

EVALUATING ACTIVE LABOUR MARKET POLICIES IN FRAGILE CONTEXTS

Insights and lessons from evaluation
studies in the ETF partner countries

Disclaimer

This report was prepared for the European Training Foundation.

The contents of the report are the sole responsibility of the ETF and do not necessarily reflect the views of the EU institutions.

© European Training Foundation, 2025

Reproduction is authorised, provided the source is acknowledged.

PREFACE

As the agency of the European Union (EU) supporting the neighbouring countries, the European Training Foundation (ETF) contributes to the development of human capital by providing advice and support on the reform of education, training and employment policies and systems. This contributes to social well-being, stability and prosperity in the neighbouring countries of the European Union.

In the aftermath of Covid-19 pandemic, multiple socio-economic and conflict crises and accelerated digital transformation of labour markets and employment, the ETF has engaged in the international debate about the effectiveness of Active Labour Market Policies (ALMPs), particularly skills formation dimension, in developing and transition countries. This review is one of several ETF actions that analyses the impact of ALMP design and delivery and supports peer learning and capacity building in the partner countries. The ETF conducted in 2023 and 2024 the review of the effectiveness of Active Labour Market Policies (ALMPs) in ETF partner countries (PCs), with particular focus on fragile socio-economic and (post)conflict contexts. The paper aims at drawing up specific recommendations on the design of ALMPs and the improvement of monitoring and assessment of implemented measures and services in contexts characterized by fragility, data scarcity and limited capacities to run fully fledged impact assessment of ALMPs.

The research was implemented by a team of researchers mobilised by the EPRD, with the contribution of ETF experts and based on several consultations and exchanges with ETF partner country experts and researchers in the field, as well as representatives of various international development institutions and organisations. The main authors of the review were Dr Nicola Duell and Marius Haulica (EPRD), with contributions from Eamonn Davern and Armen Cekic (EPRD). The ETF experts - Eva Jansova, Cristina Mereuta and Piotr Stronkowski - supervised the preparation and development of the study, including the facilitation of interviews with stakeholders from various countries and representatives of international organisations and EU Member States bilateral cooperation organisations. Iwona Ganko and Donatella Di Vozzo from the ETF contributed with an internal peer review of the draft report. Special thanks go also to Kamuran Cosar from ISKUR, the public employment service of Türkiye, for his invaluable contributions to this report and for facilitating numerous stakeholder interviews. The report underwent additional peer review by Shari Ghyselen from ENABEL and Anke Green from GIZ.

The study aimed to fill the knowledge gaps on (impact) evaluations of ALMPs, which have been carried out by public or private bodies as part of the implementation of public policies/measures and/or (donor) projects. In order to meet the objectives, the research implemented a multidisciplinary methodological framework that combined quantitative and qualitative research methods. The analysis was conducted primarily via desk research, but was supported by a survey, and interviews or other forms of communication with relevant stakeholders and experts in the EU Member States, the ETF partner countries and other countries with fragile socio-economic contexts, to identify good practices.

Representatives of the ETF partner countries and other stakeholders discussed the main conclusions and recommendations proposed in the draft report during the ETF policy learning event on ALMPs in fragile and post-conflict contexts, organised in Brussels (Belgium), on 26-27 June 2024. The ETF expresses its high appreciation and gratitude to all the researchers and experts who contributed to the reflections and shared valuable information and insights.

CONTENTS

PREFACE	3
CONTENTS	4
EXECUTIVE SUMMARY	5
1. INTRODUCTION	7
2. METHODOLOGY AND KEY CONCEPTS	9
3. ORGANISING EVALUATIONS	13
3.1 Capacity for effective evaluation	13
3.2 Budget	15
3.3 Use of results	15
4. IMPLEMENTING EVALUATIONS	17
4.1 Evaluation design	17
4.2 Data collection and data requirements	24
4.3 Outcome and impact indicators	30
4.4 Limitations for the evaluation and challenges in relation to fragile contexts	32
5. INSIGHTS FROM EVALUATIONS OF ALMPS	35
6. CONCLUSIONS	42
GLOSSARY	45
ACRONYMS	47
REFERENCES	49

EXECUTIVE SUMMARY

Creating sustainable, inclusive jobs is essential for many countries, including those affected by fragility, conflict and violence, where employment can drive economic growth, reduce poverty and promote social stability. However, job creation in fragile contexts is especially challenging due to elevated risks and persistent instability. In these environments, economic activity often slows down, limiting opportunities for workers, impeding the growth of businesses, and increasing investment risks for the private sector. Recognising these challenges, active labour market policies (ALMPs) are essential for strengthening employment and supporting transitions in the labour market. Despite their potential impact, ALMPs face constraints in funding, institutional capacity and monitoring, making it vital to rigorously evaluate their design, implementation and outcomes in order to optimise resource allocation for measures that make a difference.

To enhance knowledge and foster effective ALMP design and evaluation, the European Training Foundation (ETF) launched an initiative to assess ALMP effectiveness across its partner countries, in particular those in fragile, post-conflict contexts. This paper compiles the findings from this initiative and reviews technical aspects and design of impact evaluations used to assess existing ALMPs to provide insights for refining these programmes and better serve vulnerable populations. The assessment builds on desk research, encompassing existing evaluations by public and private institutions since 2017, and interviews with experts, representatives of public employment services and international organisations or other entities.

Key findings from the review of ALMP evaluations

Evaluation Objectives and Methodologies: impact evaluations of ALMPs are conducted to assess the effectiveness, efficiency and social value of programmes, providing insights to policymakers on which approaches yield the best outcomes for different demographic groups. In fragile settings, these evaluations are crucial for identifying context-specific barriers and informing the design and targeting of programmes. Evaluations frequently use experimental and quasi-experimental methods, involving both qualitative and quantitative data collection, which provide valuable insights into the ‘why’ and ‘for whom’ aspects of ALMP effectiveness.

Challenges in Evaluation Implementation: conducting thorough evaluations presents several challenges. National institutions, such as public employment services (PES), often lack the technical skills to design and oversee evaluations, resulting in reliance on international organisations, e.g. World Bank, International Labour Organization (ILO), United Nations Development Programme (UNDP), to conduct these studies. While international expertise brings technical rigour, it limits opportunities for domestic capacity building and continuity of impact assessment exercises. To improve local capabilities, PES and related ministries would benefit from capacity-building initiatives that focus on evaluation design, methodologies and results utilisation, as well as from building alliances with local universities and scientific communities.

Methodological Considerations and Resource Constraints: evaluation methods vary based on resource availability, programme scope and specific objectives. Large-scale ALMPs often use rigorous counterfactual impact evaluations requiring extensive data and large sample sizes, while smaller-scale or pilot projects rely on qualitative methods. Access to reliable data sources, such as household surveys and administrative records, is critical for high-quality evaluations. Extended evaluation timelines enable the measurement of short-, medium-, and long-term impacts, although this can conflict with the need for timely results to inform policy decisions.

Utilisation and Dissemination of Evaluation Results: effective use of evaluation findings depends on transparent dissemination practices, which vary across countries. In contexts where a robust evaluation culture exists, findings are shared in stakeholder workshops and reports, fostering broader engagement and trust. However, in some fragile settings, limited internal circulation of reports constrains their impact on policy design and implementation. Wider dissemination and trust-building

between evaluators and policymakers can increase the practical use of evaluations, making the results more actionable in policy cycles.

Key ALMP interventions and outcomes

ALMPs, including those implemented in fragile contexts, commonly include (re)training, employment incentives, public works, and start-up incentives. According to evidence from the reviewed evaluations, each of these interventions has shown varying levels of success across different demographic groups. On-the-job training programmes, for example, have proven especially effective for youth and women, as they improve employability and skill acquisition. Employment incentives, often in the form of cash, help reduce poverty by encouraging participation in the labour market and supporting economic stability. Public works programmes primarily benefit low-skilled workers, while start-up incentives foster entrepreneurship, proving beneficial for returning migrants and displaced individuals. Additionally, combining approaches—such as cash transfers with training and mentoring—has shown increased impact, especially in promoting self-employment.

Recommendations

To improve the effectiveness and sustainability of ALMPs, building capacity within local institutions (particularly public employment services (PES) and related ministries) is essential. Strengthening evaluation skills is a priority, as local institutions need to be capable of selecting suitable methodologies, monitoring programme outcomes and applying findings to refine ALMP design. In cases where local expertise is limited, partnering with research institutes or engaging international consultants can enhance evaluation quality. Such partnerships also foster trust and help ensure that evaluations are sensitive to local dynamics, thus producing more contextually relevant insights.

Reliable administrative data is another cornerstone for ongoing monitoring and effective evaluation. By enhancing data collection and quality, institutions can track ALMP outcomes over time, facilitating evidence-based policy adjustments. Mixed-method evaluation approaches, which integrate qualitative and quantitative data, offer a comprehensive view of ALMP impacts, especially valuable in complex environments. Additionally, adapting and disseminating specialised training and evaluation tools tailored to fragile settings is essential to support local and national institutions. Providing targeted training for policymakers, PES staff, and researchers will bolster capacity and lay the groundwork for evidence-based and sustainable ALMP practices.

European and international institutions as well as global and regional research communities should continue their efforts to consolidate capacities for policy evaluation in developing and transition countries, foster innovation in monitoring and evaluation methods, and secure faster transfer of policy impact results into design of new support programmes. Fragile and (post)conflict contexts become the new norm for ALMP implementation throughout the world therefore the ETF recommends a more decisive approach to finetuning and renewal of research methods and investments in sustainable impact evaluation frameworks.

1. INTRODUCTION

Creating inclusive and sustainable jobs is essential for economic and social development. This is particularly critical in situations affected by fragility, conflict, and violence. Jobs drive economic growth and are key to reducing poverty. They can also help promote social cohesion and stability. At the same time, job environments are particularly challenging in fragile situations. Fragility, conflict and violence can affect all levels of society – people and communities, firms and entrepreneurs, as well as institutions. For workers, this often results in reduced opportunities for wage and self-employment, and diminished human capital. As economic activity slows and firms struggle to grow, fewer jobs are likely to be created. For the private sector, fragility results in a highly risky business environment shaped by pervasive market and government failures, which increases costs, reduces demand and compromises investment returns (World Bank, 2016). In these challenging settings, active labour market policies (ALMPs) play a crucial role by contributing to job creation and supporting the transition to labour market.

At the same time, governments often face restricted budgets and increased pressures to use public funds effectively. As a result, there is a need to ensure that resources are spent on services that provide high economic and social returns. This underscores the importance of monitoring and evaluating policies and measures in a regular and rigorous way, which allows for their adjustment to improve their performance vis-à-vis set objectives. At present, regular assessment of the effectiveness or impact of different measures, active labour market policies in particular, is not an integral part of the policy cycle in many transition and developing countries, where also the engagement of public employment services (PES) in such activities is limited and/or ad-hoc, often with limited resources for the implementation of such analytical exercises. At the same time, many ALMPs are implemented in difficult socio-economic contexts which influence the employment probability and other key effects of such policy approaches.

To support knowledge consolidation and foster dialogue on the design, implementation and impact of ALMPs, the European Training Foundation (ETF) launched an initiative to assess ALMP effectiveness in ETF PCs, including those in fragile socio-economic and post-conflict contexts. The aim of this initiative was to draw up specific recommendations for the design of ALMPs and to improve the monitoring and assessment of implemented measures and services.

The aim of this report is to provide an overview of this work, based on the mapping of existing ALMPs evaluations, with a specific focus on impact evaluations. The mapping considered (impact) evaluations that have been carried out by public or private bodies as part of the implementation of public policies/measures and/or (donor) projects, particularly after 2017 (although the scarcity of rigorous impact assessment forced an extension of the timeline). While the geographical focus is on the ETF partner countries¹, other countries with fragile socio-economic contexts were included as well. The review examines several critical aspects of ALMP evaluations:

- Evaluation objectives: key performance and impact questions, information on evaluated measures or services and target groups.
- Research design: applied research methodologies, including experimental and quasi-experimental designs and data collection tools.
- Application of evaluation results: use of findings within the policy cycle.
- Quality and sustainability: issues that limit the quality and sustainability of conducted evaluations.

¹ ETF partner countries cover the following regions: Western Balkans (Albania, Bosnia and Herzegovina, Kosovo*, Montenegro, North Macedonia, and Serbia), Southern and Eastern Mediterranean (Algeria, Egypt, Israel, Jordan, Lebanon, Libya, Morocco, Palestine, Syria, and Tunisia), Eastern Partnership (Armenia, Azerbaijan, Georgia, Moldova, and Ukraine), and Central Asia (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan).

- Review of evaluation outcomes: conclusions on designing effective ALMPs in ETF partner countries, while also addressing fragile and post-conflict settings.

The report is organised as follows: Section 2 describes the methodological approach, Section 3 examines the key characteristics and technical aspects of evaluation approaches in the covered countries, and Section 4 identifies the key features of successful interventions, also considering fragile contexts. The *concluding section* summarises the key lessons learnt from the conducted mapping.

2. METHODOLOGY AND KEY CONCEPTS

The report draws from an extensive literature review that examines evaluations and meta-analyses of ALMPs implemented in ETF partner countries and other selected countries. The review specifically emphasises the needs and interests of the primary target audience for the study: policymakers and strategic managers in the field of employment policy and public employment services.

The main sources of information have been extracted from search engines and the PES websites of the target countries. Publications from organisations such as the Asian Development Bank (ADB), Belgian Development Agency (ENABEL), European Training Foundation (ETF), International Labour Organization (ILO), the German Agency for International Cooperation (GIZ), Organisation for Economic Cooperation and Development (OECD), United Nations Development Programme (UNDP), the World Bank and other organisations, as well as from academia were consulted. The studies consulted include guides, meta evaluations, research pieces on employment policies, as well as roughly 40 evaluation studies. Furthermore, 15 semi-structured online interviews were conducted with the representatives of governments, academia, non-governmental organisations (NGOs), and international organisations of particular relevance to the report. In most cases, more than one representative participated in the interview, which helped extract more comprehensive information. The aims of the interviews were to explore existing capacities and current ALMP (impact) evaluation practices, including those in fragile contexts.

The review primarily focuses on the ETF partner countries. Additionally, it includes other countries with fragile socio-economic contexts to enhance the knowledge base. Due to reasons such as a lack of evaluation culture, insufficient resources for conducting (impact) evaluations, or limited accessibility (with evaluation reports not being publicly available), not all the ETF partner countries were covered. Therefore, the analysis includes the following countries, as shown in Table 1 below:

Table 1: Geographical coverage of compiled evaluation reports

Regions	Countries
Western Balkans and Türkiye	Albania, Kosovo*, Bosnia and Herzegovina, North Macedonia, Montenegro, Türkiye
Eastern Partnership	Armenia, Azerbaijan, Georgia, Moldova, Ukraine
Central Asia	Kazakhstan, Kyrgyzstan
Southern and Eastern Mediterranean	Egypt, Jordan, Lebanon, Morocco, Tunisia
Sub-Saharan Africa	Benin, Ghana, Kenya, Niger, Uganda
Others	India, Mexico

Source: Authors

Definition of fragile contexts

While the review focuses on ETF partner countries, a specific aim was also to examine fragile contexts. In the context of development and global affairs, fragile contexts or fragile countries refer to situations where states, systems, communities or disadvantaged groups face significant challenges due to a combination of exposure to risks and insufficient coping capacities. For the purpose of the study, we consider the OECD (2022) definition that characterises fragility as the combination of exposure to risk and in some cases difficulties for state, systems and/or communities to manage, absorb or mitigate those risks. Fragility can lead to negative outcomes, including violence, poverty, inequality, displacement, and environmental and political degradation. Presently, fragility is measured

by the OECD on a spectrum of intensity and expressed in different ways across economic, environmental, political, security, societal and human dimensions:

- **Economic:** vulnerability to risks stemming from weaknesses in economic foundations and human capital, including macroeconomic shocks, unequal growth, high youth unemployment, lack of diversification of the economy, ineffective public finance management and revenue generation, etc.;
- **Environmental:** vulnerability to environmental, climatic and health risks that affect citizens' lives and livelihoods, including exposure to natural and man-made disasters, climate change impacts, pollution, environmental degradation and disease epidemics;
- **Political:** vulnerability to risks inherent in political processes, structures, events or decisions; these could be linked to the lack of political inclusiveness and representation, lack of transparency and accountability, lack of state legitimacy, etc.;
- **Security:** vulnerability of the overall security context to violence and crime, including political and social violence and specific violence by security sector forces and non-state armed groups, but also human security more generally;
- **Society:** vulnerability to risks affecting societal cohesion that stem from both vertical and horizontal inequalities, including inequality among culturally defined or constructed groups and social cleavages, structural discrimination, gender inequality, shrinking civil society spaces, and so on;
- **Human:** vulnerability to risks affecting the realisation of people's wellbeing and potential, including their ability to live healthy, long and prosperous lives, such as risks to human development in areas related to the formation of human capital, the reduction of inequalities (including gender equity) and vulnerabilities, and the provision of basic social services, including universal health and education.

