructed with DD designs can be made more robust, by combining it with other non-experimental methods (such as matching or regression discontinuity designs).

### Disadvantages



evaluation design

The results of this method are less reliable than those produced by randomized assignment. Even when DD methods are combined with other approaches, the impact may be due to the differences between the characteristics of participants and non-participants rather than the programme.

It requires at least three data collection exercises, while other methods require only two.

## **Additional reading**



Blundell, R., Costa-Dias, M. et al. 2004. "Evaluating the employment impact of a mandatory job search program", in Journal of the European Economic Association pp. 569–606, http://discovery.ucl.ac.uk

>>>>>

cont.

Card, D., Ibarrarán, P., Villa, J.M. 2011. Building in an evaluation component for active labor market programs: A practitioner's guide, IZA working paper No 6085.

Gertel, P.J., Martinez, S. et al. 2011. Impact evaluation in practice (Washington, DC, World Bank).

**Hempel, K., Fiala, N. 2011.** *Measuring success of youth livelihood interventions: Apractical guide to monitoring and evaluation* (Washington, DC, Global Partnership for youth employment).

## >>>>>

## Glossary

#### Longitudinal data

Data in which many units are observed over multiple time periods.

#### Regression

In statistics, regression analysis includes any technique for modelling and analysing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. In impact evaluation, regression analysis helps to understand how the value of the outcome indicator Y (dependent variable) changes when the assignment to treatment or comparison group P (independent variable) is varied, while the characteristics of the beneficiaries (other independent variables) are held fixed.



#### Data sources

There are three basic approaches to compiling the necessary data for a DD design. The first, widely used in developed economies, relies on administrative data from the payroll tax system. In countries with a small informal economy this is a simple, cost effective way to obtain reliable longitudinal data. In most emerging and developing countries, however, only a fraction of jobs are in the formal economy. For these reason, DD evaluations in emerging and developing countries usually rely on an ad hoc baseline and follow-up surveys (as those are used in randomized designs), or on a combination of ad hoc and existing surveys (e.g. using an existing survey as the baseline survey for the comparison group, then administering a special baseline survey to members of the participant group, and conducting a follow-up survey for both groups).

8

Difference-indifferences evaluation design

## Timing and content of the surveys

If *ad hoc* surveys are used for both groups the timing and content of these surveys are broadly similar to those for a randomized design. For a DD analysis the collection of similar outcome data in the baseline and follow-up surveys is critical, since the change in outcomes is used to build the impact estimates. The two surveys should include identical questions on labour market outcomes (employment and earnings), but also on labour market history through retrospective questions. In collecting baseline data for a potential control group, it may be necessary to survey a relatively large number of individuals to obtain a useable sample. This is the advantage of using an existing labour force survey as the baseline for the control group. The control group can be drawn from respondents who match the eligibility rules for the programme.

Where an existing survey is used, it is important to ensure that the questions on the baseline survey for the programme group closely match the questions in the existing survey. Even minor changes in question wording can lead to important differences.

Comparison of pre-programme outcomes for participant and control groups

The most important check for a DD design – and the only way to evaluate the reliability of the selected control group – is to compare the pre-programme outcomes of the participant and control groups. Whenever possible this should be conducted using the same outcome variables that form the main focus of the evaluation (e.g. earnings and employment rates). The comparison begins with a simple table of mean outcomes for the two groups in each quarter or month prior to the date when participants entered the programme. The analysis can be extended by plotting (or tabulating) regression-adjusted outcomes for the two groups. Ideally the outcomes of the two groups will follow a parallel path (i.e. a constant differential between the groups) during the pre-programme period. When this is not the case, the design is potentially compromised, since it is no longer clear that the difference between the groups would have remained constant.



Good to know...



#### Response rates to follow-up survey

Similarly to experimental design, the issue of potentially differential response bias can be addressed by a simple comparison of mean response rates of the two groups.

#### Non-compliance/crossover

The issue of non-compliance in a DD design is addressed by a simple comparison of programme participation rates between the participant and control groups. Information must be available on the programme participation of both groups. If there is significant crossover, the basic impact estimates from the evaluation will be interpreted as intention-to-treat (ITT). When there is significant non-compliance, the DD estimate of the ITT effect can be converted into treatment-on-the-treated (TOT).

### Testing the DD design

It is better to perform the difference-in-differences analysis using several plausible control groups to see whether the estimates are similar. For example, the DD may be tested against a control group known to be unaffected by the programme (which should give zero impact).

## >>> Exercise

the impact of career guidance services

In 2005, a technical assistance project financed by the European Union introduced career-counselling services for young people aged 15–18 in 20 of the 40 youth offices of the country. The Ministry of Youth continued this activity after the project ended in 2007. The Ministry now wants to measure whether career guidance had any impact on the employment prospects of young participants. They decided to carry out measurement on those who received counselling services in 2010 (2,000 young people). From 2008 onwards the youth offices collected data on young people every time they used one of the available services. The client database is quite extensive and includes personal data (e.g. age, sex, education, employment history) and household information (e.g. composition, income, social welfare status). Based on the information provided, how would you proceed to evaluate the impact of the programme with a DD design?

## Check your answer

<<<<<

- The first step is to verify the availability of the information needed to measure the outcome of interest on participants (e.g. employment) prior to programme participation. Since it appears that the data collected by the youth offices contain multiple observations on the same beneficiary, and also information on labour market status and employment history, these administrative records can be used for the pre-programme baseline of participants. If data are incomplete, then a retrospective survey has to be run to ask participants their labour market status at two points in time prior to programme participation (so as to establish the employment trend).
- Young people aged 15–18 in 2010, that received any type of service from the 20 youth offices where career guidance was not available represent the potential control group. For this group, at least two sets of observations, made at different times, will also be needed to establish the employment trends.
- To refine the potential control group, an attempt can be made to match every one of the 2,000 participants receiving career counselling services with members of the control group as closely as possible (age structure, sex, level of education, employment history and so on).
- Once the control group has been refined, the employment-to-population ratio is calculated for participants and the control group in 2009 and 2010 to establish the trend and set the baseline.
- A follow-up survey is run on both those who received the service in 2010, and those who did not, so as to measure their employment-to-population ratio after the programme and to ascertain the impact of the programme, namely:
- = (employment of participants in 2011 employment of participants in 2010) (employment of control in 2011 employment of control in 2010).
- If there are doubts regarding the validity of the control group, this can be verified through comparison with another control group to check the similarity of the estimates (e.g. a group of young people aged 19 who used any type of youth service in the same period in any youth office).





- 9.1. Regression discontinuity design (RDD)
- 9.2. Matching methods

Learning objectives

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

List the main feature of regression discontinuity (RD) and matching evaluation designs; and

Distinguish simple from propensity score matching.

None
exercises

9

Regression discontinuity design (RDD) is another quasiexperimental method to build the counterfactual in programmes that have a continuous eligibility index with a clearly defined cutoff point, such as age (e.g. a youth employment programme that provides vocational training to unemployed youth aged over 20). The assumption is that individuals just below the eligibility threshold (e.g. those aged 19) are not much different from those aged 20. The regression discontinuity measures the difference in post-intervention outcomes (e.g. a decrease in unemployment) between individuals near the eligibility cut-off. Individuals who were too young to enrol in the programme constitute the comparison group and generate an estimate of the counterfactual outcome for those young people that participated. If these two groups of individuals were similar at baseline and are exposed to the same set of external factors over time, the only plausible reason for the different outcomes must be the programme itself. This approach allows impact estimates without excluding any eligible individual. However, the estimated impact is only valid for individuals in the vicinity of the eligibility cut-off score.

When planning a RDD evaluation, the **first step** is to verify that reliable pre-programme data are available on the individuals above and below the threshold established by the programme for the outcomes to be measured. If such information is not available, it is necessary to carry out a survey with retrospective questions to identify the employment history of participants and non-participants.

**9.1**Regression discontinuity design (RDD)

The 2008–09 economic crisis considerably worsened the employment prospects of youth (aged 15-24) in the country. The government, therefore, decided to increase spending on active labour market programmes. The Youth Return to Work Act envisages that youth aged 20 and above be offered a vocational training or work experience opportunity within the first three months of unemployment. The legislative framework offers an opportunity to use RDD methods to evaluate whether the programme succeeded in decreasing the duration of unemployment for youth aged 20 and over.

Example

<<<<

RDD: Using age for eligibility ranking

## >>>> Example

## RDD: Using age for eligibility ranking

cont.

#### Source:

>>>> Example

Adapted from Hämäläinen and , K., Tuomala, J. 2007: Vocational labour market training in promoting youth employment (Helsinki, Government Institute for Economic Research).

#### Step 1

As the programme targets unemployed individuals and is managed by the Ministry of Labour, the administrative data of the employment service are screened to verify that the dataset contains all the information needed to compute the outcome of interest (decrease of youth unemployment). The records of the employment agency contain detailed information about individual characteristics (sex, age, education level, work experience). Unemployment records are cleaned monthly against the records of the Heath and Injury Insurance Fund, which collect data on registered workers. As informality is low in the country, these records can be used to verify whether participants and non-participants are employed at follow-up.

The **second step** is to verify whether the discontinuity holds, namely whether the eligibility ranking was really applied at the cutoff point.

