International Labour Organization

WORLD BANK GROUP

The impact of active labour market programmes on youth

A systematic review and meta-analysis of impact evaluations, 1990–2022

Co-funded by the European Union

Ministry of Foreign Affairs

ISBN: 9789220435212 (print); 9789220435229 (web PDF)
DOI: https://doi.org/10.54394/00034312

Printed in Switzerland

▶ Acknowledgements

This report, *The Impact of Active Labour Market Programmes on Youth,* was authored by Jonathan Stöterau and Michael Weber from the World Bank Group and Susana Puerto, Chiara Curcio and Jonas Bausch from the International Labour Organization (ILO). It was made possible through financial contributions from the Policy and Operations Evaluation Department (IOB) at the Ministry of Foreign Affairs of the Netherlands, the ILO and the World Bank Group. The finalisation and publication of the report also benefited from the support of the EC-funded project *Supporting youth employment and investment in skills in developing countries*.

The study benefited greatly from IOB's technical expertise, in particular the advice and suggestions from Paul de Nooijer and Caspar Lobbrecht.

A large team of external collaborators contributed greatly to the development of the study at various stages. We want to thank Germán Mauricio Ramírez Arciniegas for his support in the planning and implementation of the search and coding processes. We thank José Campos de Toledo Junior, Julian Cantor García, Marly Celis Galvez, José de los Heros Montori, Nourhan Hashish, Takaaki Kizu, Pilar Ouro Paz, Yamila Simonovsky and Marton Vegh for their excellent contributions to the search and screening of the studies included in this systematic review. We also thank Tharcisio Leone, Julius Rüschenpöhler and Vivan Coelho who coded numerous studies.

Special thanks go to Evangelia Bourmpoula, Yamila Simonovsky, Paulo Fernandes and Alison Gilberto for their technical expertise and significant contributions to the coding, analysis and review of programme costs and to Felix Weidenkaff for the important feedback and advice in the early stages of the study. We thank Dorothea Schmidt-Klau, Chief of the ILO's Employment, Labour Markets and Youth Branch, for her support to this study, the Decent Jobs for Youth team at the ILO for their inputs and Rosette Opiyo for her administrative support.

The study benefited from the technical advice and kind contributions from several researchers who shared their recent publications and advice on relevant studies for inclusion. In particular, we thank Johanna Kemper, Andrea Ghisletta and our co-author Jonathan Stöterau for providing us with the outputs of their recent meta-analysis of youth skills development interventions and Howard White and co-authors of the recent evidence gap map of youth employment interventions for providing access to the references of their included studies. Methodological discussions with Howard White and Hugh Waddington were very helpful during the analysis.

Finally, we are grateful for the feedback and comments received from experts from the Ministry of Foreign Affairs of the Netherlands, the INCLUDE platform, colleagues from the ILO Employment, Labour Markets and Youth Branch, including Kee Beom Kim, Anna-Karin Palm Olsson, Michael Mwasikakata and Sara Elder and colleagues from the World Bank, including Wendy Cunningham and Dhushyanth Raju.

▶ Contents

Figures

Tables

Boxes

▶ Abbreviations

ALMP	active labour market programme
ATE	average treatment effect
ATET	average treatment effect on the treated
ATT	average treatment effect on the treated
CI	confidence interval
FAT	funnel asymmetry test
ILO	International Labour Organization
ITT	intention-to-treat
IV	instrumental variable
LATE	local average treatment effect
NEET	not in employment, education or training
NGO	non-government organization
OECD	Organisation for Economic Co-operation and Development
PEESE	precision-effect-estimate with standard error
PET	precision effect test
SMD	standardized mean difference
WLS	weighted least squares

All $ references are US dollars.

Executive summary

Background

Despite improvements in global labour markets, many young people continue to face significant barriers to decent work. According to International Labour Organization estimates, the global youth unemployment rate fell to around 13 per cent in 2023, the lowest level in 15 years. Yet, progress remains uneven across regions and population groups. Young people in the Arab States, East Asia and South-East Asia and the Pacific, for example, account for unemployment rates above pre-crisis levels (ILO 2024).