Fragility can lead to negative outcomes, including violence, poverty, inequality, displacement, and environmental and political deterioration. In fragile contexts, numerous challenges emerge that affect peace, stability, and development²:

- **Political and Institutional Challenges:** institutions may be poorly equipped with resources, including poorly trained staff on both ALMP design, implementation, monitoring and evaluation. In addition, institutional frictions, political polarisation, and scarce resources may hinder ongoing reforms and limit possibilities of an effective programme implementation. In some countries ALMPs may be implemented in the first place by NGOs, who may lack sufficient stability (in financing, support, trust). The availability of training institutions to cover geographical areas of fragile contexts may be another limitation.
- **Poor Economic Conditions:** weak social welfare may limit the effective implementation of ALMPs. A private sector that does not generate enough employment opportunities, large informal sectors with poor working conditions, and a structural limited access to resources is a major challenge.
- **Endemic Corruption:** corruption hinders socio-political and economic development.
- **Migration Pressure:** emigration is widespread due to economic hardship and/or limited security. Emigration may limit the potential for economic development as it creates labour and skills shortages. Immigration of refugees may pose immense challenges for labour market integration into labour markets that show poor demand and are structurally weak. In addition, there are unresolved issues of directing much needed humanitarian aid to those displaced.
- **Ethnic Division:** persisting ethnic tensions have an impact on stability and governance.
- **Gender divide and gender inequalities:** women are found to be often in a more vulnerable position when it comes to access to jobs and capital. They often face multiple forms of

² Named during interviews, addressed in the evaluations reviewed and collected from the reports of international organisations such as GIZ (2015), the World Bank (2023), ADB (2013), and the OECD (2022).

discrimination. Empowering women in socio-economic fragile contexts is extremely challenging and overcoming gender norms is difficult.

Active Labour Market Policies

The European Commission defines labour market policies as *public interventions in the labour market aimed at reaching its efficient functioning and correcting disequilibria and which can be distinguished from other general employment policy interventions in that they act selectively to favour groups in the labour market* (European Commission, 2018).

These public interventions are classified into three broad types (further divided into 9 categories), as per the Table 2 below.

Table 2: Classification of LMP by type of intervention

Services (cat. 1)	Refer to labour market interventions where the main activity of participants is job-search related and where participation usually does not result in a change of labour market status. Services also cover functions of the PES that are not directly linked to participants. This includes placement and other services for employers, administrative functions, general overheads and other activities depending on the responsibilities of the PES.
Measures (cat. 2 to 7)	Refer to labour market interventions where the main activity of participants is other than job-search related and where participation usually results in a change in labour market status.
Supports (cat. 8 & 9)	Refer to interventions that provide financial assistance, directly or indirectly, to individuals for labour market reasons or which compensate individuals for disadvantage caused by labour market circumstance. The participants are usually persons who are out of work and actively seeking work but also persons who retire early from the labour market. Supports may be payable to persons who benefit from services, but financial assistance paid to persons participating in measures should be considered as part of the costs of the measure and not as a support.

Source: European Commission (2018)

Considering the scope of this report, and the coverage of the reviewed reports, ALMPs in public interventions classified under the category *measures* were analysed as shown in the Table 3 below.

Table 3: Definitions of active labour market programmes

Measures	Actions
Training	Classroom training, on-the-job training, internships, incentives to employers to recruit apprentices or training allowances for particular disadvantaged groups.
Employment incentives	Wage subsidies to facilitate the recruitment of unemployed persons and other target groups or help to ensure the continued employment of persons at risk of involuntary job loss.
Sheltered and supported employment and rehabilitation	Subsidies for employment in an enterprise established specifically for the employment of people with disabilities or other working limitations, subsidies for physical adaptation of the workplace (buildings and/or equipment) and the implementation of special organisational arrangements, vocational rehabilitation or training.
Direct job creation	Subsidies for creating temporary jobs, usually of community benefit or socially useful, in order to provide employment for the long-term unemployed or persons difficult to place.
Start-up incentives	Encouraging the unemployed and other target groups to start their own business or to become self-employed through direct cash benefits or indirect support including loans, provision of facilities, business advice, etc.

Source: European Commission (2018)

However, labour market services, commonly referred to as employment services, (including information, counselling and guidance, job-search assistance, etc.) may play an important role for the implementation of some of the ALMPs. Several of the evaluated programmes comprise a combination of different active labour market measures (ALMMs) and labour market services, or only labour market

services, such as the 'Finding a Job Is a Job' programme implemented in Morocco, which comprises only job-search assistance. In this case, these programmes will be included in the category of *mixed interventions*³.

Target groups

In principle, all ALMP interventions benefit disadvantaged groups that could be distinguished by labour market status in three broad categories: unemployed people, employed people at risk, and inactive people (European Commission, 2018).

Target groups of ALMPs in fragile contexts

- **Unemployed people:** persons usually without employment, available for work and actively seeking work. Persons considered as unemployed according to national definitions are always included here even if they do not fulfil all three of these criteria. Among the unemployed, some groups are affected by fragile contexts and may suffer discrimination and other forms of exclusion.
- **Employed people at risk:** persons currently in work but at risk of involuntary job loss due to the economic circumstances of the employer, restructuring, or similar. In countries with fragile socio-economic contexts this may also include underemployed people in the informal economy. Underemployment refers to a lower number of hours worked than wished and income generation at or below the poverty line.
- **Inactive people:** persons currently not part of the labour force (in the sense that they are not employed or unemployed according to the definition above). Among the inactive people, some groups are affected by fragile contexts and may suffer discrimination and other forms of exclusion, including low-skilled rural women, women victims of gender-based violence, refugees, marginalized ethnic groups, ex-combatants, etc.

Source: European Commission (2018) and own elaboration

Based on the reviewed reports and studies, the measures and services covered a broad range of specific target groups. These included economically disadvantaged individuals, such as the working poor, people in low-paid and insecure informal employment, and jobseekers receiving economic aid. Youth facing employment challenges were also a focus, including those under 35, young people with disabilities, unemployed or inactive youth aged 18–35 with no prior formal work experience, and fresh graduates. Additionally, unemployed and at-risk jobseekers were targeted, including those over 45, unqualified unemployed individuals, repatriated jobseekers (within two years), and individuals unemployed for over six months after completing vocational or on-the-job training. Finally, vulnerable and marginalized populations, such as victims of trafficking, gender-based violence, and domestic violence, as well as ex-convicts, were also included in the scope of the interventions.

³ A *mixed intervention* comprises two or more components, each of which may have a different classification by type of action.

3. ORGANISING EVALUATIONS

ALMP (impact) evaluations are essential for assessing the effectiveness, efficiency, and impact of such programmes, informing decision-makers, practitioners and improving programme design and implementation. These evaluations require financial and human resources, and often involve collaboration between Public Employment Services (PES), international donors, and research communities. Furthermore, it is necessary to make decisions regarding research design while considering the available data, the feasibility of primary data collection as well as existing budgets.

In general, when designing evaluations, several key elements must be carefully considered. First, it is essential to clearly define the evaluation objectives, including its purpose, scope, and target audience. Developing a theory of change is also crucial; this framework should outline how the programme is expected to impact the labour market and individual situations, identifying the assumptions and mechanisms through which the programme is intended to achieve its outcomes. Next, the outcomes must be clearly defined, along with the methods for measuring them. This includes identifying both the desired and undesired potential effects of the programme. Defining the target group for the evaluation is equally important, as it allows for consideration of sub-groups that may be affected differently by the programme.

Based on the mapping of available evidence, we review here the evidence gathered on practices regarding the financial and human resources and the legal frameworks that mandate these evaluations in different countries.

3.1 Capacity for effective evaluation

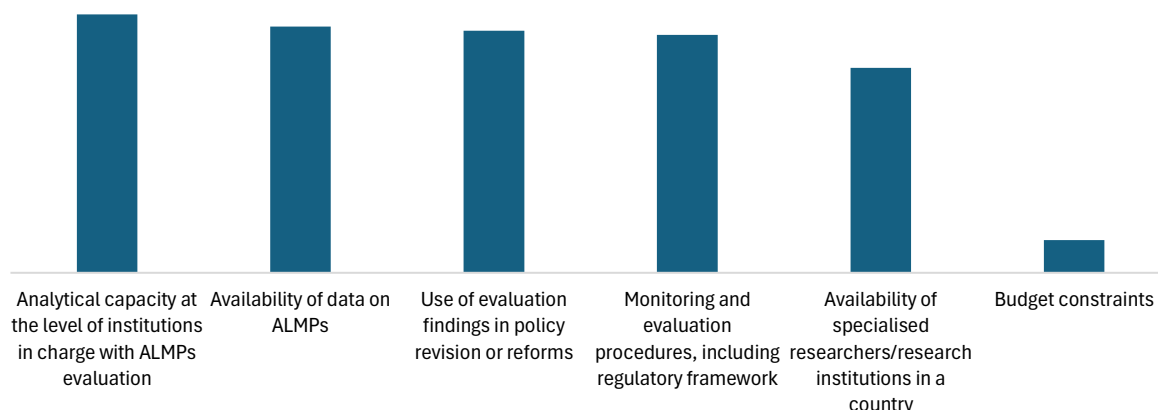
A key issue to consider in planning and conducting the evaluation of ALMPs is the financial and human resources that are available to the evaluation. In several analysed countries, such as Albania or Türkiye, conducting (impact) evaluations of ALMPs is required by law, at least for larger projects and state aid programmes as in the case of Türkiye. Being regulated by law, this ensures that the necessary budget for conducting the evaluation must be provided from the state budget. In this situation, the PES or related ministry launches international or national public procurement procedures, as the case might be, for commissioning the evaluation. In the case of Türkiye, the government has issued guidelines on conducting impact evaluations (Presidential Decree no.102/2022). In other countries, which have had PES in operation for a long time, the PES or related ministries have developed in-house capacities to commission evaluations. Examples include Egypt, Morocco, Mexico and India.

For PES and related ministries to commission evaluations there must also be a sufficient specialised research capacity in the country. However, in many cases, PES institutions face a lack of internal research capacity, making it difficult to carry out complex evaluations independently. To address this gap, PES may rely on other actors. For example, in the case of Ukraine, there is a strong institutional setting for conducting impact evaluations⁴ of ALMPs, such as through the National Scientific Institute, Institute of Economics and Forecasting, Institute of Demographics and Social Research, Institute of Professional Qualifications, etc. Another practical solution is to engage with international organisations to carry out these evaluations, as they bring their own expertise (source: interviews).

The lack of analytical capacity was highlighted also by a survey conducted in 2023 among the representatives of PES from 16 ETF partner countries, which brings an additional insight into the challenges perceived by PES that affect implementing regular ALMP evaluation and impact assessment (see Figure 1).

⁴ See Glossary of Terms

Figure 1: Importance of challenges affecting implementation of regular ALMP evaluation and impact assessment (cumulative ranking)



Source: Survey of PES representatives on flexible and inclusive ALMPs and skills development in ETF partner countries, 2023

To conduct ALMP impact evaluations, it is necessary to strengthen the capacities and knowledge of PES in this area, and within the research, academic, and consultant communities. This usually requires effective cooperation between government institutions and the research community conducting the evaluations. In many interviews, it was observed that the skills required to conduct ALMP impact evaluations were present within the country.

The following strategies have been identified for impact evaluations in ETF partner countries to improve local capacities and opportunities in this field:

In-house capacities

In most of the analysed countries, PES have a dedicated in-house unit (department/division/office) dealing with statistics and labour market analysis. Usually, the work of these units is limited in scope, dealing mainly with administrative databases, and producing various (mainly monitoring) reports.

On the other hand, in Türkiye, impact evaluations are conducted internally. However, there is no dedicated unit solely for impact assessment; instead, data processing and analysis are performed as part of the broader responsibilities of specific personnel. In the Republic of Moldova, the Public Employment Services (PES) is currently exploring the utilisation of administrative data and potential strategies that would enable the monitoring unit to operate in-house on a regular basis.

To enhance PES capacity and broaden their scope of work, one effective strategy is to make use, for example, of training courses (available online, as well) on impact evaluation of policies, programmes and projects. Thus, the staff of these units (or the programme designers, implementers, and leaders of PES) may acquire the necessary knowledge either to conduct net impact evaluations on their own or to better prepare the tendering documentation for outsourcing evaluation services.

Universities

Another strategy identified in this analysis involves developing partnerships with universities to leverage their expertise. For instance, in Egypt, national research capacities are well established, and ministries collaborate closely with universities. Additionally, university professors have been integrated into ministries to manage offices responsible for research and analysis. In other countries in the review, such as Ukraine, academic researchers actively engage in applied policy research, collaborating closely with government bodies and other relevant authorities.

However, in countries where universities and academics have limited involvement in applied work, a cultural shift becomes essential. One effective strategy involves creating dedicated research centres within universities, funded by either government agencies or international donors. These centres can

champion the adoption of rigorous evaluation methods, including full net impact evaluations. Furthermore, universities can play a pivotal role in training the next generation of evaluators through their educational programmes.

Professional community

Establishing a robust professional evaluation community within ETF partner countries is important. These professional communities not only provide mutual support and facilitate learning but also maintain quality standards. A practical approach involves fostering connections with relevant national evaluation societies, encouraging them to promote full net impact evaluations of ALMPs through training events, conferences, seminars and awareness sessions. Additionally, linking these national communities with similar international counterparts, as observed in Egypt, can significantly enhance the know-how and overall capacity for conducting this type of evaluation.

3.2 Budget

Estimating the cost of an (impact) evaluation requires an understanding of the key factors that drive those costs. The budget for an evaluation depends on the scope and complexity of both the subject under evaluation and the evaluation process itself. Importantly, while this resource primarily addresses budgeting for programme evaluation, the questions posed are also relevant to evaluations of policies and operational aspects.

In general, evaluation budgets should be aligned with the expectations and engagement of relevant stakeholders, appropriate for the research design used and key questions to be answered, adequate for ensuring quality and rigour, and in line with the level of programme and organisational resources available. The programme's complexity and scope directly affect the evaluation expenses. The population served, the number of operating sites, type and duration of services provided, and geography are programme variables that affect budget estimations.

During the interviews, it was reported that many ETF partner countries experience a general lack of PES funding, which hampers their ability to conduct surveys or other complex evaluations. These activities are typically carried out with the support of international donor organisations. Conversely, as noted previously, Albania and Türkiye must allocate resources for evaluations from their state budgets. It is generally advisable to ensure that a budget for (impact) evaluations is included in the initial design of the programme.