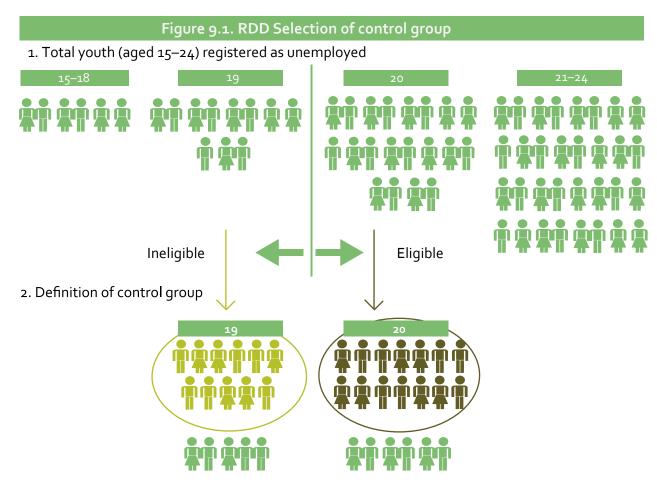
### Step 2

As participants' records are maintained by the employment services, it is necessary only to cross-tabulate the age of unemployed participants in the programme. If the cut-off point really holds at the age of 20 and over, a break in the share of participants with that age should be visible, as show in the figure below:



9

The third step is to select the sample for the evaluation. Out of the total population, those who were ineligible at the time of programme and are nearest to the cut-off point are selected as the control group (if the total number if too large a sample can be extracted). The same is done with participants. A sample of participants who were 20 at programme start are selected to minimize differences between participants and control group (e.g. those who were aged 19 at the time the programme started). It should be kept in mind that the greater the distance between the control group and participants in the ranking scale, the higher the likelihood that the groups are different.



The **fourth step** is to ascertain the outcome of interest after the programme for both participants and the control group, so as measure the difference. In the example of the vocational training programme above, this would be lower unemployment levels among participants, as compared to individuals in the control group.

## Advantages



The RDD takes advantage of existing eligibility rules to construct a valid counterfactual, and does not require the exclusion of eligible individuals from programme participation.

Good pre- and post-programme data are required to check that discontinuity has a clear cut-off point.

## Disadvantages



The results of this method are reliable, but only for those near the cut-off point. The further away participants and non-participants are from the cut-off point, the less reliable the results. Hence, this method cannot be generalized to the whole population.

The method requires large evaluation samples as only those around the cut-off are measured. If the programme targets a restricted group of the youth population, it is likely that the sample will have a number of observations too low to be valid.

When eligibility rules are not strictly enforced, a selection bias may occur. This means that the difference found in the outcome of interest is due to differences between the two groups rather than the programme.

## 9.2 Matching methods

This approach can be applied with almost any programme assignment rules, so long as a group exists that has not participated in the programme. Matching methods typically rely on observed characteristics to construct a comparison group. Therefore the method relies on a strong assumption that there are no unobserved differences between participant and control groups associated with the outcomes of interest. Because of such an assumption, matching methods are normally used in combination with other evaluation approaches.

The matching method pairs young individuals who participate in a programme with individuals who do not participate on the basis of observable characteristics (e.g. age, sex, level of education, employment history, household income and so on). The more characteristics that can be gathered, the more exact the match. The data that need to be collected, therefore, are quite extensive and can substantially increase the cost of the evaluation.

9

Instead of finding matches in the participant and control groups, when there are many variables to take into consideration, propensity score matching can be used to calculate the likelihood (propensity score) of each youth enrolling/not enrolling in the programme based on observable characteristics. In other words, propensity score matching uses the propensity score to provide a single measure of the set of characteristics that influence the probability of participating in the programme. Whereas matching pairs on each characteristic leads to problems with the amount of data, the propensity score reduces the problem to a single dimension. Once the propensity score is known (it is a number between o and 1), participants are matched to non-participants that have the closest score. This latter group becomes the control group. Once the outcome of interest has been measured, the impact of the programme is the difference found between each participant and his or her match.

This method is particularly popular for the evaluation of programmes managed by public employment services. Tipically, participants self-select into the programme, and only a fraction of those who are eligible can be served. The evaluation is designed when the programme has already started. Since the administrative data of the PES usually encompass many observable characteristics, this makes this method rather cost-efficient.

The job subsidy Special Youth Employment and Training Program (SYETP) was introduced in Australia with the aim of improving the transition of youth to work. In 1984, the SYETP comprised a flat rate subsidy of AUD75 a week, paid to employers for 17 weeks. This was equivalent in value to half the average teenage wage. It was available to youths aged 15–24 who had been claiming unemployment benefits and not studying full-time for at least four of the preceding 12 months. The impact of the programme was evaluated using propensity score matching. The following factors affected the SYETP propensity (considering only statistically significant coefficients):

## Example <<<<

## Propensity score matching:

The Australian Special Youth Employment and Training Program (SYETP)

## >>>> Example

## Propensity score matching:

The Australian Special Youth Employment and Training Program (SYETP)

cont.

- Age had a negative impact on participation, with older youths less likely to take part. Health conditions reduced the chance of participation, as did living in a country town.
- Those with 12 years old of schooling and those with longer unemployment spells were more likely to take part in SYETP, as were individuals referred to another available labour market programme.

A probit model was used to estimate the propensity score for each participant in SYETP. The same was done for those who did not participate. Members of the treatment group whose propensity score surpassed the distribution of propensities of the control group were discarded. Individuals were then identified from among the remaining control group members on a case-by-case basis, based on their similarity to an individual of the treatment group (one-to-one nearest-neighbour).

The evaluation found positive impacts of SYETP on participants, although some were not statistically significant, due in part to the small sample.

**Source**: Knight, G.M. 2002. Evaluation of the Australian wage subsidy Special Youth Employment and Training Program, School of Economics and Political Science, University of Sydney, www.psi.org. uk/docs/2004/genevievePhDfull.pdf

## Advantages



Matching allows for the comparison of outcomes between similar individuals and produces a valid counterfactual, if used in conjunction with other evaluation methods (like for example RDD).

## Disadvantages



A large sample survey may be needed to draw an appropriate comparison group. In addition, the statistical techniques needed to draw matches may be complex.

It can be performed on observable characteristics only, to the exclusion of unobservable characteristics. Since this is a very strong assumption, the use of this method alone to evaluate a youth employment programme can be easily challenged.

It may be impossible to find an exact match for each individual in the treatment group, endangering the external validity of the impact estimate.

9

### External validity in propensity score matching

In a propensity score matching all participants need to be successfully matched to a non-participant to ensure external validity. In the middle of the propensity distribution, matches are relatively easy to find because enrolees and non-enrolees have similar characteristics. However, individuals with propensity scores close to 1 cannot be matched to any non-enrolees with similar propensity scores. Individuals highly likely to enrol are so dissimilar to non-enrolling individuals that a good match cannot be found. This is called "lack of common support" of the distribution of propensity scores.



Good to know...

#### Baseline and follow-up surveys

For propensity score matching, representative and comparable surveys are needed to identify individuals that enrolled in the programme and those that did not. The two samples need to be pooled before estimating the probability of each individual enrolling in the programme, based on individual characteristics. The sample is then restricted to units for which common support appears in the propensity score distribution (i.e. those in the middle, excluding those at the end of the two ranges). Matching must be done using baseline characteristics. Finally, the matching method is only as good as the characteristics used for matching, so the presence of a large number of background characteristics is crucial.

## Additional reading

1111

**Gertel, P.J., Martinez S., et al. 2011.** *Impact evaluation in practice* (Washington, DC, World Bank).

Hämäläinen, K. Tuomala, J. 2007. Vocational labour market training in promoting youth employment (Helsinki, Government Institute for Economic Research).

**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).

**Knight, G.M. 2002.** Evaluation of the Australian wage subsidy Special Youth Employment and Training Program, School of Economics and Political Science, University of Sydney, www.psi.org.uk/docs/2004/genevievePhDfull.pdf



Timeframe, resources and dissemination of findings

10

# Timeframe, resources and dissemination of findings

Learning objectives

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

Plan the timeframe and resources to carry out an impact evaluation of a youth employment programme; and

Manage the dissemination of evaluation findings.

Learning exercises

None

Timeframe, resources and dissemination of findings

10

As mentioned in the previous modules, impact evaluations should be planned early in the programme design. The evaluation proper, conversely, needs to be carried out once the programme is mature. There is no "right" time for the measurement of results, as timing is context- and programme-specific. However, a number of considerations can guide the decision of when to collect follow-up data.

First, the impact evaluation needs to be fitted to the youth employment programme implementation cycle and be aligned to its expected duration (including implementation lags). A robust monitoring system will inform the impact evaluation about the pace of the intervention.

The timing of follow-up data collection must take into account how much time is needed after the programme is implemented for results to appear. The programme results chain helps to identify outcome indicators and the appropriate time to measure them. Some programmes (e.g. community-based works) aim to provide short-term benefits, whereas others (e.g. training programmes) aim for longer-term gains. Moreover, certain results by their nature take longer to appear (e.g. changes in skills levels from a vocational training programme) than others (e.g. earnings from a wage subsidy).

Some evaluations will collect follow-up data while the programme is still being implemented, so as to measure short-term changes and maintain contact with the evaluation sample to reduce attrition. For programmes without continuous operations, additional rounds of follow-up data collected after programme completion can help to measure longer-term changes. Follow-up data can be collected more than once, so that short- and medium-term results can be compared. Follow-up surveys that measure long-term outcomes after programme implementation often produce the most convincing evidence regarding programme effectiveness, as they also reflect unintended effects. The identification of longer-term impacts can nevertheless create difficulties. Impact may simply vanish with time or be undermined when the control group benefits from spillover effects from programme beneficiaries.

10.1
Timeframe and financial resources for an impact evaluation

## **1 2 3 4 5**

### Terms of reference

The timing of an evaluation must also take into account when certain information is needed to inform decision making and must synchronize evaluation and data collection activities to key decision-making points (e.g. decisions on budgets and programme expansion).