Globally, around 20 per cent of young people, approximately 256 million people aged 15–24, were not in education, employment or training (NEET) in 2024. Young women accounted for nearly two thirds of this group, underscoring persistent gender inequalities (ILO 2025). Opportunities for secure and stable work remain limited. In low-income countries, only one in four young workers holds a regular job, compared to three in four in high-income economies. Looking ahead, roughly 1.2 billion young people in low- and middle-income countries will enter the labour market or seek skills development over the next 15 years, making decent jobs for youth one of the foremost challenges for the world of work today (ILO 2024).

Active labour market programmes (ALMPs) remain a cornerstone of policy efforts to improve youth employment outcomes. Well-designed and well-implemented ALMPs can help young jobseekers acquire relevant skills and access jobs in wage employment or self-employment. Previous reviews (Kluve et al. 2019 and 2017) highlighted the positive impacts of ALMPs on employment and earnings outcomes of young beneficiaries and the importance of integrated and multipronged interventions that address both labour demand and supply-side constraints.

In a context of economic uncertainty, tight fiscal space and evolving labour market demands, strengthening the evidence base on what works for youth employment is critical. Understanding "what" types of ALMPs are most effective, for "whom", under what conditions and "how" their cost-effectiveness can be enhanced is central for guiding investment decisions and advancing a more inclusive and resilient future of work for young people.

Objective

This systematic review and meta-analysis examined the effectiveness of youth-targeted ALMPs in improving labour market outcomes of young people. Drawing on empirical evidence from the past three decades, it assessed the role that context, youth characteristics, evaluation features and programme design and implementation had in moderating the impact on employment, earnings and business performance outcomes.

Data and methodology

The systematic review reflects a significantly expanded database, benefiting from a surge in impact evaluations of youth employment programmes in recent years. This updated review also employed a refined and more comprehensive literature search to include studies conducted between 1990 and 2022 that met the following inclusion criteria:

1. Investigate an ALMP that was primarily targeted at young people (broadly applying the age range of 15–35 years) and designed to address their labour market constraints.
2. Assess an ALMP that included at least one of the eligible interventions: skills training (outside the formal education system), entrepreneurship promotion, employment services, wage subsidies and public works programmes.
3. Measure change in at least one primary outcome of interest: employment, earnings and business performance.
4. Estimate the causal effect of the intervention using an experimental or quasi-experimental evaluation design to identify counterfactual outcomes in the absence of the intervention.

The systematic search and selection process resulted in 5,051 treatment effect estimates, extracted from 228 reports, which assessed 220 interventions within 171 ALMPs. Treatment effects from each study were collected and coded for various subgroups, including age cohorts, gender and education level. They also covered various outcome constructs, such as employment probability, earnings or wages and business performance. Additionally, detailed information was coded regarding the design and implementation features of each intervention, beneficiary characteristics and evaluation methods. The coding process was based on a unified coding manual to ensure consistency in extracting and interpreting relevant information from studies. To complement the analysis on the impact of youth-targeted ALMPs, information on programme costs and the availability of cost–benefit analyses was captured.

The resulting database is of high quality and comprehensive regarding its geographical scope, intervention types and outcome measurements. The systematic review included studies from 62 countries. Nearly 60 per cent of the 220 included interventions were implemented in low- and middle-income countries, with Africa accounting for 33 per cent of all interventions. Skills training was the most commonly evaluated intervention type, constituting 57 per cent of the included interventions. Moreover, the analysis benefited from a large share of high-quality evaluation designs, with 47 per cent of studies conducted as a randomized controlled trial and 38 per cent published in peer-reviewed journals. Ninety-nine of the 220 included interventions provided cost information, and 65 had cost–benefit analysis.

To synthesize the evidence and assess what determined the success of ALMPs, standardized mean differences (SMDs) were computed and a meta-analysis was conducted. SMDs allow comparing the magnitude of the treatment effects across studies by capturing effect sizes in a dimensionless way. Accounting for missing or potentially erroneous information, the final analysis sample encompassed 4,060 effect size estimates, corresponding to 210 reports and 204 interventions. Random-effects meta-analysis methods were employed to synthesize and compare the SMDs. Subsequently, multivariate meta-regression models were estimated to assess how effect size magnitude correlated with study and intervention characteristics. Subsample analysis was conducted to assess whether such correlates differed by country-income level and outcome category. The empirical design addressed several challenges common to meta-analysis, such as publication bias and methodological differences between studies.