3.3 Use of results

The evaluation results may be used to decide on whether to continue or discontinue a programme, to adapt design and target groups, and to improve implementation processes. These decisions may be taken by PES, most probably in relation to operational improvements (and depending on their autonomy), the related ministries and the donors. Evaluations may also contribute to assessing the impact of a programme (or group of programmes) and strategies in relation to national priorities set in strategic planning documents with a focus on specific themes or horizontal priorities such as equal opportunities and providing good practice examples. Another purpose is to review the quality and relevance of the quantified objectives, analysing financial and physical progress and providing recommendations on improvements to the programme.

The reporting and dissemination of evaluation findings are crucial steps in sharing the results of an evaluation with various stakeholders. The evaluation team collects and analyses data in order to form conclusions and recommendations. These elements are incorporated into evaluation reports. Two common ways of reporting evaluations are through stakeholder workshops and written reports.

During the interviews it was noted that written reports are mainly disseminated internally, with line managers or the competent ministry, with the sole purpose of meeting accountability requirements. In

some cases, it was mentioned that summary reports are shared with the international donors' community to attract more funding. In general, the necessity of disseminating findings to a broader audience often depends on the organization conducting the evaluation. Research and academic institutions, as well as international organizations, typically prioritize wider dissemination. However, in other cases, sharing reports publicly is not always deemed essential and may only occur when legally required.

Effective reporting and dissemination contribute to informed decision-making and programme improvement or may support the national rolling-out of a successful pilot programme. For example, the Bab Amal pilot programme in Egypt, adapted the internationally recognised Graduation approach by targeting 2 400 low-income households of highest needs in the Assiut and Sohag governorates, and is implemented with the assistance of NGOs and the Sawiris Foundation. It consists of a comprehensive, time-bound, and sequenced set of interventions to transition people out of extreme poverty and into sustainable livelihoods. The evaluation of this programme was based on administrative data analysis and, given the good results obtained, the programme will be included in the social protection programme (J-PAL MENA, 2022).

However, during the desk research for this study, the research team frequently encountered difficulties in finding (impact) evaluation reports of ALMPs from public sources. This highlights a failure in practice to share these findings with the general public.

4. IMPLEMENTING EVALUATIONS

This section examines the various methodologies and approaches to data collection used in the evaluations examined, with a primary focus on impact evaluations.

4.1 Evaluation design

In general, there are several key approaches that can be used to assess the outputs and outcomes of ALMPs. Monitoring is a continuous way of collecting and analysing information to track the execution and performance of a project, programme, or policy against expected results. It involves tracking key indicators of progress, such as the employment status of participants sometime after programme participation. Monitoring is a valuable source of data for the performance management of an organisation as well as for impact evaluations as a basis on which to evaluate outcomes of the intervention. Monitoring results are then compared to results achieved without the intervention or are compared to available macro data. Monitoring programme implementation can be challenging in fragile contexts due to factors such as varying implementation organisations or insufficient institutional capacity to collect data reliably and coherently (see section 3.2 below).

Evaluation is then a systematic, objective assessment of an ongoing or completed project design, implementation, and result to determine its relevance and the fulfilment of objectives, efficiency, effectiveness, impact and sustainability (Hempel and Fiala, 2012). Operational evaluation (sometimes called process evaluations) examines how efficiently and effectively programmes were implemented and whether there are gaps between planned and realised outcomes, and whether the programme is ultimately well targeted and relevant. Process evaluation particularly looks into the efficiency of the organisation of processes.

Impact evaluations answer cause and effect questions to determine whether and to what extent an intervention caused observable change. Understanding impact requires the effects of the intervention to be isolated from other factors, influencing beneficiary outcomes (ILO 2018). Criteria that are looked at in evaluation studies include coherence, relevance, effectiveness, efficiency and impact (OECD 2023).⁵

To understand the programme's impact, comparing observed outcomes with a reference situation—one that represents what would have occurred without the programme—is essential. Isolating the causal relationship between the programme and the outcomes requires eliminating other factors that could have influenced the observed results. The evaluation should also determine the appropriate time period for assessing the programme's impacts, including any potential lock-in effects.

Furthermore, indirect effects, such as deadweight, substitution, and displacement effects, should be considered and accounted for in the evaluation. In the case of deadweight effects, resources invested in the programme benefit individuals who would have been hired regardless. Substitution effects occur when a programme incentivises employers to replace one type of worker with another to reduce labour costs; for example, subsidies for teenagers may encourage firms to hire teenagers over young adults (ILO, 2016). Displacement effects can arise when employment generated by the programme replaces regular employment—for instance, firms may hire subsidised workers instead of unsubsidised ones, or dismiss subsidised workers when the subsidy ends.

On the positive side, indirect effects may include multiplier effects, where additional income generated by the programme stimulates demand and creates more jobs, or spillover effects that benefit the broader economy. A net impact evaluation should aim to estimate these indirect effects and deduct them from the gross effect for an accurate assessment of the programme's impact⁶.

⁵ See Glossary of Terms

⁶ https://assets.publishing.service.gov.uk/media/5a7568e5e5274a467f7e44e0/Additionality_Guide_0.pdf

Impact evaluations that have been collected and reviewed for this report include a variety of approaches and methods. Some evaluations include process evaluations, others focus mainly on comprehensive monitoring and evaluation of gross effects, and some conduct a thorough counterfactual evaluation while also providing high-level data for comparison with general labour market development. Additionally, they may examine procedural factors that could have influenced the results. Counterfactual impact evaluations cannot be carried out for any ALMP of any size, as some requirements need to be fulfilled to carry out meaningful evaluations. Moreover, counterfactual impact evaluations require a high level of technical knowledge and expertise in the area of quantitative analysis and may be costly to assure robust results.

Data collection includes both quantitative approaches (based on statistical data and methods to analyse them, including econometric methods and models) as well as qualitative methods (interviews, written answers to questionnaires, focus groups, field visits, documentary analysis, etc). Evaluations that are publicly available have mainly been conducted for large programmes. In some instances, the approach to programme development and implementation involves initiating a pilot phase, which then requires evaluation. Some evaluation studies also include cost-benefit analysis. In addition, for this report it was possible to collect some non-published evaluations.

4.1.1 Outcome (gross impact) evaluations

Some countries analysed have conducted an assessment of outcomes aimed at determining the gross effect of a programme (without seeking to eliminate other factors that could have influenced the outcome of the programme, which is the objective of net impact evaluations).

This type of evaluation can provide data on the performance of the target group for a particular programme and can be roughly compared with other data such as on the performance of the overall unemployed or on another similar geographical area. Preference is given by PES to this type mainly because it is relatively easy to do with low costs. Most of the administrative data are available in house and other data may be obtained through data-exchange protocols with other relevant public bodies, such as the tax authority or ministry of education, as can be observed in the case of Kosovo or Kazakhstan. However, it only gives basic facts regarding the gross effect of the programme and offers little point of comparison.

Advanced monitoring tools include monitoring the labour market status (employed, unemployed, in education and training, inactive for other reasons, possibly income data) of the programme participants after participation in the programme (e.g. after a period of 3, 6 or 12 months, or longer). This requires combining different datasets and may thus require cooperation between different stakeholders (e.g. with the tax authorities). Using anonymised data for research purposes allows a summary to be made of the trajectories of participants with certain characteristics. This type of analysis delivers useful information for policymakers, although it cannot establish clear causalities between programme participation and outcomes. Challenges in practice may be linked to IT-integration of datasets from different institutions and individual data protection.

Example: Kazakhstan

The Department of Forecasting and Research of Workforce Development Centre (WDC) is responsible for conducting labour market analysis and evaluations. Based on the analysis of databases and statistical data, (quantitative analysis only), they provide recommendations to the government for developing and/or adjusting ALMPs.

Kazakhstan is a 'fully digitalised' country and, therefore, WDC have access to all the required data and information for their evaluation requirements. For example, if the system highlights a person in need, an automatic message is sent to that person asking if he/she requires support. If the answer is positive, then they contact the person and showcase what support might be provided. If support for starting a business is required, funds are disbursed automatically to the individual's account, since they have all their data in the system. If the person needs to improve their skills, then they are transferred to the 'enbek skills' platform.

To find out if the 'entrepreneurship' programme is successful, WDC checks the status of individuals after 12 months to establish if the business still pays taxes. If the result is positive, this means that business is sustainable and the project successful. Verification is completed on all participants as WDC have all their data.

For the 'short-term education at workplace' programme, WDC checks after 3 to 6 months, as appropriate, to identify if the participant has acquired a permanent job. In addition, they check if this job meets a defined 'quality criteria'. WDC has developed a set of Key Performance Indicators with the World Bank to support the assessment of job quality. A set of descriptors has been developed, for example, if a person has been employed in a permanent job for at least 6 months, is one such quality indicator.

Source: Interview

Other approaches used to monitor employment outcomes after programme participation consists of implementing tracer studies and collecting information from beneficiaries some time after the end of their support (e.g. as in Albania and see section 3.3).

Moreover, in some ETF partner countries, PES (with the support of an international organisation) or donor organisations (e.g. ENABEL) conduct client satisfaction surveys, possibly among jobseekers and employers. For example, in Kosovo, a client satisfaction survey of jobseekers registered with PES was launched by UNDP in 2021, by hiring a research institute to conduct a survey (interview). Client satisfaction surveys have, for example, also been used in evaluation reports carried out by the relevant institutions in Mexico (for example, in Mexico City, Secretaría de trabajo y fomento al empleo 2023). In this specific case, client satisfaction was one of the various tools used for the evaluation report. A limitation of client satisfaction survey for impact evaluation purposes is that clients might be satisfied that they have access to a service, even if that did not help them much to find a (better) employment. Client satisfaction surveys may nevertheless be an important tool for the PES to measure the quality of service delivery and link it to performance management tools (European Commission, 2016).

4.1.2 Counterfactual net impact evaluations

Counterfactual (net impact) evaluations focus on measuring as accurately as possible the impact of the programme by controlling for factors that could have an influence on the outcome (such as, for example, the business cycle). The results of the impact evaluation are intended to inform policymakers whether they want to reconduct the programme, scale it up or give it up. The results can, however, also be used as an indication of whether the programme design should be readjusted (for example, because the net impact is higher for a certain target group than for another) and may require additional research to understand whether the impact could be increased by a change in implementation processes.

Overview of methods to carry out net impact evaluations

Experimental design: experimental designs rely on some elements of randomisation in the allocation of participants into treatment and comparison groups. They can produce highly credible impact estimates but are often costly and, for certain interventions, difficult to implement. Within experimental design, the *Randomised Controlled Trial (RCT)* is a specific method in which individuals are randomly assigned to receive an intervention or not, ensuring statistically comparable groups in terms of observed and unobserved factors.

Quasi-experimental design: quasi-experimental design approaches are used to construct a valid comparison group by using statistical means to control for differences between the individuals treated with the programme being evaluated and those not treated.

Key quasi-experimental methods include:

Before and After Comparison of Programme Participants: this approach measures changes in participants before and after their involvement in the programme, but it lacks a non-participating group for comparison, limiting the ability to account for external influences.

Comparing participants' and non-participants' status after programme participation: this method assumes that participants and non-participants are identical except for programme participation. Comparable groups are created based on observable characteristics (e.g. age, gender, education, employment history). However, limitations arise due to unobservable factors such as motivation or socio-cultural factors, which can introduce bias, especially in fragile contexts. Some evaluations supplement this with qualitative research to better understand variable meanings and reduce potential biases.

Difference-in-differences approach (before and after comparison of participants and non-participants): this measures the change over time of outcomes of participants as compared to non-participants. In this approach it is assumed that if the programme did not exist, the two groups would have had identical trajectories over this period. One key challenge is, for a number of reasons, this assumption may not hold.

Propensity score-matching: individuals in the treatment group (participants) are matched with non-participants who have similar observable characteristics (the objective is to have as similar as possible characteristics). The average difference between matched individuals is the estimated impact. This approach needs to assume that other factors that have not been considered for the matching (either because they are unobservable, or for other reasons) do not bias the outcome. In general, applying this method to fragile contexts requires additional efforts to be made to detect correlations of these factors and the outcomes. This may be a particularly challenging exercise in fragile contexts, where a number of context-related factors (in addition to the usual socio-demographic variables) may influence the outcome.

Regression discontinuity design: individuals are ranked based on specific measurable variables, usually socio-economic background variables. It is assumed that, after controlling for the ranking criteria, any remaining differences between individuals directly below or above the cut-off score are not statistically significant, and thus, the treatment effect can be estimated without significant bias.

Source: ILO 2018

The evaluations reviewed utilised external data sources such as household surveys (e.g. Bandiera et al. 2023, Muralidhara et al. 2022) and labour force surveys to characterise the target group or perform before and after comparisons. Moreover, counterfactual impact evaluation methods were utilised mainly for major donor-funded programmes or primary national ALMPs. Additionally, administrative data and surveys specifically conducted for evaluation purposes were employed. Sampling methods are typically required for both administrative data and for surveys involving participants and non-participants.

In general, the counterfactual impact evaluation reports reviewed for this report used large sample sizes with more than 1 000 respondents and up to several thousands (with a few exceptions using sample sizes between 150 and 400 participants, especially in the case of smaller countries, such as Albania and Bosnia and Herzegovina).

For example, several evaluations of business training programmes (from various countries) have struggled to find significant impacts. A key reason for this has been the use of relatively small samples with heterogeneous firms. This lack of statistical significance has been interpreted by some as

evidence that training seldom works, rather than the conclusion that there is a lack of evidence as to whether it works or not (McKenzie et al. 2019).

Example: Kenya

Background: the International Labour Organization's (ILO)'s Gender and Entrepreneurship Together training programme (GET Ahead) seeks to enhance women's opportunities in entrepreneurship through knowledge and skills development in business and management. The programme brings gender perspective and uses an interactive approach to acquaint women, primarily from low-income settings, with business and working environments, development of business ideas, product design and management practices. The programme offered women with small businesses in rural markets a five-day training course in Kenya. A year after training, some women who had received GET Ahead were offered complementary group and individual support services over a five-month period.

The baseline sample encompassed 157 markets and 3 537 firms in four counties in Kenya, which are largely rural and with most of the population living below the poverty line. While the treatment impacts of business training programmes were not greater than in past studies, they were more precisely measured, revealing positive effects. Notably, these benefits for the trained women did not disadvantage other women in the same markets, and the impacts were stronger three years post-training compared to just one year.

Method: Quantitative Research, Randomised Control Trial.

Source: Mc Kenzie et al., 2019

Example of using available institutional surveys for carrying out a counterfactual impact evaluation of an employment incentive programme in Morocco

Background: to combat high youth unemployment, a subsidised employment programme that supports the transition of youth from school to work through company placements has been implemented (Idmaj). This is a government wage subsidy programme implemented by the Moroccan PES ANAPEC.

Method: Net impact evaluation using the propensity score matching method

A useful data source for this evaluation was a random survey conducted in 2010 by the Ministry of Labour in coordination with the Public Employment Services ANAPEC. Its objective was to provide an accurate description of the socio-economic situation of the interviewees and assess the effect of the "Idmaj" programme on their professional trajectories. While it was not completely designed for impact evaluation purposes, this survey fulfilled three characteristics that, according to the available literature, support the estimation of the causal effect by the matching method. This survey focused on two samples constituted by systematic random sampling with equal probability from (i) a nominative list of beneficiaries of the employment incentive programme "Idmaj" insertion contracts (IC) in 2006 and 2007, and (ii) a list of job seekers enrolled in ANAPEC, eligible for the programme but never having benefited from it since their enrolment until the day of the survey. The retrospective questions of this survey aimed to collect information on the activity status of interviewees during the following four phases: (i) before registering with ANAPEC; (ii) between registration and signature of the IC; (iii) during the IC period; and (iv) the post IC period. After the sampling, the survey was conducted on a sample of 2 500 beneficiaries and 500 individuals for the control group.

Source: Chatri et al., 2021

Target groups and control groups

The key challenge of any counterfactual impact evaluation is to be sure the control group has the same characteristics as the treatment group. This challenge is not specific to evaluations conducted in fragile contexts. However, more efforts have been undertaken to invest in understanding the characteristics of the target group of the ALMP (see also below).