The cost of conducting an impact evaluation must be compared to the opportunity costs of potentially running an ineffective programme. In this sense, the resources needed to implement an impact evaluation constitute a significant investment (typically ranging from 2 to 15 per cent of the total cost of a youth employment programme). In most evaluations, the highest cost is new data collection, often accounting for over 60 per cent of overall costs. The relative cost of evaluating a pilot programme is generally higher than that of evaluating a nationwide or universal programme. In addition, some evaluations require only one follow-up survey (or may be able to use existing data sources), whereas others may need to carry out multiple rounds of primary data collection.

Most youth employment programmes contract the design and implementation of impact evaluations to one or more experts, who are also able to provide advice regarding the selection of the survey firm for data collection. There are a number of expert rosters on impact evaluation that can be consulted to locate the expertise needed. However, both experts and survey partners need to be guided in their work by clear and comprehensive terms of reference. These should at a minimum include:

- 1. Programme objectives: the more background that can be supplied on the programme objectives, the better. The programme outputs and their relation to outcomes should also be described in detail.
- 2. Programme approach: it is helpful for the expert to know how the programme was implemented (e.g. agencies involved,
- 34 For a list of such rosters and also a guide to select survey firms, see 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).
- 35 Tool 3 appended to this learning material offers sample terms of reference for contracting expertise. The online resources mentioned above also offer a template of terms of reference for conducting impact evaluation.

10

decisions taken and their timeline). If there are various programme components, each should be summarized including its design and timeline, as well as any reasons that delayed/affected implementation and any variations in interventions. This section of the terms of reference usually lasts several pages, and comprises some of the most valuable information for the evaluation team.

- 3. Programme impact evaluation: it is necessary to explain the exact task which the expert will perform, namely which programme component(s) will be evaluated, and specify the research questions. There is no need to describe the methods to be used, but rather which data are available, so the expert can decide which method is most appropriate.
- 4. Data: this is probably the most important section. It needs to explain, in detail, which data are available from the programme data collection system and outside sources. Sources usually comprise the programme monitoring and accounting system, national household surveys and censuses. Detailed descriptions of the data are necessary (content, time series and level of disaggregation). If there are not sufficient data to meet the proposed analysis, and additional figures need to be collected, this section should specify the proposed content, timeframe and procedure of collection.
- 5. Work schedule: terms of reference typically lay out the required person/months of work and the key deliverables of the evaluation (e.g. interim and final report). For an impact evaluation, it makes sense to give the expert time to analyse the data and to request the presentation of a detailed research proposal as a first deliverable. In this way, the agency contracting the work can provide feedback on the proposed research and learn from the procedure of analysis. The last deliverable is the presentation of findings in a structured report.

**1 2 3 4 5** 

Production and dissemination of findings

The evaluation exercise will produce – aside from fully documented datasets – a number of reports, including the baseline report, the impact evaluation report and policy briefs that summarize the key findings.

The main objectives of a baseline report are to assess whether the selected impact evaluation design is valid and to describe the baseline (pre-programme) characteristics and outcomes of the eligible population. It also generates information about the programme and its beneficiaries. The baseline report should include a description of the programme (including eligibility criteria and benefit assignment rules), the objectives of the evaluation (including core policy questions, hypotheses and indicators) and the evaluation design. As assignment to the programme is normally done just after completion of the baseline survey, it is good practice to include such information in the baseline report. The section on sampling usually outlines the sampling strategy, power calculations, how data were collected and their quality (including non-response rates).

The final impact evaluation report is the main product of the evaluation. It presents the results and answers the policy questions set out. The report discusses the validity of estimates of the counterfactual and the impacts that are fully attributable to the programme. It includes a detailed description of the data analysis and econometric specifications.

Since the final report is the main evaluation output, it should incorporate the key information reported in the baseline report (rationale and description of the intervention, objectives of the evaluation and so on), present the results for all key policy questions and outcome indicators, and provide convincing evidence that the estimated impacts are caused by the programme itself.<sup>36</sup> This is particularly important when evaluation results show a limited or negative impact. If this is the case, it is necessary to explain how the programme was implemented to identify gaps and correct them for future programme cycles. To guarantee that results are fully objective and thus ensure legitimacy, both the baseline and final evaluation reports should be peer reviewed and subject to broad consultation before being finalized.

36 Examples of the mentioned reports can be found at the World Bank page devoted to impact evaluation of youth programmes: http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/ORGANIZATION/EXTHDNETWORK/EXTHDOFFICE

Timeframe, resources and dissemination of findings

10

Beyond delivering evaluation results, the ultimate goal of impact evaluations is to make public policies more effective and improve development outcomes. To ensure that the findings of an evaluation inform policy decisions, it must be communicated clearly, in accessible language, to all the stakeholders, including policy-makers, civil society and the media. Findings should always be made public to ensure accountability.

Before finalizing the evaluation report, a workshop with a broad range set of stakeholders should be organized to discuss results, gather feedback and outline policy changes that could be considered. These consultations can contribute to improving the quality of evaluation results.

Once the final impact evaluation report and policy briefs are available, high visibility dissemination events are critical to ensure that stakeholders become aware of the results<sup>37</sup>. Such activities can also be instrumental in soliciting a response from the responsible authorities (policy-makers and programme managers) regarding how they plan to use the findings of the evaluation to improve the design and implementation of youth employment programmes.

37 The costs of such dissemination activities should be included in the original evaluation budget.

### Additional reading

**Gertel, P.J. Martinez, S., et al. 2011.** *Impact evaluation in practice* (Washington, DC, World Bank).

**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).

World Bank, Independent Evaluation Group. 2011. Writing terms of reference for an evaluation: a how-to guide (Washington, DC, World Bank) http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/ecd\_writing\_TORs.pdf

http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/ORGANIZATION/EXTHDNETWORK/EXTHDOFFICE/o,,contentMDK:23177755~pagePK:64168445~piPK:64168309~theSitePK:5485727,oo.html

## List of tables, figures and boxes

- Table 3.1. Example of process (implementation) indicators (measurement and disaggregation)
- Table 3.2. Outcome indicators (measurement and disaggregation)
- Table 3.3. Examples of performance indicators for youth employment programme
- Table 3.4. Examples of process indicators for public (community-based) works
- Table 5.1. Records of participants to a youth employment programme
- Table 5.2. YEM-supported active labour market programmes
- Table 5.3. YEM Process and outcome indicators
- Table 5.4. Features of NES active labour market programmes under review
- Figure 1.1. Results chain
- Figure 2.1. Youth employment policy and programme development process
- Figure 2.2. Problem and objective trees
- Figure 5.1. Target group, eligible population and baseline
- Figure 5.2. Random sampling of individual with MS Excel
- Figure 6.1. Experimental and quasi-experimental approaches
- Figure 6.2. Before-and-after estimates
- Figure 6.3. Selection of evaluation approaches
- Figure 7.1. Total population (young people aged 15–24 years old)
- Figure 7.2. Randomized promotion
- Figure 9.1. RDD Selection of control group
- Box 3.1. Tips for the development of indicators
- Box 5.1. School-to-work transition surveys
- Box 5.2. Data warehousing in Austria's public employment service (PES)
- Box 6.1. Guidelines to select an impact evaluation methodology

### **TOOLKIT**

# MONITORING AND EVALUATION OF YOUTH EMPLOYMENT PROGRAMMESA LEARNING PACKAGE

## Tool 1



Sample terms of reference for the contracting of a survey firm to carry out a baseline survey <sup>1</sup>

### 1. Purpose of the consultancy

The purpose of this consultancy is to manage all aspects of the collection, entry and management of the baseline data of the programme [title of the youth employment programme] in [Country].

### 2. Background of the programme

The objectives of the programme [name of the youth employment programme] are to [for example: increase the earnings potential and living standards of disadvantaged urban youth through the provision of skills training and temporary employment opportunities. The programme will focus on disadvantaged urban youth (16–25 years old) that neither work nor study and whom have limited social and economic opportunities]. The programme focuses on the Capital District and is implemented by the [insert name of organization/agency managing the programme] with funding of [insert name of funding organization/agency] over the period [inset timeframe]. The programme [brief description]. It targets [insert number] young men and women.

The [name of the specific department responsible for the programme], which oversees the programme, will hire a survey firm/consortium to work under its direct supervision on all aspects related to the baseline data collection activities.

### 3. Scope of work

The assignment includes the following tasks:

- A) **Listing survey**. This activity requires the listing of all individuals in the catchment area of the programme, as provided by[name of the specific department responsible for the programme], through a rapid listing survey<sup>2</sup>. Enumerators will collect basic data on individuals' eligibility, basic demographic and contact information. The expected number of individuals to be listed for this survey is between [insert number] and [insert number]. The sample will be constructed on the basis of this listing.
- B) *Individual survey.* A complete survey will be conducted on a random sample of [insert number] individuals living in the catchment area. Information will be collected through one-to-one individual interviews. The survey will include:

This presumes that the agency managing the programme has access to the demographic listing (stemming from the census or administrative records of the relevant municipalities) of all individuals in the catchment area of the programme that are in the specified age group. This task envisages that the survey firm will interview only those who neither work nor study, and who have limited social and economic opportunities.

- A general information module, consisting of: individual-level demographic, educational and household information; and
- A labour and employment module, consisting of: labour market status (modelled on the information of the National Labour Force Survey), level of individual and household earnings and participation to workrelated training activities.

The survey firm will be expected to complete the following activities associated with this baseline data collection activity.

### 3.1. Gantt chart

The survey firm will develop a Gantt chart in collaboration with the [name of the specific department responsible for the programme] team, outlining the timeline for all activities. The survey firm, additionally, is responsible for designing the research protocol, including informed consent, as well as acquiring from national and/or local authorities all permissions necessary for conducting the survey. The survey firm is also responsible for adhering to local formalities and obtaining any required permits related to the survey implementation, as well as survey team health and accident insurance, salary, taxes and others as necessary.