The review also took a deep dive on Africa, leveraging the growing evidence on youth employment from that region. The analysis made detailed descriptive overviews of the studies and intervention characteristics specific to Africa and assessed the role that design and implementation features had in moderating the impact of ALMPs across the continent.

Results

The report highlights seven findings.

One, ALMPs were effective in improving the labour market outcomes of youth. The overall effect size across all studies was 0.08 SMD, with a 95 per cent confidence interval (CI) of 0.07–0.09 SMD. After adjusting for publication bias, the estimated overall effect size was 0.06 SMD and statistically significant at the 1 per cent level. Recent meta-analyses in social sciences define “medium” effects as those in the 0.05–0.20 SMD range. Hence, the findings from this study suggested a true underlying medium and positive impact of youth-targeted ALMPs, on average. Moreover, the study documented a statistically significant positive impact of ALMPs across all country income groups and outcome categories.

Two, youth-focused ALMPs were more effective in the low- and middle-income countries than in the high-income countries, on average. The study found an effect size of 0.09 SMD (95 per cent CI of 0.06–0.11 SMD) for the low-income countries, 0.10 SMD (95 per cent CI of 0.08–0.12 SMD) for the middle-income countries and 0.06 SMD (95 per cent CI of 0.04–0.08 SMD) for the high-income countries. The larger impacts in the low- and middle-income countries might be explained by the decreasing returns to investment, as suggested by Kluve et al. (2019): Youth in low- and middle-income countries often have lower levels of formal education at baseline and face more intense labour market challenges. Hence, any marginal investment to improve their situation pays off more there than in high-income countries. Interestingly, in the low- and middle-income countries, the ALMPs had larger impact on earnings than on employment outcomes. The opposite was observed for the studies from the high-income countries.

Three, the type of ALMP (referred to in the report as the "what") had an important role in determining its success.

- **Overall, entrepreneurship promotion and skills training interventions reported larger impacts than employment services and subsidized employment programmes.** This result contrasted with the previous findings by Kluve et al. (2019 and 2017) reflecting that more recent evidence better captured the differential effect of certain intervention types. Some important nuances emerged across country income groups.
- **Entrepreneurship programmes thrived where jobs were scarce.** In the low- and middle-income countries, interventions that promoted self-employment showed the largest impacts, followed by employment services and skills training interventions. These results reflected the realities of limited formal job opportunities, skills mismatches and labour market information asymmetries. By supporting entrepreneurship, such programmes helped young people generate income and move towards more productive activities.
- **Skills training interventions reported the largest impacts in the high-income countries**, particularly on employment outcomes. This finding aligned with Card, Kluve and Weber (2018), confirming that human-capital-centred ALMPs tend to achieve larger impacts in high-income country contexts, particularly in the long run. The evidence on entrepreneurship interventions in high-income countries was limited, however, pointing to an area for further research.

Four, "how" an intervention was designed and implemented was an important determinant of better labour market outcomes.

- **Comprehensive approaches delivered stronger results.** In the low- and middle-income countries, programmes that combined multiple components, such as skills training, job matching or wage subsidies, achieved better outcomes for young people than single, short interventions. Integrating soft skills training and offering certificates that validated participation or acquired competencies further enhanced effectiveness. These features helped young jobseekers not only build up their capabilities but also signalled them credibly to employers (Stöterau, Kemper and Ghisletta 2022).
- **Duration mattered.** In the low- and middle-income countries, youth employment programmes were most effective when they ran for at least four months, allowing enough time to tackle the many intertwined labour market barriers young people face. In the high-income countries, however, a longer duration did not necessarily translate into greater impact, reflecting different labour market realities and institutional capacities.
- **Partnerships were important for impact, yet context mattered.** In the low- and middle-income countries, public–private partnerships linked to larger impacts, whereas in the high-income countries, interventions that were implemented by public stakeholders alone showed larger impact. This finding speaks to the value of partnerships in developing countries, especially with local NGOs, when looking to understand and address multiple labour market constraints.

Five, for "whom" impacts were reported mattered: Impacts were larger for youth from disadvantaged backgrounds, young women and participants younger than 25 years, particularly in the low- and middle-income countries. This result highlighted the potential of youth-targeted ALMPs to promote inclusion and reduce inequalities in the labour market. In Africa, the new evidence showed significantly larger impacts on young people from disadvantaged backgrounds, highlighting the value of interventions that target and tailor support to those most at risk of being left behind.