Example: Evaluating a programme for empowering women in rural areas in Egypt

Background: in Egypt a programme was evaluated that provided an integrated approach to female economic and social empowerment to enhance the transition to work for young, marginalised women in rural Upper Egypt. The programme provided business, vocational and life skills training in addition to

actual support in starting a business or becoming employed. The programme was designed to be implemented in 30 villages.

Method: the 30 villages were chosen a priori in Upper Egypt. During the preparatory phase prior to implementation, a group of 15 control villages was selected from the same three governorates that are comparable to the 30 chosen villages. The control and intervention villages were matched in terms of village size, poverty level, education prevalence and labour market-related variables using the 2006 census of Egypt. Given the lack of randomisation in the intervention, there is a chance that the difference in labour market outcomes between treated and control groups is not a result of the intervention but could instead be due to endogenous factors that affect outcomes differently for the treated and control groups. To account for this possibility, the evaluator investigated whether the differences in labour market outcomes existed prior to the intervention. For this purpose, they exploited a detailed section on the mid-line survey on the employment history of the women as well as data from the end-line survey where the same section of questions was asked to cover the period between the two surveys.

Source: Elsayed and Roushdy (2017)

Timespan of evaluation

Depending on the methodology used and the organisation of the evaluation, a first round of surveys (and potentially additional interviews and focus group discussions) are carried out in the design phase of the evaluation, before or at the beginning of programme participation, mid-term during participation, and after participation.

It is commonly recommended to conduct evaluations after 6, 12 and up to 36 months since the end of participation in the programme, aiming to capture both short-and long-term effects on participants. The most relevant length of post-programme period to be considered depends on the type of active labour market programmes implemented. For example, training programmes are likely to have a more positive effect after a year or two, or even longer. The same would apply for self-employment support, as it usually takes some time after creating a business for it to become profitable. Employment incentives can be expected to have a measurable impact sooner.

In fragile contexts, there are more uncertainties around when the impact of a programme can be expected. This is linked to a higher uncertainty of performing economic activities through employment or unemployment. At the same time, evaluation reports conducted in fragile contexts report difficulties in following up participants over a longer period (see section 3.2). In the reviewed evaluation studies for the post-programme evaluation, the time most often spanned 10-14 months. In one case, the period was 18 months (Niger, Bossuroy et al 2021) and 68 months (Uganda, Bandiera et al. 2023). In another case, interviews on a qualitative evaluation study were conducted four years after programme participation, to deepen the understanding of a quantitative net impact evaluation that brought about unexpected results (interview, Benin). For large, multi-annual projects, one practice identified in Albania requires impact evaluations to be conducted every four years from the start of the programme implementation⁷.

In most of the evaluations reviewed for this study, at least two time periods have been selected for post-programme surveys (for example, after 2 and 4 months, 4 and 14 months, 6 and 12 months, 6 and 18 months). Short-term and medium-term impacts may indeed differ. It would often be desirable to measure the longer-term impact (in the case of training measures, positive impacts will be higher after a longer period (Card et al, 2010). However, there is a trade-off to be made between the possibilities for tracking participants and keeping the attrition rate low and the assessment of long-term impacts.

4.1.3 Assessing pilot projects

While full net impact evaluations need a large number of programme participants and are thus appropriate for evaluating large ALMPs, they are less suited for small-scale projects. In addition, to

⁷See (for example) Article 5(3), of the Decision of the Council of Ministers No. 17 dated 15.01.2020 on the procedures, criteria and rules for the implementation of employment promotion programmes through employment, on-the-job training and professional practices. (<https://www.puna.gov.al/rrethnesh/bazaLigjore>)

design and implement large programmes in a most effective way, pilots may be implemented first, before roll-out over the whole country. Pilots may also be implemented to test innovative approaches and procedures and are intended to inform policy-making. Evaluating the pilots thus has a high policy relevance in practice. The methods used include a generally detailed documentary review of processes, analysing administrative data and carrying out interviews both with implementing institutions and beneficiaries, performing fieldwork and using other qualitative research tools.

One limitation is that the impact of programme participation can only be roughly assessed, therefore, the results will be less robust than in the case of counterfactual impact evaluations. One advantage is that this approach allows to identify both promising approaches for implementation and implementation challenges.

Example: Albania

Under the framework of the 'Promoting Inclusive Labour Market Solutions in the Western Balkans II' regional project, UNDP Albania worked with the National Agency for Employment and Skills (NAES) and municipal actors to design an Integrated Case Management model (ICM), a targeted and inclusive package of social and employment services, in line with the specific and individual needs of the marginalised communities. The package was piloted in the municipalities of the region of Elbasan, through financial support to third parties. About 120 beneficiaries were selected and coached toward employment opportunities and social integration for a four-month period. The main target of this intervention were individuals who received social assistance and support, aiming to create an empowering environment for their sustainable transition from social assistance to employment. The ultimate objective of this intervention was to inform policymakers on preparing a new Employment Promotion Programme, which would diversify the portfolio of NAES and ultimately improved the employability of unemployed jobseekers.

The chosen capitalisation of the knowledge process aimed to build up capital from information or knowledge available in an organisation or project to make it available to others - institutions or actors. This capitalisation process was organised in the following steps:

During the preliminary phase, general information was collected about the project. Introductory meetings were held with each organisation's project management teams. During the planning phase, monitoring pillars were carried out. The capitalisation process focused on seven main pillars: i) participating organisations; ii) target group of beneficiaries; iii) collaboration with NAES and other stakeholders; iv) implementation of coaching cycles; v) employment support services; vi) social protection services; and vii) success indicators. This involved a detailed documentary review (e.g. project proposals, technical databases, job descriptions, lists, agreements, agenda of activities, meeting minutes, attendance sheets, monitoring reports, coaching for employment manual and respective templates, participants' development plans, and other relevant internally produced documents).

During the execution phase, the fieldwork activities for collecting, organising, and analysing information were performed. Project-related documents were collected and analysed, including questionnaires, checklists, interviews, brainstorming, longitudinal analysis, field observations, mid-term, and final reviews. For the dissemination of the experience, best practices from this experience were identified, described and recommendations were formulated.

This approach led to the following results: the quantitative and qualitative data indicated that credible and satisfying results toward employment and social support services could be provided to 120 beneficiaries. The approach allowed areas of improvements for the implementation to be identified. It highlighted implementation challenges, such as the selection of participants, the reluctance of the participants to complete the coaching cycle processes, project duration, the participants' lack of trust toward state institutions and programmes, or the participants' perception of group meetings as unnecessary.

Source: Austrian Development Cooperation, AKPA, UNDP (2022)

4.1.4 Analyses of cost-benefits and cost-effectiveness

Cost-benefit analyses are rarely conducted in ALMP evaluations (as asserted in Card et al. 2018), although cost-benefit analysis can contain precious information for policymakers.

In two examples of the reviewed evaluations, net impact evaluations also contained a cost-benefit or cost-effectiveness analysis (Albania (UNDP 2019), Türkiye (impact evaluation of employment

incentives in small companies)). In both examples, the costs of the programme included mainly expenditure for the programme (or foregone social security contributions in the case of Türkiye, where an employment incentive was implemented consisting in an exemption of social security contributions with the objective of supporting employment contracts being given in micro and small enterprises). Benefits included the payment of social security contributions for those who were employed after finishing the programme (identified on the basis of data retrieved from the counterfactual impact evaluations that were carried out).

In the case of Türkiye, the approach had limitations, such as not accounting for or monetizing all costs and benefits. For example, it overlooked the increased income of informal workers who became contracted and any rise in tax revenues from increased consumption due to higher incomes. Additionally, it did not consider government costs related to providing social assistance to those without formal employment. While the subsidy's benefit to the government was the increase in social security premiums and tax payments from more registered workers, benefits like higher tax revenues from consumption and reduced social assistance costs were excluded. As a result, the analysis likely underestimated the subsidy's benefits (Aşık et al., 2022).

In the case of Albania, survey results among participants on their programme participation and post-programme employment situation were collected. Based on this data, the costs (such as employment subsidies and training expenses) and benefits were calculated. The benefits for society included income generated by participants who found employment after the program, as well as social security contributions and income taxes paid. However, the report noted some limitations of this approach, as it could not assess implicit costs and benefits like leisure time, home production, fringe benefits, or the time and effort spent by National Employment Service employees. Additionally, the analysis did not fully account for additional employment generated by the program (UNDP 2021).

4.2 Data collection and data requirements

Running impact evaluations, including those in fragile settings, requires careful consideration of data collection methods and specific data requirements to ensure accuracy and relevance in measuring programme outcomes. Key data collection tools include surveys, interviews, focus groups, participatory observations, and expert opinions. Administrative data, including registries, are especially valuable for labour market programmes, but when such data are unavailable, which is a likely situation in fragile settings, surveys often serve as a primary data source. In fragile contexts, however, unstable institutional support and limited data completeness can complicate the collection process, particularly when assessing vulnerable populations. Thus, qualitative research methods can be preferred instead.

Indeed, the evaluation studies reviewed often incorporated qualitative research methods, such as interviews and focus groups (in addition to large surveys and the exploitation of administrative data). In line with a theory-based approach, some evaluation studies have used qualitative methods as the main research method to deepen understanding of results obtained through quantitative measures. For instance, in Armenia, the World Food Programme used qualitative research to understand socio-economic integration barriers for refugees, leveraging focus groups and informant interviews.

Particularly in fragile contexts, employment issues faced by target groups are often under-researched, making it particularly beneficial to employ both qualitative and quantitative methods. The use of qualitative techniques helps to capture the nuances of labour market dynamics and the lived experiences of individuals, which may be overlooked in purely quantitative studies. In the following boxes different examples are shown that mix quantitative and qualitative approaches to data collection.

Example of a qualitative study evaluating an entrepreneurship programme in Benin

An impact evaluation of a youth employment programme implemented in Benin (PEJ), conducted by the World Bank's Gender Innovation Lab (GIL), was designed to measure the impact of a programme that

offered life-skills, entrepreneurship training and financial support to self-employed and micro-enterprises. The GIL impact evaluation showed that the effects of the programme differed depending on the package of support. The life skills and entrepreneurship training generally had a positive impact on socio-emotional well-being, women's control over their own resources, and productive investments, and translated into a consistently positive effect on earnings for both women and men. In contrast, the cash grant alone had no overall impact for men and had a short-term negative impact on earnings for women, despite evidence of business investments. The objective of the qualitative study was to understand why cash grants were found to be less effective.

The chosen method aimed to capture personal stories about finances and labour market activities through in-depth individual interviews carried out in French and local languages, allowing for unrestricted narratives.

Limitation: interviews were conducted only in the South of the country (not in the vulnerable zones in the North). The qualitative interviews gave some useful insights but could not fully explain the quantitative results. Questions of 'why' remained. Based on the authors, going into more detail would have been needed to fully understand the complexity related to how gender roles in the household played a role.

Source: Interview with World Bank – Gender Innovation Lab

Example from an evaluation of a training and empowerment project for refugees and Lebanese population in Lebanon

Background: Lebanon has faced an additional crisis with the influx of Syrian and Palestinian refugees since 2011. UNICEF developed the Skills Training project with funding from the Government of Germany, implemented from mid-2020 to mid-2023. The programme supported youth (aged 15–24) to build skills they needed to access employment and income-generating opportunities. The skills training programme, which operated under the umbrella of UNICEF Lebanon's Youth and Adolescent Development Programme, had three objectives, including (1) increased access to competence-based training (CBT); (2) increased employability and income-generating activities; and (3) enhanced empowerment and personal development. It targeted both Lebanese and refugee (Syrian and Palestinian) youth. Young people participated in (1) basic literacy and numeracy training (BLN) Level 1, Level 2, or both; (2) Competence-based Training (CBT); (3) life skills training; and (4) employment support services or cash for work programme.

Evaluation approach: a combination of quantitative and qualitative approaches. Qualitative interviews were conducted with UNICEF, training institute staff, Directorate General of Vocational Training and Education (DGTVE) staff, and drop-out youth. In addition, focus group discussions (FGDs) with youth (including six story circles), parents of youth, and training facilitators were held. The research team also conducted Most Significant Change (MSC) analysis workshops with youth to review the stories of change and selected the most significant stories. While the quantitative approach identified the statistics prevalence, significance, and differences in intended programme outcomes, the qualitative approach further explored these dimensions from the personal experiences of beneficiary youth, their families, programme trainers and employers, implementing partners, and key stakeholders.

Source: Morey et al., 2023

Example: Jordan and Türkiye

Background: in early 2016, the German Government launched the 'Partnership for Prospects in the Middle East'. This includes cash-for-work projects that create short-term jobs and income-generating opportunities, and the payment of salaries for teachers to teach Syrian child refugees. A further component added recently involves interventions for vocational training and the promotion of business start-ups.

Initially, the plan was to create 500 000 cash-for-work (CfW) jobs within two years, operating based on a multi-donor trust fund. This involved substantial time pressure and a stringent reporting system. Given the difficult and very divergent framework conditions in Syria's four neighbouring countries (Türkiye, Jordan, Lebanon and Iraq), this entailed considerable challenges which applied both to appropriate conceptual preparation and to the implementation. The cash-for-work programme had a duration of

mostly six weeks, in some cases up to eight months. Typical tasks included waste collection and work in the recycling market.

Objectives of the evaluation: to investigate how short-term assistance and development cooperation can be effectively combined. It aimed to determine which interventions and instruments were suitable for mitigating the effects of crises on both refugees and the host communities, and for improving people's lives sustainably. It was conceived as 'strategic' evaluation as it served to inform the Ministry's strategy design.

Methods: the evaluation applied a combination of different methods, including rigorous impact measurements in conjunction with a quasi-experimental design involving two phases of surveys among a large number of refugees and other vulnerable target groups in Jordan and Türkiye. A difference in difference method was applied.

The evaluation team applied a **context-based method, which involved a triangulation of methods:** desk research, pilot case study (interviews, to prepare survey questions) in Jordan, qualitative country profiles in Türkiye and Jordan (focus groups, expert interviews), quasi-experimental panel survey with treatment and a control group. Desk research and the pilot case study helped to build the model of the impact evaluation.

The research team evaluated the impact of the programme on different empowerment dimensions for the beneficiary, derived from a theory of change of approach: (i) material empowerment (being in employment, earnings; however, long-term impacts were not measured); (ii) social empowerment (relations between the refugees); (iii) cognitive empowerment (learnings through the measures); and a perceptual dimension (how the Turkish people looked at the Syrian and vice versa).

Limitations / barriers / challenges

- The choice of Türkiye and Jordan was linked to the extreme crisis and security situation in the other countries. Therefore, the most fragile contexts in Lebanon and Iraq were excluded.
- Getting access to people was challenging as well as formulating panel questions in a context-sensitive way.
- Construction of the control group was more challenging in Türkiye as here the evaluation was conducted after termination of the programme. Constructing the control group was easier for the ongoing programme, as the control group was constructed from the waiting list.
- There was a need to understand gender aspects in the fragile context.

Source: Interview and Roxin et al. 2021

Example: India

Background: a reform to the National Rural Employment Guarantee Scheme (NREGS) in India introduced in 2016 was evaluated. Over 600 million rural residents are eligible to participate in the NREGS scheme. The reform introduced a new system (biometric 'Smartcards') for making payments. Previous work showed that introducing the Smartcards reduced leakage of funds, increased programme earnings, reduced payment delays and the time required to collect payments, and increased real and perceived access to work, without changing fiscal outlays on the programme. One important finding of the evaluation was to show the effectiveness of a better implementation of an ALMP, in particular in fragile settings with high corruption and instability.