The deliverables for this component are: (i) Gantt chart, (ii) research protocol and informed consent forms, and (iii) evidence of insurance and permits needed to implement the survey.

### 3.2. Design and pre-testing of the questionnaire

The survey firm will design, in consultation with the [name of the specific department responsible for the programme], an individual questionnaire in the English language. Such questionnaire will include the phrasing of questions and response codes appropriate to the study context. Once the survey firm has developed the questionnaire to reflect the needs of the programme, it will translate it into the local language(s) and pre-test it, adapting any components that are necessary to accurately capture the intended information on the study populations.

The deliverables for this component are: (i) an initial English-language questionnaire, (ii) translation of the initial questionnaire into the local language; (iii) pre-testing of the questionnaire including the timing of modules, comments from interviewers and supervisors, and necessary changes to the questionnaire; (iv) the final local language questionnaire; and (v) the final corresponding English questionnaire.

### 3.3. Preparation of field sampling plan

The [name of the specific department responsible for the programme] will provide the survey firm with an outline of the defined sampling strategy. Based

on this, the survey firm, in consultation with the [name of the specific department responsible for the programme] team, shall propose a plan for selecting the survey sample. The sampling plan must detail:

- A plan for listing and selecting the individuals to be interviewed;
- Definition of the probabilities of selecting the target population;
- Guidelines for selecting replacement individuals; and
- A coding strategy to ensure each individual has a unique code for data entry that merges seamlessly across datasets.

The deliverable for this component is a written individual field-sampling plan approved by the [name of the specific department responsible for the programme].

### 3.4. Final field workplan

The field procedure plan should outline in detail all aspects of the fieldwork to be conducted by the survey firm, including:

- A final updated Gantt chart;
- Composition of a field team, including:
  - o Number of enumerators
  - o Number of field-supervisors
  - o Number of field data-entry clerks
  - o Qualifications and training
- Expected tasks, responsibilities and delivery schedule for each team member;
- Number of scheduled interviews;
- Sample control file for data collection;
- Supervision and spot check plans to ensure adherence to data-collection protocols, including a minimum number of random checks of the baseline sample to confirm the validity of data;
- Protocols and procedures for addressing data inconsistencies/miss-reporting when identified;
- Protocols for Computer Assisted Field Entry (CAFE), whereby questionnaires are captured and validated immediately and the results transmitted back to the field teams to conduct quality checks as needed; and
- Paper questionnaire and data transmission protocols.

This field workplan should be presented to [name of the specific department responsible for the programme] for comments, and revised as necessary prior to commencing fieldwork. The survey firm must then implement the survey, adhering as closely to the plan as local conditions allow. In the event of changes to these plans, the survey firm's field supervisors shall inform the [name of the specific department responsible for the programme] in writing.

The deliverable for this component is a written field workplan approved by the name of the specific department responsible for the programme].

### 3.5. Final data entry programme

The survey firm must develop (or adapt) a robust data entry programme approved by [name of the specific department responsible for the programme]. The survey firm will be responsible for translating the user interface into the local language, if needed, and adding modules for any additional data collection required for the survey. Specifically, the data entry programme:

- Adapts data entry range and consistency checks to values appropriate for the country context;
- Conducts consistency checks during entry of the questionnaire;
- Ensures that violations initiate an immediate alert to the data entry clerk, along with a practical method for correcting errors and documenting any answers that violate the range and consistency check rules;
- Allows valid open-ended and "other" textual responses;
- Envisages variable names that correspond logically to the question labels used in the questionnaire; and
- Has a coding system in place that maintains unique identifiers for individuals.

The deliverable for this component are: (i) a written data entry protocol; (ii) a final data entry programme adapted for the local questionnaire; and (iii) a dataset dictionary with all variables labelled and defined.

### 3.6. Recruitment of qualified field staff

The personnel requirements include:

The core survey team. The survey firm must provide a minimum of:

- (insert number) Project manager
- (insert number) Field manager
- (insert number) Data manager

Field team. The survey firm will determine the number of field teams in consultation with the PMU. Each field team should comprise:

- (insert number) Supervisor
- (insert number) Field data entry clerks
- (insert number) Interviewers

The deliverable for this component is a roster of recruited personnel with their corresponding qualifications.

### 3.7. Training of field staff

The survey firm will design training materials for field staff in consultation with [name of the specific department responsible for the programme]. Comprehensive general training will be provided to supervisors, interviewers and data entry clerks.

Supervisors should receive supplemental training as needed. Training should be scheduled for a minimum of [insert number] weeks. The training programme should include:

- Theoretical knowledge. Training should review the theory underlying the questionnaire and each question, as well as standard quantitative interviewing techniques and field protocols.
- Classroom practice. Training should include individual and group exercises to familiarize trainees with techniques for administering and completing questionnaires.
- *Pilot-testing*: Following theoretical and classroom practice, the interviewers should administer the full questionnaire to a small number of individuals (outside the study sample) in a field situation. This pre-test should focus on administration of the questionnaire under normal circumstances.

The deliverable for this component is the final version of the locally adapted training materials and field manuals.

### 3.8. Pilot testing

Indicators of success include the following:

- Interview teams correctly list, sample and interview individuals in the enumeration area;
- Interview team members understand and correctly follow interviewing protocols; and
- Data from [insert number] individuals in [insert number] enumeration areas (outside the study area) are successfully collected, entered and supervised for quality control without major data entry programme problems.

The deliverable for this component is a documented pilot test and data successfully transferred to the [name of the specific department responsible for the programme]..

### 3.9. Field work management and supervision

A completed sample location includes:

- A dataset containing all data coded, including complete data from the listing exercise and individual survey;
- A field manager's report that documents:
  - o Dates of completion of survey areas
  - o Difficulties or deviations from the standard field plan
  - o A record of each substitution of individuals that may have been required, including the reasons for substitution

- o Any other notable occurrences; and
- A report on real-time validity checks upon receipt of data.

The deliverable for this component is the project manager's written report on baseline data collection.

### 3.10. Final database

The final data delivery report include the following:

- Identify incomplete entries and redundant observations;
- Ensure all components are correctly linked datasets can be merged cleanly;
- Final completion numbers; and
- Completion of the inventory.

The deliverable for this component is the completed database, including the listing and individual data, with data correctly organized, variables named and labelled, and appropriate identifiers that permit seamless merging between databases.

### 3.11. Reporting

[Weekly, bi-weekly] progress reports of the numbers of individuals successfully interviewed.

The deliverable for this component is the timely delivery of the project manager's progress reports to the [name of the specific department responsible for the programme].

### 4. Estimated schedule

The work is estimated to take [insert number] months, from[insert date] to[insert date].

Deliverable	Date	Payment
Signature of contract	Month 0	10%
-Gantt chart with all proposed activities, deliverables and timeframe -Research protocol and informed consent forms -Evidence of insurance and permits needed to implement the survey	Month X	20%
<ul><li>Initial questionnaire in English</li><li>Translation of questionnaire into local language</li></ul>	Month X	
-Pre-testing report including timing of modules, comments from interviewers and supervisors, and necessary changes to the questionnaire -Final local language questionnaire -Final corresponding English language questionnaire	Month X	20%
-Written individual field sampling plan approved by the PMU	Month X	
-Written field workplan approved by the PMU	Month X	]
-Written data-entry protocol -Final data-entry programme adapted for the final questionnaire -Dataset dictionary with all variables labelled and defined	Month X	
-Roster of recruited personnel with corresponding qualifications	Month X	
-Training materials and field manuals	Month X	20%
–Final pilot report and data successfully transferred to the PMU	Month X	
-Project manager's final baseline data collection report	Month X	
-Final databases and final data delivery report	Month X	30%

### 5. Required skills/experience

The survey firm must meet the following minimum requirements:

- Legal status recognized by the Government of [Country], enabling the organization to perform the above-mentioned tasks;
- Minimum [insert number] years experience managing individual-level surveys of similar scale (2000+ individuals); experience working on labour-related surveys preferred;
- Capacity and experience in planning and organizing survey logistics;
- Good network of experienced enumerators, supervisors and data entry clerks;
- Proven capacity in data management and statistics; and
- Strong knowledge of survey-specific software (SPSS and STATA);

### 6. Payment

The work shall be remunerated upon delivery of the agreed outputs, at the agreed fee rate and the agreed number of days of professional fees chargeable.

## **Tool 2**



Sample terms of reference for technical assistance to conduct an impact evaluation<sup>3</sup>

### 1. Purpose of the consultancy

The purpose of this consultancy is to provide technical assistance for a prospective, rigorous impact evaluation of the programme [title of the programme] in [country]. The consultancy includes the following services:

- A. Research services: design and oversight of all aspects of implementation of the impact evaluation;
- B. Data quality assurance services: direct technical support to the survey firm(s) on data quality assurance related to the design and implementation of the impact evaluation; and
- C. Learning, knowledge management and dissemination services: help with the translation of lessons from the impact evaluation for the benefit of [country] employment policy and practice.

### 2. Background of the programme

[Describe the objectives of the programme, including its components and the key research question].

The scope of the evaluation is to assess the impacts of the intervention in [country], as well as to compare these impacts across other countries with similar interventions. It should be able to provide externally valid advice on the feasibility and effectiveness of these approaches.