Six, regional or local programmes tended to achieve larger impacts on youth labour market outcomes than those implemented nationally. The finding suggested that localized approaches may allow for a closer engagement with employers and community organizations, more homogeneous participant groups and programme design that is better adapted to local labour market realities. This offers promise for municipal or regional programmes to address youth employment challenges, for example through tailored targeting and outreach strategies.

Seven, understanding costs and benefits strengthened the case for investment. The updated systematic review found an encouraging trend in the inclusion of cost information and cost–benefit analyses in impact

evaluations since the mid-2010s. Nearly three quarters of the available cost–benefit analyses reported that programme benefits surpassed the programme costs. These were overrepresented among skills training interventions (at 95 per cent) and thus call for greater attention to assessing cost-effectiveness in other intervention types. At the same time, this systematic review found limited evidence on relative effectiveness across intervention components or from the combination of different intervention types.

Conclusion

This systematic review and meta-analysis collated robust evidence on how youth-targeted ALMPs are effective policy tools for improving labour market outcomes of youth around the globe, particularly in low- and middle-income countries and among young people with the greatest risk of exclusion. The analysis distilled three decades of impact evaluations, revealing that programmes designed with a comprehensive approach, certification of participation and skills acquired and with public–private partnerships yielded the strongest outcomes in terms of employment and earnings.

Yet, these promising results come at a pivotal moment when young people face an arduous transition into the labour market worldwide. The rise of remote and hybrid work models, advancements in artificial intelligence and automation and the rapid growth of the gig economy have fundamentally reshaped the labour market. These shifts have made the experience of young workers today distinct, offering both new opportunities and fresh challenges. This transformation unfolded amid the global pandemic, which, compounded by recent wars and economic uncertainty, left an indelible mark on an entire generation. High rates of informality, low-quality jobs and persistent inactivity remain major hurdles, even as youth unemployment rates have declined in some regions.

In light of these findings, several implications emerge for policy and practice:

- Well-designed ALMPs should be scaled up and tailored to the needs of marginalized youth, especially in low- and middle-income countries.
- Youth employment programmes that integrate multiple components (such as skills training, job matching and entrepreneurship support), integrate soft skills, offer certification and involve employers, civil society and communities tend to be most effective, especially in developing contexts.
- Investing in localized, regional and context-specific interventions and strengthening outreach to disadvantaged groups, young women and younger youth should be prioritized to maximize impact.

Despite progress in advancing the evidence base on youth employment and in incorporating cost–benefit analyses, more efforts are needed to bridge knowledge gaps and provide practical guidance for ALMP design and implementation. Persistent gaps remain regarding long-term effects, business performance outcomes, entrepreneurship programmes in high-income contexts and overall evidence from Arab States. More comparative analyses and relative cost-effectiveness studies are needed to assess the effects of different intervention components and delivery models. Rigorous evaluation and cross-country learning should continue to guide future investments.

Overall, ALMPs are a vital instrument for advancing youth employment and inclusion. Continued targeted investments, with attention to comprehensive design, partnerships and certification, are essential to address the persistent challenges confronting young people in today’s rapidly changing world of work.

Introduction

In 2023, the global youth unemployment rate reached its lowest point in 15 years, standing at 13 per cent. This represented a recovery from the COVID-19 crisis, yet it reflected a reality for nearly 65 million youth actively seeking a job. The recovery has been uneven, with youth unemployment rates in the Arab States, East Asia and South-East Asia and the Pacific still exceeding pre-pandemic levels. Moreover, gender disparities in youth unemployment remain a pressing concern. During the pandemic, the rate of unemployment among young women rose to match that of young men, but their recovery has been slower (ILO 2024; Elder and O'Higgins 2023).

Although global youth unemployment has declined, the proportion of young people not in employment, education or training (NEET) continues to be large, holding at 20.4 per cent in 2024 (ILO 2025). Young women are particularly affected, with NEET rates more than double those of young men. In many low- and middle-income countries, the lack of productive employment remains the biggest labour market challenge. Needing to take up any employment they can find to make a living, many young people who do get a job endure poor quality employment and low wages. In low-income countries, only one in four young workers holds a regular, secure job, compared to three in four in high-income countries. This precarious situation has led to increasing anxiety among young people about job security and their ability to achieve financial independence (ILO 2024), with potential adverse impacts on their mental well-being (ILO 2020a). Targeted and urgent investments in youth employment are therefore essential to facilitate successful transitions into decent work.