The objective of the evaluation was to assess the effects of the programme on poverty reduction, both through the direct income provided to participants and through broader economic effects. The counterfactual impact evaluation, using an experimental approach, was applied to estimate the effect on poverty reduction, including the impact on income, wages and employment outcomes. Key results showed that the reform raised beneficiary households' earnings by 14%, and reduced poverty by 26%. Importantly, 86% of income gains came from non-programme earnings, driven by higher private-sector (real) wages and employment.

The fact that both wages and employment increased is central to the large income gains estimated: wage gains were not offset by reduced employment but instead were amplified by increased employment. To analyse this further, the researcher considers three broad sets of (non-mutually exclusive) mechanisms for the increase in wages and employment: (1) an increase in labour productivity; (2) an inward shift in labour supply in the context of imperfectly competitive labour markets (e.g. oligopsony); and (3) increases in the aggregate demand for locally produced goods and services.

Source: Muralidharan, 2022

Using a mixed-method approach to data collection in net impact evaluations offers significant advantages, particularly in complex contexts such as fragile labour markets. Combining qualitative and quantitative methods allows researchers to leverage the strengths of each approach; quantitative data provides measurable and statistically generalisable results, while qualitative data adds depth and context to these findings. For instance, quantitative surveys can assess the economic impact of an employment programme, while qualitative interviews reveal personal experiences and nuances, such as why certain aspects of the programme resonate with participants in different ways. This comprehensive approach has proven useful in various case studies, such as youth employment programmes in Benin and Lebanon, where the combination of methods has highlighted not only the measurable outcomes of training programmes but also participants' personal growth and empowerment.

However, there are limitations and challenges to mixed-method approaches. Integrating data from both methods requires time, resources and expertise, especially in sensitive or resource-constrained environments. For example, in Lebanon, conducting evaluations during economic instability may influence participants' responses and the reliability of qualitative data, as contextual challenges could distort participants' focus on programme-related issues. Additionally, while qualitative insights help explain quantitative results, they may fall short in fully capturing complex issues such as gender dynamics, requiring further specialised research.

Evaluating programme impact also involves specific challenges related to data reliability and respondent engagement. For example, in Türkiye's ESSN programme, vulnerability among refugee populations was measured using a multidimensional survey approach, but limitations arose due to its phone-based format, which restricted data depth. Similarly, in Egypt's youth employment programme, data collection through baseline and exit surveys helped assess employment outcomes, but short follow-up periods and implementation challenges limited the evaluation's impact. Strategies such as tracer studies, used in programmes such as Morocco's youth employment initiative and Mexico's Evolucionaria internship programme, aim to track long-term outcomes by repeatedly contacting participants post programme. Yet, these studies often face obstacles such as low response rates and challenges in reaching vulnerable participants, underscoring the importance of robust data management tools and flexible data collection strategies to navigate the complexities of fragile environments effectively.

Based on the reviewed literature, it is clear that conducting surveys and interviews prior to programme participation may be beneficial for different reasons. For example, the World Food Programme (WFP) mentioned above involved focus group discussions with 162 refugees and 50 local residents across five regions in Armenia, as well as 13 key informant interviews, and provided valuable insights into the challenges faced by both groups, along with potential pathways to establish sustainable livelihoods for refugees and local communities (World Food Programme, 2024).

Surveys and interviews also play a crucial role in supporting net impact evaluations, as they provide the baseline data necessary for constructing control groups and correcting biases. This information is essential to accurately measure the programme's impact. Additionally, conducting surveys and interviews helps to define the desired programme outcomes, as seen in previous evaluations (Roxin et al., 2021). Evaluators can further use these tools to assess how effectively the programme targets its intended population by comparing key indicators between beneficiaries and non-beneficiaries. An example of this approach is seen in the Türkiye Emergency Social Safety Net (ESSN) Refugee Survey, conducted by the World Bank Group and WFP, which analysed differences in indicators as a primary research outcome (World Bank Group and World Food Programme, 2024; Cuevas et al., 2019).

Example: Türkiye

Background: Türkiye was hosting 4 million refugees living in the country by December 2018. The overwhelming majority, 3.6 million, were from Syria. The remainder originated mainly from Afghanistan and Iraq. In November 2016, the Emergency Social Safety Net (ESSN) programme was introduced with the objective of supporting the most vulnerable refugees in meeting their basic needs through monthly cash transfers (with 1.5 million beneficiaries as of December 2018). The programme was funded by the European Union member states and implemented nationwide in partnership with the Ministry of Family, Labour and Social Services (MoFLSS), the World Food Programme (WFP), and the Turkish Red Crescent (TRC). A study by the World Bank and the World Food Programme was carried out with the objective of getting a comprehensive view of the vulnerability situation of refugees eligible for ESSN and an assessment of how well ESSN targets, supports and protects the most vulnerable refugees.

Methodology: a central element of this study consisted in defining a multidimensional concept of vulnerability in the context of a programme targeting refugees: the evaluators believe that the following dimensions provide a relatively comprehensive account useful for a study of household vulnerability: poverty and resources; access to key resources (education, housing, health services, legal status); food security; and capacity to cope.

An important indicator for evaluating the performance of any assistance programme is its coverage of the target population. In this case, it is measured by the share of eligible population among the poor. Another key indicator is the adequacy of the transfer size. Benefit adequacy represents the ratio between the monthly value of transfers that eligible households will receive and their pre-transfer monthly budget.

The main data source for the analysis is the Pre-Assistance Baseline (PAB) survey, which was collected by Turkish Red Crescent (TRC) as part of the monitoring and evaluation of the programme and is representative of refugee households that applied to ESSN between the start of the programme and May 2017. The PAB sample size is 8 690 households. In population terms, the survey is representative of 1.6 million people. The survey was collected by the TRC call centre and was designed by the World Food Programme with inputs from the World Bank. The TRC call centre team was recruited and trained to conduct interviews in three languages: Arabic, Turkish, and English, to minimise the number of surveys that could not be completed due to a language barrier.

Issues and limitations: although the PAB data bring important added value, there are two limitations. First, the PAB survey does not sample the non-applicant population. Therefore, the vulnerability and targeting assessment is based on those who applied. Second, the phone-based modality poses an important limitation on the amount of information that can be collected from survey respondents. A face-to-face survey can collect information about every household member and can ask about living standards with a fine level of detail through hundreds of questions. However, it is also relatively expensive, which should be considered.

Source: World Bank Group and World Food Programme, 2024

Example: Egypt

Background: the Emergency Employment Investment Project (2014-2017) was a grant for EUR 67.6 million financed by the European Union (EU), administered by the World Bank and implemented by the Micro, Small and Medium Enterprise Development Agency (MSMEDA) (formerly the Social Fund for Development). Specifically, the 'Improving Youth Employability' component financed the piloting of youth employment projects aimed at facilitating young people's sustainable transition into salaried and self-employment.

Evaluation method: monitoring of outcomes (3 outcome indicators defined) and benchmarked against the target, randomised approach.

Research questions of the evaluation: what is the effect of training/employment support on the labour market outcomes of youth, such as employment status and income? How does the addition of individualised counselling impact the employment outcome of participants? Do the impacts of this intervention differ based on the gender of the participants? Does the intervention have impacts on non-labour market outcomes, such as female empowerment?

Data Collection: Surveys were conducted at various stages of the programme and post-programme period:

- Baseline Survey (10 min, paper-based) at registration before training, measuring employment, self-employment, loan repayment, education/training participation, and self-confidence.
- Exit Survey (10 min, paper-based) at graduation, assessing completion rates, job/internship placements, small business support, satisfaction, and job offers.
- Post-Participation Polling (6 months post-training) with data separated by governorate, tracking job/business retention, loan repayment, and re-enrolment in education.
- Post-Programme Discussions: (i) Focus groups (10–15 youth) and (ii) group polling (6 months post-graduation).
- Employer Questionnaire (5 min, web-based) at 6 months, evaluating employer satisfaction with graduates.
- 12-Month Focus Groups: Surveying youth who had started businesses at exit.

Limitations and Issues:

- Short Evaluation Period: End survey conducted 4–10 months post-programme, considered too brief for assessing long-term impact.
- Implementation Challenges: Modifications to fit the randomised evaluation reduced vocational industries, and stakeholders found business training too short, affecting operations. Malfunctioning equipment (e.g., sewing machines) hindered skills acquisition and business performance.
- Staffing and Budget Issues: High staff turnover (three project managers replaced) delayed data collection. Budget constraints led to hiring less qualified personnel.
- Delays and Dropouts: Some participants dropped out due to delays between random assignment and intervention. One NGO's outreach efforts were ineffective.
- Soft-Skills & Coaching Challenges: Training was too short and theoretical. Coaching faced obstacles as employers restricted sessions during work hours, and many women avoided evening sessions due to household constraints. Coach quality was inconsistent.

Source: Elsayed et al., 2018, Interview

Example: Morocco

Background: youth unemployment is high in Morocco, in particular among young graduates. Donor-funded project targeted at 12 000 youth: Job Search Training 'Finding a Job Is a Job' implemented by Education for Employment (EFE Maroc). The evaluation was carried out with ILO guidance (Taqeem initiative).

Evaluation approach: monitoring participants and a small tracer study / follow-up approach.

Data Collection: (i) Application Forms & Interviews – An online form with ~15 questions on socio-demographics, work status, education, and motivation determines eligibility, followed by a short interview. (ii) Attendance Lists – Tracked for participation. (iii) End-of-Training Questionnaire – A 20-question survey (10–12 min) assessing satisfaction, initial skill gains, and short-term job search plans. (iv) Follow-up Surveys – Conducted online at two and four months (20 questions, 10–12 min) covering job-search strategies, applications, interviews, and programme impact. (v) Focus Groups – Moderated by EFE Maroc's M&E manager, gathering 8–10 graduates for discussions (4–6 key questions) to complement survey data. Sessions, recorded by a note-taker, are held 2–3 times per year, lasting 60–120 min.

Outcome/result indicator: gaining knowledge and skills; job-search behaviour; finding a job.

Issues / challenges / limitations and approaches to overcome them

(i) Low response rates (24%). Planned improvement: send end of training survey only to part of participants and deploy techniques to get better response rates. Strategies for increasing response rates include offering an incentive, sending several reminders, personalising communication, and choosing alternative survey tools (e.g. phone calls instead of web-based surveys). To prevent 'survey fatigue' among beneficiaries, EFE decided to suspend its two-month follow-up survey.

(ii) Organising focus groups amongst FJJ graduates was difficult. Focus groups tend to be labour- and time-intensive for staff, as graduates are often unavailable due to other commitments such as employment. Graduates are also sometimes sceptical as to why they are being contacted, even though careful precautions are taken regarding the wording of the invitation message (email and telephone). As a result of these challenges, the EFE's first pilot focus group had to be cancelled because no participants

showed up, and a second one was conducted with only four of eight confirmed participants in attendance. In the future, EFE plans to overcome this issue by offering a larger incentive to youth.

Source: ILO 2015

Example: Mexico, State of Guanajuato

Background: Guanajuato is one of the most economically dynamic states in the Centre of Mexico, an industry-based economy, with the presence of multinational companies. At the same time, it is one of the states in Mexico with the highest crime rates. JuventudEsGTO is a state agency, resulting from the merger of two former government institutions: Educafin and Guanajoven⁸. The agency pursues innovative approaches and has established strategic alliances with public and private actors (including municipalities, universities and other training institutions, employers). The agency also has close contacts to 46 municipalities. The municipalities run local institutes for youth, which help to reach out to young people. The agency runs public policy programmes for young people. One project is an innovative internships scheme, called Evolucion GTO with a duration of six months. This programme is open to young people aged 18-30 in higher education (private or public) or who have already graduated and have less than one year work experience. Interns at companies receive a monthly support of EUR 430.

Method: monitoring of the internship programme Evolucion

- A tracer study is conducted during six months after participation: a survey questionnaire is sent to participants every month to collect information on the employment status, retention in the company that they did the internship in, formal employment in other companies, informal employment, salary, job title and assessment on matching education level and acquired skills with those skills requested in the job. The ILO provides methodological support for developing the survey.
- Get information on whether young people are eventually employed by the company for six months and not less.

Issues / limitations

Problems in implementing the monitoring approach for the Evolucion programme:

- The more vulnerable young people do not always have access to internet. They are the most difficult to follow up.
- Difficulty in following up on companies to ensure compliance with the 6-month internship rule (some only implement 3 months, and it is very difficult to understand the reasons).

Source: interview and Instituto Latinoamericano de Planificación Económica y Social (ILPES) de la Comisión Económica para América Latina y el Caribe (CEPAL), (2021).

4.3 Outcome and impact indicators

Evaluation studies should define several types of outcomes: first, the outcome that directly reflects the main objective (often the main outcome indicator), and then additional output indicators that allow contextualisation and a better understanding of the primary impact. In addition, evaluations may measure both direct and indirect impacts (impact indicators).

In fragile contexts, specific factors may have an influence on whether an ALMP works in the specific context. Measurement of relevant input and output variables – such as changes in attitudes, beliefs, prejudices, or power relationships – may be difficult, as these factors are not always easily quantifiable. Therefore, it may be necessary to conduct additional research to specify and define relevant variables and indicators.

Another issue to consider, for example, in many Sub-Saharan countries, is the insecurity and fluctuation in employment, income generation and life perspectives ('surviving from one day to the other'). In general, this scenario is common, as compared to more economically developed countries,

⁸ <https://juventudesgto.guanajuato.gob.mx/>

and affects programme design (social protection programmes and ALMPs) as well as outcome indicators for evaluations.

In general, the evaluations reviewed emphasise both ‘hard outcomes’, such as employment status following a specific programme or measure, and ‘soft outcomes’ related to motivation, self-perception, and similar factors. The significance of considering ‘soft outcomes’ has been extensively highlighted during the conducted interviews. Table 4 presents selected examples of indicators measured during the evaluation exercises.

Table 4: Examples of indicators used in net impact evaluation studies mapped in this study

	Type of intervention	Evaluation approach/technique	Outcome and impact indicators
Morocco	Training and internship programme (with a focus on young people)	Randomised Control Trial	<ul style="list-style-type: none"> - Labour market outcomes, including labour market statuses (employment, NEET, unemployment, inactive), reservation wage, and the number of months worked. - Education outcomes, including the enrolment in education, attainment level, aspiration level. - Self-perception of skills and self-efficacy. - Savings and financial behaviour (maintaining a savings account).
Morocco	Subsidised government employment programme, which supports the transition of youth from school to work through company placements (Idmaj)	Propensity score matching	Direct impact on employment as well as working conditions, including: labour market status (unemployed/employed) after having participated in a programme; level of salary; benefit or not of social cover; number of hours worked per week.
Egypt	Empowering young women (who can read and write) through an integrated approach to female economic and social empowerment to enhance the transition to work for young, marginalised women in rural upper Egypt. The programme provides business, vocational and life skills training in addition to actual support in starting a business or becoming employed.	Quasi-experimental design; difference-in-difference and propensity score matching (PSM) (treated women and women in control villages); estimating spill-over effects (to control villages).	<ul style="list-style-type: none"> - Labour market outcomes of young women looking, in particular, at income-generating activity, salaried employment, and self-employment. - Business knowledge index, on a scale of zero to one based on an unweighted index of six items capturing women's business knowledge. - Economic aspirations: estimated by women's economic goals for the future (women are asked about whether they have plans to set up/continue a project, and/or to secure salaried employment). - Social empowerment is estimated using two indexes: (1) Gender equality index: women were given different statements about the role of women and asked if they agreed with each statement; and the (2) Decision-making index (DMI): women are asked whether they usually have the final say in making different decisions within the family.
Bosnia and Herzegovina (FBiH)	Start-up incentive targeted at youth and women	Propensity Score Matching and descriptive statistics analysis	<ul style="list-style-type: none"> - Number of persons from the target group who kept the activity / small business until the end of the contract on co-financing self-employment. - Number of persons from the target group who remained employed until the end of the contract and are employed in the newly established activity. - Planned / contracted / spent funds under the programme.