The [name of organization/agency commissioning the evaluation] is seeking an experienced firm to provide comprehensive technical assistance to the impact evaluation of the programme [title]<sup>4</sup>. The firm shall be responsible for the design and implementation of the impact evaluation, and analysis and dissemination of its results. The firm will collaborate with [name of organization/agency commissioning the evaluation] and key country counterparts (e.g. government, relevant ministerial departments and others) to develop an impact evaluation design. The design should measure the causal effects of the programme, be aligned to operational requirements, and ensure that results are comparable with other similar impact evaluations.

<sup>4</sup> Typically, impact evaluations are contracted to a consultancy firm that works closely with a (local) survey firm responsible for data collection.

### 3. Scope of work

The assignment includes the following tasks:

#### A. Research services

- (a) Establish an in-country advisory panel. The firm shall work with [name of organization/agency commissioning the evaluation] to establish an in-country advisory panel comprising key stakeholders (programme management team, government counterparts, local researchers). This team shall guide the impact evaluation team on policy relevance, country context on other country-specific issues, and generate buy-in from the relevant stakeholders on the impact evaluation design.
- (b) Develop the evaluation design. The firm will develop an impact evaluation design in consultation with key skateholders that isolates the causal impact of the intervention on key outcome indicators [name of indicators, for example, probability of employment and earnings]. The firm will propose a preliminary impact evaluation design for the programme [name]. The evaluation design will be developed with the country counterparts. The firm will seek stakeholders' feedback on the preliminary evaluation design and revise it based on feedback.

The firm will develop a detailed impact evaluation concept note and Gantt chart detailing the agreed upon identification strategy, data collection and analysis plan, staffing, budget and timeframe. This concept note will be integrated into the overall implementation plan. The impact evaluation design shall include collation and analysis of secondary data sources in order to inform power calculations and a rigorous identification strategy.

- (c) On the basis of the draft impact evaluation design, the firm shall prepare a detailed impact evaluation Gantt chart. The Gantt chart will be integrated into the overall implementation plan.
- (d) Identify and help contract a qualified survey firm(s). The firm will advise on the recruitment and contracting of the survey firm(s) required to complete data collection and entry activities. The firm will:
- Draft specifications and finalize terms of reference;
- Provide recommendations on short-listed firm(s); and
- Review the technical and financial proposal with regard to:
  - i. Survey team composition
  - ii. Total number of field teams and field team structure
  - iii. Proposed timeframe and budget allocated to training

For examples of concept notes, baseline and impact evaluation reports, see Impact evaluation cluster: Active labour market and youth employment, (Washington, DC, World Bank), at http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/ORGANIZATION/EXTHDNETWORK/EXTHDOFFICE/o,,contentMDK:23177755~pagePK:64168445~piPK:64168309~theSitePK:5485727,oo.html

- Negotiate and recommend detailed survey firm budget(s).
- (e) Together with the [name of organization/agency commissioning the evaluation], identify key outcome indicators and other variables to be measured as part of the impact evaluation. This will include defining the data sources and methods of collection and preferred measurement methods. The firm will:
- Develop questionnaire templates adapted to the country-specific context;
- Provide in-country support on pre-testing of the questionnaire(s), track necessary revisions to the questionnaire(s) based on lessons learned from pre-test, and document all recommendations for questionnaire content and execution;
- Review questionnaires with an advisory panel;
- Finalize the questionnaire based on feedback from pilots and the advisory panel;
- Review the questionnaire(s) for consistency, accuracy and content, and ensure that skip patterns are correct and that formatting is consistent and print-ready;
- Review the English version of the questionnaire following translation into the local language for consistency and accuracy in content; and
- Develop a detailed research protocol.

The firm will collaborate with the survey firm(s) to identify and fulfil all requirements for ethical clearance of the impact evaluation activities. This includes development of the research protocol, which will include an overview of the methodology, data collection, analysis plan, measures taken for ensuring informed consent and respondent confidentiality.

- (f) The firm will refine the *definition of the sample sizes* required to estimate programme impact, as well as the sampling strategy. This will include:
- The outcome indicator(s) used to determine sample size(s);
- The minimum impacts that the study should be designed to measure;
- The number of individuals of both the control and treatment groups;
- The number of individuals in each community and (if relevant) the number of repeat visits before and after the interventions;
- Definition of the sample inclusion criteria; and
- Definition of central or field-based sampling methodology in collaboration with the survey firm(s).
- (g) The firm will perform all necessary analysis leading to assignment of treatment and control status, specifically:
- Compile the list of geographic locations;
- Perform any necessary stratifications;

- Write the programming code to perform the assignment and run the code;
- Produce the final list of treatment and control areas; and
- Produce the sample control file.
- (h) The firm will oversee all activities related to preparation and implementation of the baseline data collection, including:
- Develop the field workplan with the survey firm(s);
- Design the survey training materials;
- Recruit and train interviewers;
- Run a pilot test with field team(s);
- Produce progress reports; and
- Collect and compile any required secondary data.
- (i) Data cleaning and analysis and validation of the evaluation design (baseline). The firm will perform an initial analysis of the baseline data in order to validate the evaluation design and verify the quality of data produced. This includes production of STATA (or other appropriate statistical software) code, descriptive tables and a report. The firm will:
- Review data files and basic data cleaning to ensure all observations are uniquely identified and reported data is internally consistent;
- Conduct merging across data files;
- Construct variables in accordance with international and local definitions;
- Construct baseline difference in means tests and export to Excel;
- Produce summary statistics of key outcome and covariate variables;
- Produce a written descriptive report; and
- Produce PowerPoint presentations summarizing the report.
- (j) Dissemination of baseline results. The firm will be responsible for organizing dissemination of the results from the baseline surveys. This will include workshops in the capital district.
- (k) Manage data documentation and storage (baseline data). The firm will be responsible for ensuring proper, comprehensive documentation of the impact evaluation activities and ensuring all baseline data and corresponding documentation are in order. The firm will be expected to follow international metadata standards in the documentation of data.
- (I) Monitor and document the intervention between the baseline, midline and follow-up surveys. The firm will be responsible for monitoring and documenting implementation of the intervention. In particular, the firm shall:
- Conduct periodic qualitative work to monitor implementation of the programme;
- Conduct qualitative work to identify the channels of causal impact of the programme, with a view to reviewing data collection

- instruments for follow-up surveys; and
- Track and analyse any available administrative data that could help estimate impact between the baseline and follow-up surveys.
- (m) The firm will oversee all activities related to preparation and implementation of follow-up data collection. In particular, the firm will:
- Recruit a survey firm(s);
- Develop the field workplan with the survey firm(s);
- Design survey training materials;
- Recruit and train interviewers;
- Run a pilot test with field team(s);
- Produce survey firm progress reports; and
- Collect and compile any required secondary data.
- (n) Data cleaning and analysis (endline surveys). The firm will perform an initial analysis of the data to determine the impact of the programme. This includes production of STATA (or other appropriate statistical software) code, descriptive tables and a report. The firm will:
- Review data files and basic data cleaning to ensure all observations are uniquely identified and reported data are internally consistent;
- Conduct merging across data files;
- Construct variables in accordance with international and local definitions;
- Construct difference in means tests, run regressions to estimate impact and export to Excel;
- Produce summary statistics of key outcome and covariate variables;
- Produce a written report; and
- Produce PowerPoint presentations summarizing the report.
- (o) Dissemination of endline results. The firm will be responsible for organizing dissemination of results from surveys. This will include workshops in the capital district.
- (p) Manage data documentation and storage (endline data). The firm is responsible for ensuring proper, comprehensive documentation of the impact evaluation activities and ensuring all data and corresponding documentation are stored adequately. The firm will be expected to follow international metadata standards in the documentation of data.

### B. Data quality assurance services

The firm will provide data quality assurance services for the baseline, midline and endline round of surveys.

- (a) *Technical support on training programme and materials.* The firm will provide support on the following:
- Review the survey firm(s) training programme and provide recommendations on structure and content;
- Develop/adapt training programme materials, such as the supervisor manual, interviewer manual and PowerPoint presentations; and
- Deliver/observe the training programme and provide feedback to the survey firm(s).
- (b) Technical support on data entry programme and management. The firm will provide support on the following activities:
- Develop/adapt the data entry programme, including necessary consistency and out-of-range checks, as well as proper labelling of data;
- Organize data entry;
- Provide remote/in-country support to the survey firm(s) on export to STATA;
- Determine a means to transfer the entered data to a central computer, or design an alternate transference channel;
- Identify any security threats and recommend methods to mitigate threats to data during data entry and management;
- Provide a clear protocol to the data entry supervisor for planning and controlling the progress of data entry, as well as managing data coming from field (proper labelling, storage); and
- Provide estimates on the duration of the entire data entry process and an assessment of the expected data quality based on direct observation of staff performance.
- (c) *Technical support on fieldwork.* The firm will provide support on the following:
- Develop/adapt or advise on the survey firm's sample control file for data collection; and
- Develop/adapt or advise on the field team structure, transportation and data management plan.
- (d) The firm is responsible for *ensuring that data quality* assurance mechanisms are in place for data collection, entry and management, including field supervision, questionnaire-checking in the field, re-surveying a sample of individuals and so on. The firm will supervise the data quality assurance mechanisms put in place by the survey firm and report on a regular basis.

### C. Learning, knowledge management and dissemination services

The firm will support learning, knowledge management and dissemination of results generated around the impact evaluation of the programme. This is to help translate lessons from the impact evaluation results to the benefit of [country] employment policy. In particular, the firm will:

- (a) Analyse the diverse types of information generated by the programme on a regular basis, create policy briefs and disseminate them to stakeholders; and
- (b) Provide technical support to the [Government] for results monitoring of the [programme].