Well-designed and well-implemented active labour market programmes (ALMPs) can help young people access jobs and prevent long-term scarring from a crisis. Not surprisingly, ALMPs were an essential component of COVID-19 recovery plans (Asenjo, Escudero and Liepmann 2022; ILO 2022; ILO 2020b; Barford, Coutts and Sahai 2022). Kluve et al. (2019 and 2017[1]) documented the effectiveness of ALMPs on youth labour market outcomes, showing positive impacts on both employment and earnings. However, despite these positive results, the overall effect of ALMPs has generally been considered modest, with significant variation depending on the type of intervention and influenced by contextual and design factors.[2]

To assist policymakers in addressing the employment challenges of young people, the systematic review presented in this report analysed empirical evidence on youth employment policies and programmes of the past three decades (from 1990 to 2022). The study examined the effectiveness of youth-targeted ALMPs to improve labour market outcomes of beneficiaries, assessing the moderating role of youth characteristics, context, evaluation features and programme design and implementation.

Relative to Kluve et al. (2019 and 2017), this systematic review more than doubled the sample for analysis: from 113 to 228 reports, from 107 to 220 interventions and from 3,629 to 5,051 treatment effect estimates. The vast increase in the number of studies was due to the surge in impact evaluations of youth employment programmes in recent years, a refined search approach and the input from many researchers who shared their recent evaluations and literature reviews.

The study found that ALMPs had been effective in improving labour market outcomes for young people. The overall effect size across all outcomes was 0.08 standardized mean difference (SMD), with a 95 per cent confidence interval (CI) of 0.07–0.09 SMD.[3] After adjusting for publication bias and concentrating on employment and earnings outcomes, the estimated overall effect size was 0.055 SMD and statistically significant at the 1 per cent level. Based on recent meta-analyses in social sciences, this effect is considered "medium" (in the 0.05–0.20 SMD range),

[1] Kluve et al. published accounts of the same data in 2017 and in 2019. The 2017 publication with the Campbell Collaboration Group focused on the systematic review of youth employment interventions and the resulting effect sizes. The 2019 publication in the *World Development Journal*, built on the previous report and added analyses emerging from multivariate meta-regressions. Both publications relied on a sample of studies published between 1990 and 2014.

[2] Other related meta-analyses of ALMPs did not focus only on young people and/or offered a regional angle, including Levi Yeyati, Montané, and Sartorio, 2021; Vooren et al., 2019; Card, Kluve and Weber 2018; and Escudero et al. 2018. Other studies only include specific types of interventions or outcomes, such as Stöterau, Kemper and Ghisletta 2022; Piza et al. 2016; Grimm and Paffhausen 2015; Valerio, Parton, and Robb, 2014; and Tripney et al., 2013.

[3] The overall SMD of 0.08 is larger than the one reported in Kluve et al. (2019 and 2017) at 0.04, with a CI ranging between 0.02 SMD and 0.06 SMD.

suggesting a true underlying medium and positive impact, on average. The summary effect size for employment outcomes was 0.053 SMD (95 per cent CI of 0.37–0.068 SMD) and 0.062 SMD for earnings outcomes (95 per cent CI of 0.038–0.085 SMD).

On average, ALMPs were more effective in the low- and middle-income countries than in the high-income countries. In the former, impacts on earning outcomes were greater on average than on employment, while the reverse was true in the latter.

The review findings highlighted the important role of the type of intervention as well as its design and implementation features – what is described as the "what" and the "how". The reviewed ALMPs that focused on delivering entrepreneurship promotion interventions – and to some extent also those focusing on employment services – were significantly more effective than other types of interventions in the low- and middle-income countries. Overall, skills training interventions yielded better labour market outcomes than other intervention types in the high-income countries.