	Type of intervention	Evaluation approach/technique	Outcome and impact indicators
			- Number of persons who kept the activity 6 or 12 months after the expiration of the co-financing period.
Uganda	Programme offering 3 types of interventions, including vocational training; vocational training combined with a light touch matching intervention that passes worker's details to local firms; and matching only.	6-year experiment, randomised experiment with 3 treatment groups and one control group.	Being called back by a company for an interview was the main result. Other outcome indicators were considered such as skills acquisition, earnings, and job-search behaviour (including motivation).

Source: Own elaboration based on Dyer et al. 2018, Chatri et al., 2021, Dyer et al. 2017, Meštrović, and Markuš (2021), Bandiera et al., 2023

4.4 Limitations for the evaluation and challenges in relation to fragile contexts

Evaluating ALMPs in fragile contexts involves specific considerations related to participant tracking, security risks, programme implementation, data limitations, and contextual factors. This section summarises these key aspects based on the collected data.

Tracking and participation attrition

Successful impact evaluations rely on tracking programme participants over time. However, attrition rates⁹ can be high. One way to address this issue is through close in-person follow-up, although it can be costly. Visiting participants in their home villages and collecting contact details of both the participants and their relatives/parents (who tend to be more stable in their location). Another strategy is to provide incentives for being tracked (interview). Providing incentives, such as mobile phone credits, has proven effective. For example, in Egypt, training and incentivising enumerators with mobile credit for participants resulted in an attrition rate of just 6% (Elsayed et al., 2018).

Security and accessibility constraints

Interviews and surveys carried out in person may not be implemented in regions with political conflicts and crimes, as this would be too dangerous or difficult to organise from a political point of view for the research teams (interviews). In such cases, data are not collected at all or are collected through local stakeholders or researchers on the ground, without, however, the possibility for the core research team to verify the quality of data collected.

Programme implementation issues

Implementation challenges frequently impact the evaluation process in fragile contexts. Issues such as lower-than-expected enrolment and participation rates, high dropout rates, delays in implementation and staff fluctuation, amongst other factors, have been cited in interviews and reviewed evaluations. For instance, training institutions located in unsafe areas may hinder participant attendance, especially among vulnerable groups such as women in rural areas.

For example, one of the implementation problems involves reaching out to the target group, such as, for example, women in rural areas in Egypt. Hence, the women were not randomly selected for participation in the programme. A main empirical problem with the analysis was the self-selection of

⁹ Share of people not responding to the survey.

participants arising from the voluntary nature of the programme participation. To account for the selective nature of the intervention, the evaluators combined a difference-indifference (DD) approach with propensity score matching (PSM). A combination of DD and PSM allowed circumvention of the self-selection problem by drawing on the assumption that, depending on observable characteristics of women, unobservable characteristics that might affect self-selection into the programme and subsequent changes in outcomes are similar between treated and control groups (Dyer et al. 2017).

Example: Morocco

Background: The donor-funded programme '100 Hours to Success' targeted 20 000 young people. It provided youth with 100 hours of training in life skills, entrepreneurship, and financial education. The participants of the training were required to open a bank account at Al Barid Bank. After the training, between 5 and 10 percent of youth received internship placements. The length of the training programme was three months. The evaluation was coordinated by the international implementer of the programme Mennonite Economic Development Associates (MEDA) Morocco. Guidance was provided by the ILO Taqeeem initiative.

Method: The research team chose a randomised controlled trial (RCT) design for the evaluation. The main research questions included: (i) Do training participants demonstrate greater financial knowledge and heightened awareness of banking institutions and their services? (ii) To what extent has training changed beneficiaries' perception of their own capacities in a broad array of life skills, such as confidence, teamwork and problem solving? (iii) Has the intervention influenced educational choices? (iv) Has it placed participants in a better position to enter the labour force? (v) Has it increased their chances of securing employment or starting new businesses as youth entrepreneurs?

Data collection: the study involved a baseline survey (conducted in October and November 2012), and a follow-up survey conducted 12–14 months after the end of the training. *Constructing control group:* Over the three-week span of the baseline survey, 1 817 youth were surveyed. A random selection initially placed 600 in the participant group (equalling the number of available spaces in the training programme during the time under study) and 1 217 in the comparison group. The bigger size of the comparison group was decided to both increase the power of the calculations (given the small population) and help remediate some of the enrolment take-up issues that MEDA foresaw at the time (furnishing back-up candidates for the participant group). The summary of statistics shows that randomisation was successful in achieving balance across participant and comparison groups. The surveys were conducted in 13 youth centres where MEDA Morocco's trainings were carried out.

Limitations and challenges encountered

Attrition rate: the research team could only locate and interview 871 youth (427 from the treatment and 444 from the control group) for the follow-up survey, implying an attrition rate of just over 50 per cent. The remaining sample showed balanced characteristics at baseline, and attrition rates did not differ between the treatment and the control group overall. The results were robust to various sensitivity checks, including inverse probability weighting.

Another difficulty encountered was linked to **programme implementation**. Therefore, a second random selection of 300 youth was taken from the comparison group and brought into the participant group, bringing both into alignment at about 900 individuals each. This means that for the purpose of the evaluation, programme implementation had been reviewed. The second wave of training was completed in August 2013, with a combined total of 505 youth having participated in the training overall. While this random selection should help reduce bias in results, certain heterogeneities in the results could not be observed because of decreased statistical power. At the same time, the implementing body MEDA was tracking down dropouts to ensure they could be available for the follow-up survey where 'intention to treat' analysis could be applied.

Source: Dyer et al., 2017

Deadweight and substitution effects

A critical challenge in assessing the net impact of ALMPs is determining whether observed outcomes are solely attributed to the programme. Deadweight and substitution effects can limit net impacts.

For example, an ALMP evaluation in Bosnia and Herzegovina encountered potential deadweight effects, where resources were provided to individuals who would likely have achieved the

programme's goals independently. In the case of self-employment measures assessed here, an example of a deadweight effect is a situation when a person who already has an operative business 'under the radar' waits for the opportunity to legalise it by becoming a participant in a self-employment measure. The deadweight effect might not be entirely avoidable, but it should be considered in the process of designing the eligibility criteria for the measure (Meštrović and Markuš, 2021).

Data limitations and the need for comprehensive data collection

Evaluation studies frequently encounter constraints related to the type and extent of data gathered (see above). This is particularly problematic in fragile contexts where reliable administrative data is often unavailable, and the ability to gather primary data may be restricted. Thus, a robust and comprehensive data collection framework that incorporates both quantitative and qualitative methods is essential for effective impact evaluation. It is important to note that this requirement is not exclusive to fragile contexts but is broadly applicable across all evaluations.

Reliance on administrative data and monitoring capacity

Evaluations often depend on administrative data, which requires a mature and reliable monitoring system (interview). In fragile contexts, institutions may lack the infrastructure or consistency in data collection, which needs additional resources and coordination. Another common issue is the possibility of accessing administrative data. This requires establishing a relationship of trust between the institution and the evaluator (interviews).

Adapting established evaluation models to fragile contexts

Conventional theories and causal models used in impact evaluations may not always hold in fragile socio-economic environments. Evaluators may need to invest additional effort in adapting theories of change to better align with local conditions. However, resource and time constraints often limit the scope for such adaptation, impacting the reliability of findings.

Example: Ghana: training and financial support to micro-entrepreneurs

Background: theories on poverty traps suggest that credit and managerial capital constraints inhibit investment and thus profits. The evaluated programme provided large grants of capital and management consulting, separately and combined, to urban microenterprises of tailors in Ghana.

Method: This evaluation used a randomised trial. The study was designed based on the hypothesis that large infusions of financial and managerial capital could be transformative.

Limitations and challenges encountered

The results of the evaluation showed that the micro-entrepreneurs invest the cash grants in their businesses and adopt the management practices advocated by the businesses consultants. But there was no evidence that these changes were associated with increases in profits after one year. One explanation is that these microenterprises are unsure about the profitability of expanding their business or of using innovative managerial practices and face capital and managerial constraints on investment. The authors of the study argue that a more detailed analysis would have been needed to understand and explain the findings. In particular, for further research they suggest a more detailed measurement of attitudes, including measuring confidence and willingness to take further business risks, preferences and expectations, both from surveys and from field experiments, to shed light on mechanisms. Heterogeneity of firms is also perceived as biasing results. An analysis by gender (including on risk-taking mechanisms) would have enabled the impacts to be differentiated and would shed light on differences in the constraints and barriers. They also suggested looking in more detail into the dynamics and determinants of learning (including differences in programme design).

Source: Karlan et al. 2014

5. INSIGHTS FROM EVALUATIONS OF ALMPs

Designing effective ALMPs, including those in fragile settings, requires a nuanced approach grounded in evidence and adaptive strategies. Building on a review of evaluation outcomes collected, in this section key insights are provided into what works. By examining successes and failures across diverse contexts, this review may guide policymakers and practitioners in creating effective and context-sensitive ALMPs.

It became clear that developing a standard for measuring programme impact that could be compared across studies would be unlikely to succeed, due to the wide variation in methodological approaches in the reviewed reports. While we were unable to extract standardised effect sizes from our sample of studies, we were able to extract an estimated programme effect and the associated employment rate of the comparison group.

The Table 5 below provides a summary of the programme types and geographical coverage based on the availability of (impact) evaluation reports reviewed by our team. For ease of discussion, we have categorised the analysed countries by region.

Table 5: Programme types and geographical coverage

Regions	Training	Employment incentives	Sheltered and supported employment and rehabilitation	Direct job creation	Start-up incentives
Western Balkans and Türkiye	Albania, Kosovo, Montenegro, North Macedonia, Türkiye	Albania, Kosovo, North Macedonia, Ukraine		Türkiye	Bosnia & Herzegovina, Kosovo
Eastern Partnership	Moldova, Armenia			Georgia	
Central Asia	Kazakhstan, Kyrgyzstan	Kazakhstan		Kazakhstan	Kazakhstan
Southern and Eastern Mediterranean	Morocco, Egypt, Lebanon	Morocco			Morocco, Egypt, Tunisia
Sub Saharan Africa	Benin				Ghana, Benin, Kenya
Others	Mexico, India	India			Mexico

Source: Authors

In several countries, we have identified two types of mixed interventions. The first type ensures that all participants benefit from each component of the intervention. The second type either offers different actions for distinct target groups or acts as a budget reserve for various types of actions, with individual participants benefiting from only one component at a time, although they may progress to another component subsequently.

The Table 6 below presents a summary of the programme types and participant characteristics reviewed in this study.

Table 6: Programme types and participant characteristics

Regions	Training	Employment incentives	Sheltered and supported employment and rehabilitation	Direct job creation	Start-up incentives
Western Balkans and Türkiye	<ul style="list-style-type: none"> • young people under the age of 29; • long-term unemployed jobseekers; • jobseekers over the age of 45; • unqualified unemployed jobseekers; unemployed jobseekers who have been repatriated for no more than two years; • jobseekers that receive economic aid; • jobseekers that receive unemployment benefits for more than 3 months; • persons with disabilities; • victims of trafficking, gender-based violence and domestic violence; • jobseekers from the Roma communities; • persons who have completed the vocational training course or on-the-job training programme and are still unemployed 6 months after completion. 	<ul style="list-style-type: none"> • jobseekers in difficulty; • jobseekers from the most vulnerable groups (youth, minorities, women); • repatriated women and men. 	<ul style="list-style-type: none"> • persons with disabilities. 		<ul style="list-style-type: none"> • jobseekers registered with PES, with a focus on youth; • persons with disabilities; • ex-convicts.
Eastern Partnership	<ul style="list-style-type: none"> • unemployed women; • fresh graduates; • young people with disabilities; • war veterans. 	<ul style="list-style-type: none"> • registered unemployed. 		<ul style="list-style-type: none"> • recipients of social assistance. 	
Central Asia	<ul style="list-style-type: none"> • registered unemployed; • refugees; • long-term unemployed; • unemployed women; • newly graduated; • young people with disabilities; • war veterans. 	<ul style="list-style-type: none"> • registered unemployed. 		<ul style="list-style-type: none"> • registered unemployed. 	<ul style="list-style-type: none"> • registered unemployed.
Southern and Eastern Mediterranean	<ul style="list-style-type: none"> • unemployed youth; • disadvantaged youth; • youth aged 18-29 (in some cases up to 35), unemployed or inactive (with a focus on the poor, young women, low levels of education, disability, no prior work in the formal sector, etc.); • immigrants; • women. 	<ul style="list-style-type: none"> • Fresh graduates; • long-term unemployed graduates. 			<ul style="list-style-type: none"> • youth aged 18-29 (in some cases up to 35), unemployed or inactive (with a focus on the poor, young women,

Regions	Training	Employment incentives	Sheltered and supported employment and rehabilitation	Direct job creation	Start-up incentives
					low levels of education, disability, no prior work in the formal sector, etc.); •women.
Sub-Saharan Africa	<ul style="list-style-type: none"> •low skilled youth •informal young workers; •women working in the informal economy; •inactive women 				<ul style="list-style-type: none"> •vulnerable youth; •low skilled youth; •women; •micro-entrepreneurs (tailors); •poor women living in rural areas.
Others	<ul style="list-style-type: none"> •youth aged 18-30. 	<ul style="list-style-type: none"> •poor workers living in rural areas. 			<ul style="list-style-type: none"> •unemployed aged 15+

Table source: Authors

Each ALMP category will be examined in this section, and its effectiveness will be evaluated through a qualitative analysis of the literature in terms of what works, for whom, and why. It should be noted that many ALMPs now include several (active) measures, sometimes combined with job-search assistance services, so classifying programmes can be problematic.

Training

Training is the most common active labour market measure, being accepted as one of the most effective interventions that may enhance the labour market prospects of jobseekers, including within fragile settings. In practice, training can take place in a wide range of arrangements, which are classified by the EC methodology into four categories, based on the location where the training mostly takes place (institutional, workplace and alternate training). A fourth category is the special support for apprenticeship schemes, either by providing incentives to employers to recruit apprentices, or by providing training allowances for particular disadvantaged groups.

One of the most popular types of action from this category is on-the-job training programmes, where most of the countries report good results in increasing the retention rate of participants by about 63% in Kazakhstan or 69% in Armenia, to name a few.

Training for employment programme was found effective in Egypt

Researchers conducted a randomised evaluation to assess the impact of a soft skills training programme and matching with private firms providing technical on-the-job vocational training. The programme was offered by two NGOs. The programme duration included a three-day business training and industry-specific vocational training lasting 6-17 days (NGO1), and a four-day soft skills training and 1-6 weeks of on-the-job training (NGO2). The impact on labour supply (probability of working or hours worked) and earnings was:

- Share of those reporting that they are currently working increased by 45 percentage points (a 300% increase relative to the average of 15% employment in the comparison group), (NGO1).
- Share of those currently working increased by 12 percentage points, (NGO2).
- Monthly personal income increased by 66% (EGP 91; USD 5.82) (NGO1). The share of individuals who report having no income decreased by 17 percentage points, (NGO1).
- Monthly personal income increased by 34.6% (EGP 99; USD 6.33), (NGO2).

Source: Interviews with representatives of EIL & J-PAL MENA, 2024

The successful design and implementation of training within active labour market programmes rely on several key elements, as mapped in this review. In Kyrgyzstan, short-term training programmes proved highly effective, with approximately 70% of young participants securing employment (interview). However, limited outreach made it difficult to engage individuals in remote areas, highlighting the need for more robust dissemination efforts.

For refugee and returnee support, a structured system for assessing participants' skills is essential, as it enables targeted and impactful interventions that address specific needs (interview). In general, the collected evidence shows that matching of trainees to programmes that best fit their skills and their career aspirations is crucial. Building on these insights, a comprehensive curriculum that combines both hard and soft skills, coupled with recognised certification, enhances employability and aligns training outcomes with labour market demands. Blending classroom-based instruction with on-the-job training further enriches the learning experience, fostering both theoretical knowledge and practical expertise. Furthermore, offering financial assistance can help mitigate financial barriers that may hinder individuals from participating in and succeeding in training programmes.