### 4. Team composition

Applicant firms shall propose a team composition that they consider adequate. However, at the minimum, the following requirements should be observed:

- Lead investigator: substantial presence in the country. Proposals should include proposed time allocation in the country and time commitment overall (in country and remote, if applicable);
- Senior investigator: preferably full-time presence in the country. Proposals should include proposed time allocation in the country;
- Data quality assurance expert: substantial presence in the country during survey preparation and implementation time. Proposals should include proposed time allocation in the country;
- Field coordinator/research assistant: full-time presence in the country.

### Lead investigator: required skills and experience

- PhD in a relevant field, preferably labour market economics;
- Minimum [insert number] years of impact evaluation experience [alternative formulation: worked as lead researcher on at least (insert number) impact evaluations];
- Minimum [insert number] years experience in designing and implementing quantitative impact evaluations using randomized or other evaluation designs [alternative formulation: has designed and implemented at least (insert number) impact evaluations using randomized or other evaluation designs];

- Relevant experience in measurement of labour market outcomes through individual/household surveys;
- Relevant experience designing and coordinating field work for large household/individual surveys;
- Relevant experience analysing quantitative data (using statistical analysis software);
- Relevant experience in coordinating implementation of impact evaluation fieldwork;
- Excellent written English communication skills, with focus on research protocols, research papers and descriptive reports for diverse audience;
- Ability to facilitate communication between various levels of management and work independently in order to meet deadlines; and
- Evaluations published in peer-reviewed journals.

### Senior investigator: required skills and experience

- PhD in relevant field, preferably labour market economics;
- Minimum [insert number] years of project impact evaluation experience [alternative formulation: contributed to at least (insert number) impact evaluations];
- Minimum [insert number] years experience in designing and implementing quantitative impact evaluations using randomized or other evaluation designs; [alternative formulation: has experience in at least (insert number) impact evaluations using randomized or other evaluation designs];
- Relevant experience in measurement of labour market outcomes through household/individual surveys;
- Relevant experience designing and coordinating field work for large household/individual survey;
- Relevant experience analysing quantitative data using statistical analysis software;
- Relevant experience in coordinating implementation of impact evaluation field work;
- Excellent written English communication skills, with a focus on research protocols, research papers and descriptive reports for diverse audience;

- Ability to facilitate communication between various levels of management and work independently to meet deadlines; and
- Ideally, evaluations published in peer-reviewed journals.

### Data quality assurance expert: required skills and experience

- [Insert number] years experience managing large-scale data collection including household surveys. Experience with labour market data collection is a plus;
- Extensive experience with data quality assurance mechanisms (training of enumerators, design and adaptation of questionnaires, field work management, data entry programming, data management);
- Ability to communicate in English and local language a plus;
- Ability to submit all written documentation in English; and
- Proven ability to facilitate communication between different levels of government, and between government and development partners.

### Field coordinator: required skills and experience

- Master's degree in relevant field, preferably labour market economics;
- Experience of managing projects, including procurement and field work;
- Relevant experience analysing quantitative data (household and individuals) using statistical analysis software;
- Ability to communicate in English, local language a plus;
- Ability to submit all written documentation in English;
- Familiarity with the labour market of the country a plus; and
- Proven ability to facilitate communication between different levels of government, and between government and development partners.

### 5. Deliverables

А	Deliverables	Delivery date	Days	Responsable team member
	a. Establish advisory panel			
	b. Impact evaluation concept note			
	c. Gantt chart			
	d. Survey firm TOR(s)			
	Final written summary of comments and			
	recommendations regarding survey firm			
	proposals			
	e. Final English questionnaire(s) with			
	comments and recommendations from pre-			
	test			
	Research protocol			
	Final English questionnaire			
	f. Ethical clearance package(s)			
	g. Power calculations			
	Survey inclusion criteria			
	Central or field sampling plan			
	h. Compilation of geographic locations			
	Assignment to treatment and control			
	including programming code			
	Sample control file			
	i. For baseline surveys:			
	Field workplan			
	Final training materials			
	Summary of pilot test			
	Any required secondary data			
	j. Baseline report			
	Report on validation of evaluation design			
	Includes STATA do files, tables and reports			
	and PowerPoint presentations.			
	k. Baseline dissemination workshops			
	I. Final baseline data and documentation			
	m. Six-month updates on programme			
	implementation			
	Reports on qualitative work regarding			
	channels of causal impact			
	Analysis of administrative data to estimate			
	impact at preliminary stages			
	n. For midline surveys:			
	Field workplan			
	Final training materials			
	Summary of pilot test			

		·	
	Any required secondary data		
	o. Midline report		
	Midline impact report		
	Includes STATA do files, tables and reports		
	·		
	and PowerPoint presentations		
	p. Final midline data and documentation		
	q. Midline dissemination workshop		
В	a. For baseline survey round:		
	Final written summary of comments on		
	training programme and materials		
	Final manual(s) and PowerPoint presentations		
	Final written summary of comments and		
	recommendations for training		
	<u> </u>		
	b. For baseline survey round:		
	Final data dictionary		
	Written summary of data entry protocol,		
	including management of data from field,		
	threats to security and recommendations,		
	description of data management, estimation		
	of productivity (questionnaires per day, total		
	days)		
	c. For baseline survey round:		
	Final sample control file with comments and		
	recommendations		
	Written summary of recommendations		
	related to field work management, including		
	field team structure and data management		
	plan		
	d. For baseline survey round:		
	Detailed description of quality control		
	measures implemented and assessment of		
	success		
C	a. Policy briefs		
	b. Results monitoring of the [name of		
	programme]		

### 6. Estimated timeline and management arrangements

The consultancy will be required from [start date] to [end date] subject to the availability of funds. This is expected to include one round of baseline surveys as well as the endline surveys.

The evaluation team will be located at [name of organization/agency hosting the consultancy]. The firm will report directly to [name of responsible organization/agency]. All data, tools and reports produced for the impact evaluation in the context of this contract will be the property of [name of responsible organization/agency] and the Government of [country].

# >>>> Tool 3

Sample follow-up questionnaire (treatment group)\*

### A. General information

A1. F	full name of respond	lent:				-
A2. E	Birth date: day:	month:	yea	ar:		
A3. S	Sex:					
	Woman Man	1 2				
A4. A	Address of responde	nt:				
A5. L	evel of formal scho	oling before ento	ering the	programme	<u> </u>	
	Primary school (in Primary school (c Secondary schoo Secondary schoo	ompleted) I (general, incom		)	1 2 3 4	
A6. F	Programme attende	d				
	Work-training pro Employment sub Work trial progra Institution-based On-the-job traini Self-employment	sidy mme training ng	2 3 4 5	Go to A7 Go to A8		
A7. V	Vhat was your wage	level during pro	gramme	participatio	n?	SRD
A8. F	Programme attende In (occupation)	d from				i)
	How did you find cipation?	the guidance,	/advice/in	nformation	provided duri	ng programme
	Very useful Useful Not useful		1 2 3			
	What is your situati best describes your			yment in the	e last month? (	Choose the one
	Attending educat Available and act Working for salar	ively looking for		1 2	Go to que Go to que Go to que	stions C

Working as self-employe Engaged in home duties Unable to work owning t Rentier (independent, ov	(including child care o sickness and disa	e) 5	End of interview
B. In training			
B1. Have you ever been employ you attended?	ed during the perio	d since the e	nd of the programme
Yes 1 <b>Go to B2</b> No 2 <b>Go to B10</b>	,		
B2. In which occupation/sector v	vas this job?		
In (occupation)		(sector,I	NACE)
B3. What kind of employment w	as it?		
Unlimited duration Limited duration (12–36 Seasonal (under 12 mont Other (please specify)	ths)		1 2 3 4
B4. How long did it last? (If less t	than two years indic	ate the numb	er of months.)
	(months,	years)	
B <sub>5</sub> . How much was your average	e wage in this job? _		_
B6. How much was your average i	ncome during progr	amme partici	pation?
B7. What was the reason for leav	ving the job?		
I was dismissed I left the job of my own w		Go to B8 Go to B9	
B8. Why do you think the compa	any dismissed you?		
Unsuitable, low level of s Requirements for job hig Insufficient work experie Enterprise had to reduce Considered too young/to Being male/female Discrimination, prejudice Other (please specify)	her than skills poss nce production/dismiss oo old		1 2 3 4 5 6 7

GO TO B<sub>10</sub>

Bg. Why did you leave the company?		
Wanted to undergo additional training Found another job Low wage level Poor working conditions Discrimination, prejudices Personal/family reasons (personal/family responsibely)	-	1 2 3 4 5 6
B10. After how many months/years will you complete you	r currer	t training?
months/years		
B11. What do you plan to do after completing your current	t trainin	g programme?
Look for a job Enter self-employment Stay at home (personal or family responsibilities) Immediately enter further education/training Do not know	1 2 3 4 5	Go to B12 Go to B14 End of interview End of interview End of interview
B12. Have you already started looking for a job?		
Yes Go to B13 No End interview		
B13. How are you looking for a job? (Mark up to four meth	nods)	
Through education/training institution Through employment service Through job fairs Through advertisements (newspaper, Internet, etc Through friend and/or relatives Through a private employment agency Other (please specify)	)	1 2 3 4 5 6 7
B14. Which steps have you taken to enter self-employmen	nt? (Mar	k up to four methods)
Attended self-employment training Applied for a credit in commercial bank Applied to a grant/credit scheme Borrowed money from family/relatives Contacted business partner Other (please specify)		1 2 3 4 5 6

**End of interview** 

## C. Unemployed

C1.Have	you	ever	been	employed	during	the	period	since	the	end	of	the
program	ne y	ou att	endec	1?								
			_	_								
Ye	es :	1	Go	to C2								

No 2 <b>Go to C10</b>	
C2. In which occupation/sector was this job? In(occupation)(sector,NACE)	
C <sub>3</sub> .What kind of employment was it?	
Unlimited duration	1
Limited duration (12–36 months)	2
Seasonal (under 12 months)	3
Other (please specify)	4
C4. How long did it last? ( <i>If less than two years indicate the num</i> (months/years)	ber of months)
C5. How much was your average wage in this job?	
C6. How much was your average income during programm	e participation?
C7. What was the reason for leaving the job?	
I was dismissed 1 <b>Go to C8</b> I left the job of my own will 2 <b>Go to C9</b>	
C8. Why do you think the company dismissed you?	
Unsuitable, low level of skills	1
Requirements for job higher than skills possessed	2
Insufficient work experience	3
Enterprise had to reduce production/dismiss workers	4
Considered too young/too old	
Being male/female	5 6
Discrimination, prejudices	7
Other (please specify)	8

GO TO C10

C9. Why did you leave the company?	
Wanted to undergo additional training Low wage level Poor working conditions Discrimination, prejudices Personal/family reasons (personal/family responsibility) Other (please specify)	1 2 3 4 5 6
C10. How long have you been available for work and actively looking for months	or a job?
C11. How are you looking for a job? (Mark up to four ways)	
Through employment service Through job fairs Through advertisements (newspaper, Internet, etc) Through friend or relatives Through a private employment agency Other (please specify)	1 2 3 4 5
C12. What is, in your opinion, the main obstacle to finding a job?	
No/unsuitable education No suitable training opportunities Requirements for job higher than education/training received No work experience Not enough jobs available Considered too young Being male/female Discrimination, prejudices Low wages in available jobs Poor working conditions in available jobs Other (please specify)	1 2 3 4 5 6 7 8 9 10
C13.Do you feel the programme you attended will be useful in getting	a job?
Very useful 1 Somewhat useful 2 Not useful 3 Do not know 4	

**End of interview** 

### D. Workers

D1. Are you still working for the same employer you had during the programme?					
Yes Go to D8 No Go to D2					
D2. Who was the employer during your programme?					
Name of company:Industry:					
D <sub>3</sub> In which occupation/sector was the programme you attended?					
In (occupation)(sector, N	ACE)				
D4. How much was your average income during programme particip	pation?				
D <sub>5</sub> . What was the reason for leaving the employer?					
I was dismissed 1 <b>Go to D6</b> I left the job of my own will 2 <b>Go to D7</b>					
D6. Why do you think the company dismissed you?					
Unsuitable, low level of skills Requirements for job higher than skills possessed Insufficient work experience Enterprise had to reduce production/dismiss workers Considered too young/too old Being male/female Discrimination, prejudices Other (please specify)	1 2 3 4 5 6 7 8				
GO TO D8					
D7. Why did you leave the company?					
Found another job Low wage level Poor working conditions Discrimination, prejudices Personal/family reasons (personal/family responsibility) Other (please specify)	1 2 3 4 5 6				

D8. Please provide the details of your current employer:	
Name of company: Industry: Number of workers:	
Dg. In which occupation/sector is the job you have?	
In (occupation)	(sector,NACE)
D10.What kind of employment is it?	
Unlimited duration Limited duration (12—36 months) Seasonal (under 12 months)	1 2 3
Other (please specify)	4
D11. When did you start your current job? (If less than two (months/years)  D12. In your current job, are you entitled to the following	
D12. In your corrent job, are you entitled to the following	y:
Facility/benefit  1. Transport or allowance for it  2. Meals or meal allowance  3. Annual paid leave (holiday time)  4. Paid sick leave  5. Pensions/old-age insurance  6. Severance/end-of-service payment  7. Medical insurance coverage  8. Bonus/reward for good performance  9. Social security contribution  10. Educational or training courses  11. Occupational safety/protective equipmer  12. Child care facilities  13. Others (please specify)	
D13. How much do you earn now?	
D14. How much did you earn before the programme?	
D15When you finished the programme, did you ex	spect to earn:
More than now? 1 Less than now? 2 About the same? 3	

	s your imme?	presen	t occupation relat	ted to the training/wo	ork you did during the
	Yes No				
-	hat typ currer		owledge do you thi	nk is primarily needed fo	or performing efficiently
		cal ience etical ar	nd practical nd experience	1 2 3 4 5	
D18. A	re you	applyin	g in your work wha	at you learned during th	e programme?
			Go to D20 Go to D19 Go to D19		
		e things or your v		g the programme, wha	t is the one thing that is
	Theor		nery/tools struction vorkers	1 2 3	
D20. lr	n your v	vork, are	you using equipm	ent similar to that used	during the programme?
	Yes No				
			ence gained durir in your current job		table to enable you to
	Yes No		Go to D23 Go to D22		
D22. W	/hy?				
	Progra Progra Progra Progra	amme d amme d amme w amme w	id not provide suit id not provide theo id not provide user vas too short vas too long specify)	oretical knowledge ful skills	1 2 3 4 5 6

D23Do you think participating in the programme has paid off?					
Yes Go to D24 No Go to D25					
D24. If you think it has paid off, why?					
I got a good job I am earning more money than without the programme I have more prestige than somebody without the training It opened up possibilities for employment and advancement	1 2 3 4				
D <sub>25</sub> . If you do not think it has paid off, why?					
I did not get a good job I am not earning more than I would have earned otherwise I have little status at work It did not open up possibilities for advancement	1 2 3 4				
End of interview					
E. Self-employed					
E1. Did you ever have a salaried/wage job prior to entering self-employn	nent?				
Yes Go to E2 No Go to E9					
E2. Who was your employer? Name of company:					
E3. In which occupation/sector was your prior job?					
In (occupation)(sector,NACE)					
E4. What kind of employment was it?					
Unlimited duration 1 Limited duration (12–36 months) 2 Seasonal (under 12 months) 3 Other (please specify)4					
E5. How long did it last? (If less than two years indicate the number of mo	nths)				
(months/years)					
E6. What was the reason for leaving the job?					
I was dismissed 1 <b>Go to E7</b> I left the job of my own will 2 <b>Go to E8</b>					

E7. Why do you think the company dismissed you?	
Requirements for job higher than skills possessed Insufficient work experience Enterprise had to reduce production/dismiss workers Considered too young/too old Being male/female Discrimination, prejudices	1 2 3 4 5 6 7
E8. Why did you leave the company?	
Low wage level Poor working conditions Discrimination, prejudices	1 2 3 4 5 6
Eg. Why did you choose to be self-employed?	
Greater independence as self-employed/own account worker More flexible hours of work Higher income level	1 2 3 4 5
E10. When did you establish your present business?	
Month:Year:	
E11. In which industry does your enterprise/organization operate?	
(Sector, NACE)	
E12. Where do you mainly undertake your work?	
Farm or individual agricultural plot Homes or workplaces of clients Construction sites Markets, trade fairs, street stall	1 2 3 4 5 6 7

E13. W	hom do you mainly sell the goods or	service	es to?		
	Private individuals or households Small businesses, traders Middlemen, agents, contractors Large shops or enterprises in domestic market International market, export products Government agencies/public enterprises Others (please specify)				
E14. D	oes anyone help you in your busines	s/econo	mic activity?		
	Paid employees Family members No help, working alone	1 2 3	/	<b>-</b>	
	om where did you get the money to ode all that applies)	start yo	our current business?		
Own savings Savings of other family members Loan from family and/or friends Loan from bank or commercial institution Grant from NES Loan/assistance from government institution Loan/assistance from NGO, donor project, etc. Funds from savings and credit/group Other (please specify)					
E16. D	o you still have outstanding loans fo	r the bu	siness?		
	Yes ONO				
E17. Do	oes your business/enterprise have a	trading	licence?		
	Yes ONO				
E18. Is	your business/economic activity:				
	Making a profit? Making a loss? Breaking even?			1 2 3	

E19. How much is your monthly incom	SRD		
E20. Do you feel the programme you p your business?	participated in was useful to you in	n establishing	
Very useful Somewhat useful Not useful Do not know	1 2 3 4		
E21. How long were you looking for w		ent business?	
months			
E22. What was, in your opinion, the m	ain obstacle to finding a good job	)?	
No/unsuitable education No suitable training opportunir Requirements for job higher the No work experience Not enough jobs available Considered too young Being male/female Discrimination, prejudices Low wages in available jobs Poor working conditions in available, please specify	nan education/training received	1 2 3 4 5 6 7 8 9 10	

**End of interview** 



# Sample follow-up questionnaire (control group)\*

A1. Fu	ll name	of resp	ondent:				·
A2. Bi	rth date	e: day: _	month	າ:	year:		
A <sub>3</sub> . Se	X:						
	Woma Man	ın	1 2				
A4. Ac	ddress o	f respo	ndent:				
A5. Le	vel of fo	ormal so	hooling att	ained			
	Prima Prima Secon Secon	ry schoo ry schoo dary scl dary scl	ol (incomple ol (complete nool (gener nool (vocati	ete) ed) al, incomple onal, incom		1 2 3 4 ring the	last month?
					s your current		
	Availa Worki Worki Engag Unabl	ble and ng for sang ng as se led in ho e to wo	alary/wage If-employe ome duties	oking for wo with an emp d/own accou (including ch o sickness an	oloyer unt worker	1 2 3 4 5 6	Go to questions B Go to questions C Go to questions D Go to questions E End of interview End of interview End of interview
B. In	trainir	ng					
B1. H	ave you	ever be	en employe	ed?			
	Yes No	1 2	Go to B2 Go to B10				
B2. In	which o	occupat	ion/sector v	was this job?	•		
	In (occ	cupation	n)		(sect	or,NAC	E)
B <sub>3</sub> . Wl	nat kind	l of emp	loyment w	as it?			
	Limite Seaso	nal (und	ion (12—36 r Ier 12 mont			1 2 3 _4	