Lastly, the recruitment and careful screening of high-quality trainers is crucial, as skilled instructors significantly influence programme effectiveness. Together, these elements create a strong framework for impactful training initiatives, especially when further combined with outreach strategies to reach diverse participant groups.

Employment incentives

Employment incentives refer to demand-side active labour market measures. These incentives can include wage subsidies or targeted reductions in social security contributions for employers. Additionally, employment incentives aim to favour the conversion of temporary contracts into open-ended ones. Typically, they focus on reactivating long-term unemployed individuals or supporting groups at risk of labour-market exclusion, such as young people, people with disabilities, women, migrants and older workers. Furthermore, we can distinguish between two types of employment incentives:

- *Recruitment Incentives*: these facilitate the hiring of unemployed individuals by providing work experience and improving employability.
- *Employment Maintenance Incentives*: these assist in maintaining employment for individuals at risk of job loss due to restructuring or economic pressures.

The results of consulted reports suggest that employment incentives can effectively reduce poverty in developing countries and may also improve economic efficiency. As observed from the consulted counterfactual impact evaluations, the probability of being employed after the programme increased between 7.5% (Morocco) and 33.8% (Albania).

Internship programmes were effective in several countries for women and young people under 24 years old, but due care and consideration should be paid to the negative effects, such as lowering the wages of beneficiaries, compared to the wages of the control group, as being reported in Morocco. Moreover, it was observed that 24 months of internship is an extended period that can produce counterproductive effects (ILO, 2015).

It is worth noting that several other negative effects, such as displacement and deadweight, were observed in relation to employment incentives. The risk is that by subsidising new entrants into jobs, others may lose work (displacement). A deadweight loss then occurs when employers would have created these jobs anyway. Employment incentives then become a form of artificially cheap employment for employers who abuse the measure (Sienkiewicz, 2021).

Employment subsidies are also considered as one of the most expensive measures. The cost-effectiveness analysis of the wage subsidy programme needs to be accompanied by cost-benefit analysis in order to assess its beneficial effects relative to the costs (Nikoloski, 2021).

North Macedonia

The evaluation of outcomes from the wage subsidy programme shows an improvement in 2019 relative to 2018. In 2018, wage subsidies were associated with increasing unemployment and an increased intention to emigrate. This can be attributed to possible job closures after the expiration of the period for receiving wage subsidies. However, in 2019, wage subsidies had a reduced impact on unemployment, a positive effect on salaries, and a decreased intention to emigrate. Although the incremental cost-effectiveness ratio indicates improved efficiency in 2019 compared to 2018, this ALMM is still considered as one of the most expensive measures. In this context, it was recommended to redesign this measure by improving its targeting and conditions for retaining the subsidised jobs.

Source: Nikoloski, 2021

Furthermore, the reports consulted identified several limitations that must be addressed for women to benefit from this measure. These include issues related to childcare availability, the role of families in specific cultural contexts, and (the lack of) self-confidence.

Finally, in donor-driven projects, ENABEL's experience in implementing the First Employment Facility projects in Morocco and Palestine, aimed at specific groups of graduates and migrants, has highlighted the significance of cooperation and execution through strong local partners, such as (public or private) employment services or similar entities. This approach is essential for the success of these initiatives.

Sheltered and supported employment and rehabilitation

This type of action covers measures that aim to promote the labour market integration of persons with reduced working capacity through sheltered or supported employment or through rehabilitation. Sheltered employment refers to employment in an enterprise operating in a commercial market, with or without public support, and established specifically for the employment of people with disabilities or other working limitations, but which may also employ able-bodied people in a limited proportion. Supported employment refers to employment in a regular working environment where people with disabilities or other working limitations are enabled through public support (financial or otherwise) to work alongside able-bodied employees (European Commission, 2018).

This measure is the least frequently reported active measure among the reviewed reports. This measure has significant potential in supporting individuals with disabilities and other work limitations.

For example, in Kosovo, the piloted grant scheme ‘Enhancing Employment of Vulnerable Groups’ demonstrated positive outcomes; however, the initiative was not extended further (interview).

Direct job creation

For years, this type of action, often referred to as public works, was considered the last option for supporting jobseekers facing difficulties in finding employment. It primarily targeted low-skilled, older workers, and certain ethnic minorities. A specific example for implementation of this type of measure is the Cash-for-Work programme implemented by GIZ (multi-donor trust fund) in Syria’s four neighbouring countries (Türkiye, Jordan, Lebanon and Iraq), which had a duration of mostly six weeks, with some cases extended up to eight months. However, in recent years, there has been a shift from traditional infrastructure projects (such as park and riverbed cleaning) to what is now termed ‘digital public works’.

The current review found a limited number of relevant impact evaluation reports, which does not necessarily imply that public works programmes are not implemented in the analysed countries. For example, in Kazakhstan public works schemes are expanded and reduced to reflect cyclical labour market fluctuations and numbers of long-term unemployed persons, but the country does not conduct counterfactual impact evaluations of ALMPs. The small number of examples might also be the result of two dominating opinions. On the one hand, the PES staff do not perceive it as being within their mandate to create jobs, therefore they display some resistance to using this measure. On the other hand, the literature broadly agrees that public works are not an effective policy tool for unemployment in the long term, or at least without a mixed intervention that would include vocational training or job-search assistance.

Major risks related to the practical implementation of the public works scheme

In Georgia, three major risks related to the practical implementation of the public works (PW) scheme were identified as follows: being trapped in low-skilled jobs, trapped in low-paid jobs, and job substitution. To avoid being trapped in low-skilled jobs, it is recommended to gradually introduce a two-tier system, allowing for higher skilled PW jobs, to allow for a training component for PW participants, and to place PW into an overall system of PW services – building links to other measures and a conceptual approach. Increasing the PW allowance could help to mitigate the risk of being trapped in low-paid jobs. Finally, the risk of job substitution seems to be the most serious at the onset of the PW programme. The risk could be decreased by changing the rules in a way that would prevent easy switching from regular jobs to PW vacancies.

Source: Lubyova, Martina and Diakonidze, Ana (2022)

Start-up incentives

Interventions from this category are aimed at promoting entrepreneurship by encouraging the unemployed and other target groups to start their own business or to become self-employed. Assistance may take the form of direct cash benefits or indirect support including loans, provision of facilities (business incubators or shared workspaces), business advice, training etc. They are targeted at a wide group of workers. Frequently targeting broad groups of workers, these interventions are sometimes integrated with poverty alleviation programmes, particularly in fragile settings where economic opportunities are limited. A notable example is the Self-employment programme—Start-up in Bosnia and Herzegovina, which demonstrated a 99% success rate in 2018, with 397 out of 400 participants gaining employment through this initiative. Most participants remained employed 12 months after the programme concluded, highlighting the programme's effectiveness and sustainability across all cantons (Meštrović and Markuš, 2021). Nevertheless, it should also be noted that start-up initiatives often entail a certain level of risk, including high rates of business failure and related challenges.

Tunisia

An experiment conducted in Tunisia demonstrates the complexity of employment barriers for women. The programme aim was to address women's financial and human capital. Treatment group 1: 1 000 women were offered an unconditional cash grant of TND 634 (USD 768 in PPP terms, USD 240 in nominal terms). This amount is relatively large: about four times the median monthly income of respondents with an income generating activity at baseline. Before getting the cash grants, the women had to take part in a one-day training course, which covered the basic concepts of money management and investment. The training was co-produced and co-organised by Tunisia's Centre of Arab Women for Training and Research (CAWTAR) and the World Bank. It aimed at encouraging women to invest their money productively in physical capital (e.g. starting an income-generating activity) or human capital (e.g. paying for vocational training of their choice). The training covered three main modules: i) financial planning and budgeting, ii) savings, and iii) debt management. Each module included a series of videos, exercises and guided discussions.

Treatment group 2: joint training with male partners. Half of the 1 000 women receiving cash grants (i.e. 500 women) were invited to bring their male partners to the financial training described above. Partners were asked to sit next to each other and do the exercises together. The take-up rate of the joint training programme was also high: 444 partners attended the training from a total of 486 invited partners (and 502 women were randomly selected for this treatment group). The take-up rate of this component was therefore 88.4%.

The results show the positive impact of the cash grant and training programme on the women's likelihood of having an income-generating activity and on their income; however, this was the case only if they participated without their partners. When partners were involved, women were less likely to have income-generating activities and they earned less. The positive impacts of income-generating activities could also be seen on other household members. The authors concluded that traditions interfere with programmes aimed at stimulating the employment of women.

Source: Jules Gazeaud et al, 2022

Kosovo

In 2023, the final evaluation of the ALMP2 programme was conducted and the programme report includes a special section on impact evaluation. This was developed using standard UNDP methodology, which includes a survey of the beneficiaries of the measures. That programme implemented three ALMPs, namely on-the-job training, a wage subsidy scheme and support for self-employment. The report concluded that the most effective measure, especially for returnees, was self-employment. This measure included counselling, coaching, mentoring, training and grants for equipment. It was initially piloted for a group of beneficiaries and the design was jointly reviewed together with the PES through a working group, thus building long-term sustainability. Additionally, a network among the beneficiaries was also established.

Source: Interview

In addition to various measures and services, the following observations were made based on the collected evidence and feedback. Firstly, to implement effective measures, it is essential to adopt a holistic approach that considers the diverse needs and circumstances of clients, ensuring that the support provided is both relevant and impactful. Moreover, the selection of target groups is critical, as it allows interventions to be directed toward individuals who stand to benefit the most, thereby maximising the programme's effectiveness.

Success also depends on strong institutional support and the active involvement of capable partner organisations, which reinforce the programme's foundation and implementation capacity. Additionally, in certain cases, involving the private sector from the beginning and ensuring their commitment is crucial for improving both the effectiveness and sustainability of initiatives. This approach aligns the projects with labour market needs, particularly when considering long-term impacts.

6. CONCLUSIONS

This study underscores the essential role of impact assessments in shaping ALMPs. Such evaluations provide critical insights into what strategies are effective, for whom, and under what circumstances, thereby enabling policymakers and implementing institutions to refine the design, targeting, and execution of different measures based on solid evidence.

By examining the existing evidence, several insights can be gained regarding the implementation of impact evaluations in ETF partner countries, as well as the design of effective ALMPs.

Evaluation objectives and organisation of evaluations

ALMP impact evaluations are aimed at assessing the effectiveness, efficiency and impact of programmes. They are also meant to improve the design and implementation of programmes as well as to justify the use of public resources. In the context of fragile and post-conflict settings, impact evaluation studies can be incredibly useful not only to understand what works, but also to investigate details of what works for whom and why. It thus helps policymakers as well as institutions that implement ALMPs to improve the design of measures, targeting and the implementation mechanisms on the basis of evidence. To meet the objectives, evaluations often need to be complex to mirror the complex realities and deepen the understanding of institutional, structural and person-related barriers. Conducting such evaluations demands a high level of expertise.

In a few countries included in this study, institutions such as PES and/or related ministries have developed capacities to commission evaluation studies (and have thus developed an evaluation culture), or to implement some improved monitoring elements, including conducting surveys among participants or using other tracing methods. However, in most cases, particularly when it comes to fragile contexts, international organisations have either been commissioned to organise the evaluation or international organisations have proceeded to conduct the evaluation of the programmes that they funded.

Contracting evaluations requires having knowledge of the applied methodologies and methodological issues that may arise when conducting the evaluation. The staff in PES and/or related ministries may not have the required specific skills. International organisations may have their own procedures and methodological preferences. For PES, related ministries, and for international donors to commission evaluations, there must be sufficient specialised research capacity in the country, otherwise they will have to contract international experts. In practice, international experts leading evaluations need to cooperate with local experts and research institutes, for example, for data collection.

One caveat of the leading role of international organisations (such as the World Bank, ILO, UNDP, and others) is that, in practice, PES and related ministries do not have many opportunities to develop their own technical capacities in designing and out-sourcing evaluations. On the positive side, international organisations have the technical capacity and understanding of complex net impact evaluations to carry them out themselves, or to contract all or part of it out. However, in cases where international organisations evaluate their own large-scale programmes, the objective is often to inform the policy design and programme implementation of their own organisation and not primarily the policy-making of the country.

Methodology, research design and data collection tools

The chosen methodology depends on a number of factors: available resources, the way the programme is implemented (for example, if it is implemented in some regions and not in others), the nature of the programme (including the size of the programme), and the objective of the evaluation.

A key issue is available resources. This will determine which methods can be chosen and how well the method can be implemented (for example, carrying out additional research and surveys to grasp with

more granularity the main issues that may affect the impact). Again, funding for evaluation is often provided through international organisations.

When funded with their own and fewer resources, institutions tend to opt for less complex evaluations and improved monitoring systems (although this is not a general rule). In some cases, methodological support was received from international organisations. Capacity building in the long run is key. Experience shows that trust building between the international organisation, experts, and the policymakers and implementer in the country is beneficial for all involved parties.

Research to detect counterfactual evaluation reports, including those applied in fragile and post-conflict contexts, shows that only a few evaluations have been carried out overall, covering mainly large-scale programmes. Mostly counterfactual impact evaluation methods were applied, including experiments and quasi-experimental approaches. Most studies have encompassed and devoted many resources to constructing the control groups and correcting biases. Conducting counterfactual impact evaluations usually requires a large sample size to be statistically relevant. Some of the studies have combined qualitative and quantitative approaches, which has proven most valuable to understanding the 'why' and the 'for whom' questions better.

The examples show that defining the impact indicators and a system of key outcome indicators and relevant variables requires good data sources, including the existence of large surveys such as household surveys for the time period considered in the evaluation, mature monitoring systems and administrative data, and skilled staff to conduct surveys on the ground. Another key decision to make is the time span of the evaluation. Ideally, the analysis of short-term, medium-term and long-term (> 1 year) impacts are desirable, however, there is a conflict between the timeliness of an evaluation, the time span covered and the comprehensiveness of the evaluation approach.

Use of evaluations

Two common ways of reporting evaluations are stakeholder workshops and written reports. Depending on the evaluation culture in the country, as well as the culture of transparency, evaluation reports may be published or not. Even if not published, they may be used by policymakers and implementers, if circulated internally. Building trust between the institution organising the evaluation and policymakers or implementing institutions increases the probability of results being used. A wider dissemination of results will quite likely increase their usage. The preliminary findings and recommendations are usually communicated to the local partners and stakeholders during a workshop at the end of the field mission. It is difficult to establish how much evaluation reports are effectively used in the policy design and implementation cycle.

ALMPs and Key Success Factors

- Training is the most common active labour market measure, being largely accepted as one of the most effective interventions that may enhance the labour market prospects of jobseekers.
- On-the-job training is most effective when focused on specific target groups, such as youth and women.
- One of the key elements for supporting refugees and returnees is the development of a proper system for assessing their skills, which will enable a more effective intervention.
- Employment incentives can effectively reduce poverty in developing countries, and may also improve economic efficiency.
- Internship programmes were effective for women and young people under 24 years old.
- Sheltered and supported employment and rehabilitation is the least common active measure and more efforts are required to test and learn from evaluations of these kinds of initiatives.

- Public works are effective for low-skilled, older workers and certain ethnic minorities, especially when combined with training and job-search assistance. It is, however, not effective in creating employment in the long-term and should be seen as a social protection instrument.
- Support for self-employment has been found to be an effective measure for specific target groups, for example returning migrants.
- Cash transfers for entrepreneurship and self-employment work best when combined with training, mentoring, coaching and other support services (such as shared workspaces, etc.).