A. General information

B4. How long did it last? (If less than two years indicate	the numbe	er of months)						
(months/yea	(months/years)							
B5. How much was your average wage in this job?								
B6. What was the reason for leaving the job?								
I was dismissed 1 I left the job of my own will 2	Go to Go to							
B7. Why do you think the company dismissed you?								
Unsuitable, low level of skills Requirements for job higher than skills possesse Insufficient work experience Enterprise had to reduce production/dismiss wo Considered too young/too old Being male/female Discrimination, prejudices Other (please specify)		1 2 3 4 5 6 7 8						
GO TO B <sub>9</sub>								
B8. Why did you leave the company?								
Wanted to undergo additional training Found another job Low wage level Poor working conditions Discrimination, prejudices Personal/family reasons (personal/family responded)		1 2 3 4 5 6 -						
B9. After how many months/years will you complete yo months/years	our curren	t training?						
B10. What do you plan to do after completing your curi	rent trainir	ng programme?						
Look for a job Enter self-employment Stay at home (personal or family responsibilitie Immediately enter further education/training Do not know	1 2 s) 3 4 5	Go to B11 Go to B13 End of interview End of interview End of interview						

ртт. г	iave yc	ou alleac	iy Start	ed looking for a job!			
	Yes No		Go to End i	B12 nterview			
B12. How are you looking for a job? (Mark up to four methods)							
	Throu Throu Throu Throu Throu	ugh eduough empugh job f ugh job f ugh advough frier ugh a pr r (please	5 6				
B13.W	/hich st	teps hav	e you t	aken to enter self-employn	nent? (Mark up to fo	ur methods)	
	Appli Appli Borro Conta	nded self led for a led to a g lowed mo acted bu r (please	1 2 3 4 5 				
				End of interview			
C. Ur	nempl	loyed					
C1. Have you ever been employed?							
	Yes No		1 2	Go to C2 Go to C10			
C2. In which occupation/sector was this job?							
	In (oc	cupatio	(sector,NACE)_				
C3. W	hat kin	d of em	ployme	ent was it?			
	Unlimited duration Limited duration (12—36 months) Seasonal (under 12 months) Other (please specify)						
C4. H	ow long	g did it la	ast? ( <i>If</i>	less than two years indicate	the number of mont	hs)	
	(months/years)						

C <sub>5</sub> . How much was your average wage in	this job	?					
C6. What was the reason for leaving the	job you h	nad?					
I was dismissed I left the job of my own will	1 2	Go to C7 Go to C8					
C7. Why do you think the company dismissed you?							
Unsuitable, low level of skills Requirements for job higher than skills possessed Insufficient work experience Enterprise had to reduce production/dismiss workers Considered too young/too old Being male/female Discrimination, prejudices Other (please specify)							
GO TO C <sub>9</sub>							
C8. Why did you leave the company?							
Found another job Low wage level Poor working conditions Discrimination, prejudices Personal/family reasons (personal/family responsibility) Other (please specify)							
Cg. How long have you been available fo month		nd actively looking for	a job?				
C10. How are you looking for a job? (Mar	k up to f	our ways)					
Through employment service Through job fairs Through advertisements (newspa Through friend or relatives Through a private employment a Other (please specify)		ernet, etc.)	1 2 3 4 5 6				
C11. What is, in your opinion, the main o	bstacle to	o finding a job?					
No/unsuitable education No suitable training opportunities Requirements for job higher than No work experience		on/training received	1 2 3				

GOTO D8

7

Discrimination, prejudices

Other (please specify) \_\_\_\_

D7. Wr	ny did you leave the company?		
	Found another job Low wage level Poor working conditions Discrimination, prejudices Personal/family reasons (personal/family resp Other (please specify)	onsibility)	1 2 3 4 5 6
D8. Ple	ease provide the details of your current employ Name of company: Industry: Number of workers:	=	
Dg. In	which occupation/sector is the job you have?		
	In(occupation)(	sector,NACI	Ξ)
D10. W	Vhat kind of employment is it?		
Limite Seaso Other	ited duration d duration (12–36 months) nal (under 12 months) (please specify) /hen did you start your current job? ( <i>If less than t</i>	wo vears ind	1 2 3 4 icate the number
of mor	, ,	woyearsina	reace are normoer
	(months/years)		
D12. lr	n your current job, are you entitled to the follow	ving?	
14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26.	Facility/benefit Transport or allowance for it Meals or meal allowance Annual paid leave (holiday time) Paid sick leave Pensions/old-age insurance Severance/end-of-service payment Medical insurance coverage Bonus/reward for good performance Social security contribution Educational or training courses Occupational safety/protective equipment, clichild care facilities Other (please specify)	Yes	No
D13.H	ow much do you earn now?	_	

D14. What type of knowledge do you think is primarily needed for performing efficiently in your current job? Theoretical 1 Practical 2 Experience 3 Theoretical and practical Theoretical and experience 5 End of interview E. Self-employed E1. Did you ever have a salaried/wage job prior to entering self-employment? Go to E2 Yes Go to E8 E2. Who was your employer? Name of company: E3. In which occupation/sector was your prior job? In (occupation) (sector, NACE) E4. How much did you earn?\_\_\_\_\_ E5. What was the reason for leaving the job? I was dismissed Go to E6 I left the job of my own will 2 Go to E7 E6. Why do you think the company dismissed you? Unsuitable, low level of skills Requirements for job higher than skills possessed 2 Insufficient work experience 3 Enterprise had to reduce production/dismiss workers Considered too young/too old Being male/female Discrimination, prejudices Other (please specify)

**GOTOE8** 

E7. Why did you leave the company	/?				
Wanted to set up own busin Low wage level Poor working conditions Discrimination, prejudices Personal/family reasons (pe Other (please specify)	rsonal		1 2 3 4 5 6		
E8. Why did you choose to be self-e	employ	ved?			
Could not find a wage or sal Greater independence as se More flexible hours of work Higher income level Other (please specify)	elf-emp	oloyed/own account worker	1 2 3 4 5		
Eg. When did you establish your pro	esent l	ousiness? Month:Year:			
E10. In which industry does your en	nterpris	se/organization operate?			
(Sector, NACE)					
E11. Where do you mainly undertak	ke youi	rwork?			
Business, space inside or att Factory, office, workshop, s Farm or individual agricultu Homes or workplaces of clie Construction sites	Markets, trade fairs, street stall				
E12. Whom do you mainly sell the g	goods	or services to?			
Private individuals or households Small businesses, traders Middlemen, agents, contractors Large shops or enterprises in domestic market International market, export products Government agencies/public enterprises Other (please specify)					
E13.Does anyone help you in your b	ousines	ss/economic activity?			
Paid employees Family members No help, working alone	1 2 3	How many? How many?			

(Code all that applies) Own savings 1 Savings of other family members 2 Loan from family and/or friends 3 Loan from bank or commercial institution 4 Grant from NES 5 Loan/assistance from government institution 6 Loan/assistance from NGO, donor project, etc. 7 Funds from savings and credit/group 8 Other (please specify) \_\_\_\_\_ 9 E15. Do you still have outstanding loans for the business? Yes 🔘 No E16. Does your business/enterprise have a trading licence? Yes No E17. Is your business/economic activity: Making a profit? 1 Making a loss? Breaking even? 3 How much is your monthly income? \_\_\_\_\_ SRD E18. How long were you looking for work before establishing your current E19. business? months What was, in your opinion, the main obstacle to finding a good job? E20. No/unsuitable education No suitable training opportunities 2 Requirements for job higher than education/training received 3 No work experience 4 Not enough jobs available 5 Considered too young 6 Being male/female 7 Discrimination, prejudices 8 Low wages in available jobs 9 Poor working conditions in available jobs 10 Other (please specify) \_\_\_\_\_ 11

E14. From where did you get the money to start your current business?

## TOO 5 Sample tabulation plan follow-up surveys

### **Cumulative tables**

- Number and percentage over total of respondents holding a job at any time since leaving the programme;
- 2. Number and percentage over total of respondents who are in education, unemployed, in wage employment, in self-employment and inactive.

### For those in school or training:

- 3. Number and percentage who plan to look for a job or to enter self-employment;
- 4. For those already looking for a job, number and percentage by job-searching method;
- 5. For those planning for self-employment, number and percentage by steps taken.

### For those who are unemployed:

- 6. Number and percentage looking for a job for more than six months;
- 7. Number and percentage by method of job-searching;
- 8. Number and percentage by obstacle to finding a job.

### For those in wage employment:

- 9. Number and percentage by job length (1–3 months, 3–6 months, more than 6 months);
- 10. Number and percentage in a permanent, temporary and seasonal job;
- 11. Number and percentage in full-time work;
- 12. Number and percentage by enterprise type and size;
- 13. Number and percentage by economic sector;
- 14. Number and percentage by occupation;
- 15. Number and percentage in informal working arrangements;
- 16. Number and percentage by net monthly earning;
- 17. Number and percentage using learning acquired during the programme in the current job.

### For those who are self-employed:

- 18. Number and percentage by length of business operation (< 6 months; > 6 months);
- 19. Number and percentage by economic sector;
- 20. Number and percentage employing additional workers;
- 21. Number and percentage with good prospects;
- 22. Percentage in informal self-employment;
- 23. Number and percentage by range of monthly net income;
- 24. Number and percentage using learning acquired during the programme in the current job.

Knowledge Management Facility on Youth Employment and Migration

c/o Youth Employment Programme 4, route des Morillons CH12-11 Geneva 22, Switzerland Email: youth@ilo.org

Visit our website: www.ilo.org/yep

ISBN 978-92-2-128075-0