Some key recommendations for the implementation of (impact) assessment of ALMPs

- Strengthen the capacities within the Public Employment Services (PES) and related ministries to organise impact evaluations, real-time evaluations and output evaluations. They should be able to clearly identify output indicators and possess the knowledge to select appropriate methodologies. These skills will enable them to use evaluation results to improve the design and delivery of PES services and Active Labour Market Policies (ALMPs).
- Engaging research institutes and consultancies, preferably with contributions from both national and international consultants, should be considered when there is a lack of capacity within the PES and associated agencies.
- Explore the use of administrative data for better monitoring and evaluation, including to enhance the quality of administrative data for net impact assessments.
- Conduct additional interviews and surveys to gather better context and other qualitative aspects of ALMP implementation and measurement of effects. This includes analysing implementation conditions through stakeholder interviews and focus group discussions.
- Track beneficiaries and work on reducing attrition rates by setting incentives and exploring innovative methods to update participants' contact details over time.
- Establish trust among key partners involved in ALMP evaluations, including public institutions, researchers, and external organisations that support impact assessments. This will facilitate smoother implementation of impact assessments and their application in policy.

European and international institutions as well as global and regional research communities should continue their efforts to consolidate capacities for policy evaluation in developing and transition countries, foster innovation in monitoring and evaluation methods, and secure faster transfer of policy impact results into design of new support programmes. Fragile and (post)conflict contexts become the new norm for ALMP implementation throughout the world therefore the ETF recommends a more decisive approach to finetuning and renew of research methods and investments in sustainable impact evaluation frameworks.

GLOSSARY

Coherence	The compatibility of the intervention with other interventions in a country, sector or institution. Note: the extent to which other interventions (particularly policies) support or undermine the intervention, and vice versa, including internal and external coherence. Internal coherence refers to synergies and interlinkages between the intervention and other interventions carried out by the same institution/government, as well as the consistency of the intervention with the relevant international norms and standards to which that institution/government adheres. External coherence refers to the consistency of the intervention with other actors' interventions in the same context. This includes complementarity, harmonisation and coordination with others, and the extent to which the intervention is adding value while avoiding duplication of effort.
Counterfactual	The situation or condition that hypothetically may prevail for individuals, organisations, or groups where there was no intervention (the status quo). <i>Note: this is used for counterfactual evaluation approaches. It can be estimated by creating a control group, a comparison group or a hypothetical counterfactual.</i>
Effectiveness	The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups. <i>Note: the analysis of effectiveness involves taking account of the relative importance of the objectives or results.</i>
Efficiency	The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way. <i>Note: 'economic' is the conversion of inputs (funds, expertise, natural resources, time, etc.) into outputs, outcomes and impacts in the most cost-effective way possible, as compared to feasible alternatives in the context. 'Timely' delivery is within the intended timeframe, or a timeframe reasonably adjusted to the demands of the evolving context. This may include assessing operational efficiency (how well the intervention was managed).</i>
Impact evaluation	An evaluation that assesses the degree to which the intervention meets its higher-level goals and identifies the causal effects of the intervention. Impact evaluations may use experimental, quasi-experimental and non-experimental approaches. <i>Note: the term is also sometimes used to refer only to evaluations that use explicit counterfactual analysis to determine the effects (including outputs and outcomes) caused by an intervention.</i>
Monitoring	A continuing process that involves the systematic collection or collation of data (on specified indicators or other types of information). Provides the management and other stakeholders of an intervention with indications of the extent of implementation progress, achievement of intended results, occurrence of unintended results, use of allocated funds and other important intervention and context-related information.
Outcome	The short-term and medium-term effects of an intervention's outputs. <i>Note: outcomes are often changes in the institutional and behavioural capacities for development conditions that occur between the completion of outputs and the achievement of impacts.</i>
Process evaluation	An evaluation of the internal dynamics of implementing organisations, their policy instruments, their service delivery mechanisms, their management practices, and the links between these.
Randomised Control	A type of evaluation that randomly assigns access to the intervention, to control influencing variables and limit bias, generating internally valid estimates of results.
Relevance	The extent to which the intervention objectives and design respond to beneficiaries, global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change. <i>Note: 'respond to' means that the objectives and design of the intervention are sensitive to the economic, environmental, equity, social, political economy, and capacity conditions in which it takes place. 'Partner/institution' includes government</i>

	<i>(national, regional, local), civil society organisations, private entities and international bodies involved in funding, implementing, and/or overseeing the intervention. Relevance assessment involves looking at differences and trade-offs between different priorities or needs. It requires analysing any changes in the context to assess the extent to which the intervention can be (or has been) adapted to remain relevant.</i>
Results	The outputs, outcomes, or impacts (intended or unintended, positive, or negative) of an intervention.
Sustainability	The extent to which the net benefits of the intervention continue or are likely to continue. <i>Note: this includes an examination of the financial, economic, social, environmental, and institutional capacities of the systems needed to sustain net benefits over time. Involves analyses of resilience, risks and potential trade-offs. Depending on the timing of the evaluation, this may involve analysing the actual flow of net benefits or estimating the likelihood of net benefits continuing over the medium and long term.</i>
Theory of change	The way the intervention is expected to achieve or achieves change. It represents how people understand change to occur in a given context, including explicit (or implicit) assumptions about the causal links between inputs, activities, and results. Often also includes evidence and risks for these elements of the results chain.
Theory-based evaluation	A theory-based evaluation approach is a structured way of understanding a given programme or policy, which involves formulating hypotheses about how and why it works, testing them, and making decisions about the programme or policy based on the results. This approach can help to identify what in a programme or policy is helping and what is not working, and suggest how to improve it, so that meaningful changes can be made. The theory-based evaluation approach involves understanding the programme or policy's purpose and the context of the intervention in order to identify the key components and any possible areas of improvement. This allows for a more comprehensive understanding of the programme or policy and enables stakeholders to better assess its success. Ultimately, this approach will help those creating and administering programmes and policies to effectively evaluate their research and refine their strategies.
Triangulation	<p>The use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an assessment.</p> <p><i>Note: it seeks to overcome the bias that comes from single informants, single methods, single observers, or single theory studies by combining multiple data sources, methods, analyses, or theories, monitoring and evaluation.</i></p>

Source: Mainly based on OECD 2023, for theory-based evaluation see <https://www.evalcommunity.com/career-center/theory-based-evaluation-approach/>

ACRONYMS

ADB	Asian Development Bank
ALMPs	Active Labour Market Policies/Programmes
ALMMs	Active Labour Market Measures
CPIA	Country Policy and Institutional Assessment
DD	Difference-in-Difference
EIL	Egypt Impact Lab
ESSN	Emergency Social Safety Net
ETF	European Training Foundation
FCAS	Fragile and Conflict-Affected States
GIZ	German Development Cooperation
IDA	International Development Association
ILO	International Labour Organization
J-PAL MENA	Abdul Latif Jameel Poverty Action Lab Middle East and North Africa
LATEs	Local Average Treatment Effects
MEDA	Mennonite Economic Development Associates
NGO	Non-Governmental Organisation
OECD	The Organisation for Economic Co-operation and Development
PAB	Pre-Assistance Baseline Survey
PES	Public Employment Services
PSM	Propensity Score Matching
RCT	Randomised Control Trial

RCC	Regional Cooperation Council
SEMED	The Southern and Eastern Mediterranean Region
TRC	Turkish Red Crescent
UN	United Nations
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Fund
WFP	World Food Programme

REFERENCES

- Asian Development Bank (2013), “Working Differently in Fragile and Conflict-Affected Situations”.
- Aşık, Güneş and Laurent Bossavie, Jochen Kluge, Efsan Nas Özen, Metin Nebiler, Ana Maria Oviedo (2022), “The Effects of Subsidizing Social Security Contributions. Job creation or Informality Reduction?”, World Bank, Policy Research Working Paper 9904.
- Assaad, Ragui and Mona Amer (2022), “Bab Amal Program Impact Evaluation and Initial Findings: Graduating the Ultra-Poor in Egypt”, Abdul Latif Jameel Poverty Action Lab, 2022.
- Austrian Development Cooperation, AKPA, UNDP (2022), “Towards Integrated Case Management approaches for vulnerable jobseekers, supporting pathways from social assistance to sustainable employment. Capitalisation report of the pilot implementation in the region of Elbasan”. Tirana.
- Bandiera, O. and Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, Anna Vitaliy (2023), “The Search for Good Jobs: Evidence from a Six-year Field Experiment in Uganda”, July.
- Bossuroy et al (2021), “Pathways out of Extreme Poverty Tackling Psychosocial and Capital Constraints with a Multi-faceted Social Protection Program in Niger”.
- Card D, Kluge J, and A. Weber (2010), “Active labour market policy evaluations: A meta- analysis.” The Economic Journal 120(548): F452-77.
- Card D, Kluge J, and A. Weber (2018), “What works? A meta analysis of recent active labor market program evaluations.” Journal of the European Economic Association 16(3): 894-931.
- Card, D.; Ibarrarán, P.; Villa, J.M. (2011), “Building in an evaluation component for active labor market programs: A practitioner’s guide”, IZA Discussion Paper No. 6085 (Bonn, Institute for the Study of Labour (IZA)).
- Chatri, Abdellatif and Khadija Hadeef and Naima Samoudi (2021), “Micro-econometric evaluation of subsidized employment in Morocco: the case of the ‘Idmaj’ program”, Journal for Labour Market Research, <https://doi.org/10.1186/s12651-021-00300-5>
- Cuevas, P. Facundo, O. Kaan Inan, Aysha Twose, and Çiğdem Çelik (2019), “Vulnerability and Protection of Refugees in Turkey: Findings from the Rollout of the Largest Humanitarian Cash Assistance Program in the World”. © World Bank and World Food Programme.
- Dyer, P, Kluge, J, Bausch, J, Gardiner, D and Mizrokhi, E. (2017), “The impact of skills training on the financial behaviour, employability and educational choice of youth in Morocco”, 3ie Impact Evaluation Report 54. New Delhi: International Initiative for Impact Evaluation (3ie).
- Elsayed, Ahmed and Kevin Hempel, Adam Osman (2018), “Overcoming Youth Unemployment in Egypt: Randomized Evaluations Showcase the Promise of Active Labor Market Programs”, The World Bank, MSME Projects Development Agency, Working paper.
- Elsayed, Ahmed and Rania Roushdy (2017), “Empowering Women under Social Constraints: Evidence from a Field Intervention in Rural Egypt”, IZA Discussion paper series No 11240, December, <https://docs.iza.org/dp11240.pdf>
- ETF, Cedefop, ILO (2016), “Carrying out Tracer Studies. Guide to anticipating and matching skills and Jobs”. Volume 6., https://www.etf.europa.eu/sites/default/files/m/45A4CE81F3398029C1258048005BEFB8_Vol.%206%20Carrying%20out%20tracer%20studies.pdf
- European Commission (2016), “Measuring customer satisfaction with PES. Increasing PES effectiveness by meeting customer needs”. Analytical paper. Written by Lukasz Sienkiewicz.
- European Commission (2018), “Communication on Education in Emergencies and Protracted Crises”.

European Commission (2018), “Labour Market Policy – Methodology”.

Gazeaud, Jules and Nausheen Khan, Eric Mvukiyehe, Olivier Sterck (2022), “With or Without Him? Experimental Evidence on Gender-Sensitive Cash Grants and Trainings in Tunisia”, World Bank Policy Research Working Paper 10132.

GIZ (2015), “Employment Promotion in Contexts of Conflict, Fragility and Violence Opportunities and Challenges for Peacebuilding”.

GIZ (2019), “Identifying employment effects in GIZ interventions Guidelines to support measuring and reporting employment effects”.

Hempel, K. and Mohamed Ali Midani, Souk At-tanmia. “Independent evaluation of the 2012-2013 pilot edition in Tunisia.” <https://tunisia.iom.int/sites/g/files/tmzbd11056/files/inline-files/Souk%20At%20Tanmia%20Independent%20Evaluation%20Report%20Final%202015.pdf>

Hempel, K., Fiala, N. (2012), “Measuring Success of Youth Livelihood Interventions. A Practical Guide to Monitoring and Evaluation”. The World Bank.

ILO (2015), “Making Learning Count. Effective Monitoring and Evaluation of Youth Employment Programs in the Arab World. Taqueem Initiative’. <https://www.ilo.org/publications/making-learning-count-effective-monitoring-and-evaluation-youth-employment>

ILO (2016), Indirect (negative) effects of Active Labour Market Policies (ALMPs), PPT presented in Riga 12-13 April

ILO (2018), “Guide on Measuring Decent Jobs for Youth - Monitoring, evaluation and learning in labour market programmes. Note 5: Identifying an appropriate impact evaluation method/ International Labour Office”. Geneva, 2018.

Instituto Latinoamericano de Planificación Económica y Social (ILPES) de la Comisión Económica para América Latina y el Caribe (CEPAL) 2021, Evaluación de diagnóstico y diseño programa becas y apoyos para la población vulnerable – e057/q3161 Informe final instituto para el desarrollo y atención a las juventudes del estado de guanajuato Consultora: María Claudia Ormazábal Abusleme 15/12/2021.

Karlan, D. and Ryan Knight, Christopher Udry (2014), “Consulting and Capital Experiments with Microenterprise Tailors in Ghana”.

Lauringson, A. and M. Lüske (2021), "Institutional set-up of active labour market policy provision in OECD and EU countries: Organisational set-up, regulation and capacity", OECD Social, Employment and Migration Working Papers, No. 262, OECD Publishing, Paris, <https://doi.org/10.1787/9f2cbaa5-en>

Lubyova, Martina and Diakonidze, Ana (2022), “Evaluation of Public Works Programme in Georgia”, EU-funded project “Technical assistance to skills development matching labour market needs” in Georgia, December 2022, not published.

McKenzie, D. and Susana Puerto, Frank Odhiambo (2019), “Unpacking the determinants of entrepreneurship development and economic empowerment for women in Kenya”; 3ie.

Meštrović, Branka and Ranko Markuš (2021), "Evaluation of the Active Labour Market Measure Startup 2018", Federation of Bosnia and Herzegovina Federal Employment Institute, Niras, Gopa, Pôle emploi.

Morey, M. and Hannah Ring, Ella Hutzler, Haya Breish, Jochen Kluve, Anna Wessendorf, Leonard Dublatz (2023), “Impact Evaluation of the Integrated Skills Training Programme for Disadvantaged Adolescents and Youth in Lebanon.” Final Evaluation Report. AIR, Unicef, Statistics Lebanon, KfW.

Muralidhara, K. and Paul Niehaus, Sandip Sukhtankar (2022), “General equilibrium effects of (improving) public employment programs: experimental evidence from India”.

OECD (2020), “Impact evaluation of labour market policies through the use of linked administrative data.” Final report. VS/2019/0261- Joint OECD-EU analysis of labour market policies.

OECD (2022), “States of Fragility 2022”, OECD Publishing, Paris, <https://doi.org/10.1787/c7fedf5e-en>

OECD (2023), Glossary of Key Terms in Evaluation and Results Based Management for Sustainable Development (Second edition), OECD Publishing, Paris <https://doi.org/10.1787/632da462-en-fr>

Presidential Decree no.102/2022 on the implementation, coordination, monitoring and evaluation of state aids, Türkiye.

Regional Cooperation Council (2021), “Impact Assessment of the Active Labour Market Measures in North Macedonia”.

Roxin, H; A. Kocks, R. Wedel, N. Herforth, T. Wencker (2021), Die Wirksamkeit deutscher Entwicklungszusammenarbeit bei Konfliktbedingten Fluchtkrisen. Die Beschäftigungsoffensive Nahost, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Secretaría de trabajo y fomento al empleo (Ciudad de Mexico) (2023), “Programa social, ‘fomento al trabajo digno’. Evaluación Interna 2023”, Dirección general de empleo, Dirección de programas de apoyo al empleo.

UNDP et al (2019), “Impact Evaluation of Employment Promotion Programmes in Albania”, May.

UNEG, (2013), “Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management”.

World Bank (2023), “Revised Classification of Fragility and Conflict Situations for World Bank Group Engagement”.

World Food Programme (2024), “Barrier Analysis for the social-economic integration of refugee population and host communities in Armenia”. February.