Regarding the "how", interventions that included a component of soft skills training appeared more successful in most settings, as did programmes that provided certification to their participants, thus pointing to the importance of signalling strategies. The effects from youth employment programmes in the low- and middle-income countries appeared to differ in other important ways from those in the high-income countries: ALMPs in the low- and middle-income countries were considerably more successful when multiple components were combined, when the programme duration exceeded four months and when implementation was carried out through a partnership between public and private actors, such as the collaboration between a public employment service agency and a company or a non-government organization. The interventions implemented across Africa were, on average, as effective as the interventions in other low- and middle-income countries, yet there was no evidence that success was driven by soft skills components or certification for participants.

The subsample analysis of youth from disadvantaged backgrounds, young women and participants younger than 25 years revealed larger effect sizes, particularly in the low- and middle-income countries. This result showed that ALMPs can help youth groups that typically find it harder to integrate into the labour market.

The review also offered some insights on the scale of interventions, suggesting that interventions with a local or regional scope linked to better outcomes than national-level interventions, particularly in the high-income countries and the Africa region. This offers promise for municipal or regional level efforts to address youth unemployment, likely through better targeting and outreach strategies.

Consistent with previous reviews, the studies based on randomized experiments showed smaller effect size estimates than quasi- or non-experimental studies. By contrast, the availability of peer-reviewed publications or intention-to-treat estimates did not systematically correlate with better effects. Regarding the duration between intervention exit and outcome measurement, the systematic review found that in high-income countries, effect sizes measured at least one year after the end of the intervention led to higher effect size magnitude. In the case of the low- and middle-income countries, however, effects materialized within 12 months after the end of the intervention, but not beyond that period.

The cost analysis showed increased availability of cost information and cost–benefit analyses, which was more pronounced after the mid-2010s. The trend was observed mainly among skills training interventions. Nearly three of four available cost–benefit analyses reported net gains, with programme benefits surpassing programme costs.

This report is a collaborative effort between the International Labour Organization (ILO) and the World Bank Group, with financial support from the Policy and Operations Evaluation Department at the Ministry of Foreign Affairs of the Netherlands, the ILO and the World Bank Group. It is structured as follows: Section 2 outlines the inclusion criteria and important adjustments made to the review's population, interventions, comparisons and outcomes, in comparison with Kluve et al. (2019 and 2017). Section 3 introduces the search strategy and data collection process. Section 4 discusses the characteristics of the included studies and interventions, along with cost information, and provides an overview of the studies and interventions from Africa. Section 5 explains the methods used for the meta-analysis, after which Section 6 shows the results from a univariate unconditional approach and a bivariate approach as well as through multivariate meta-regressions. It also offers in-depth analysis of findings specific to Africa. Section 7 concludes with the insights and recommendations for policymakers and practitioners on youth-targeted ALMPs and further areas for research. The study is complemented with an extensive set of annexes describing the elements of the search process, coded variables and included studies.

2 Inclusion criteria

2.1 Population and context

The systematic review covered all countries, regardless of their level of development. To be included, studies must have investigated ALMPs that were designed to address employment constraints for young people. The review allowed for various youth definitions adopted by individual programmes, including national definitions and broadly applied the age range of 15–35 years.

2.2 Interventions

Considering the emergence of new interventions to support the transition of young people into employment, such as mentoring and volunteering, and the trend towards further integration of ALMPs within public employment services, the typology of interventions chosen in the 2017 Kluve et al. was reassessed. The assessment compared selective studies and meta-analyses with the Classification for Public Expenditure and Participants in Labour Market Programmes (OECD 2022), as shown in Table 1.

Table 1. Comparison of active labour market programme categories in relevant and recent studies

Review or source	Includes a classification closely matching the following programmes				
	Public employment services and administration	Training	Employment incentives	Direct job creation	Start-up incentives
Heckman, LaLonde and Smith 1999	Yes	Yes	Yes	Yes	No
Kluve and Schmidt 2002	Yes	Yes	Yes	Yes	No
LaLonde 2003	Yes	Yes	No	Yes	No
Betcherman et al. 2007	Yes	Yes	Yes	Yes	Yes
Card, Kluve and Weber 2010	Yes	Yes	Yes	Yes	No
Eichhorst and Rinne 2015	Yes	Yes	Yes	Yes	Yes
Caliendo and Schmidl 2016	Yes	Yes	Yes	Yes	No
Kluve et al. 2019 and 2017	Yes	Yes	Yes	Yes	Yes
Card, Kluve and Weber 2018	Yes	Yes	Yes	Yes	No
Escudero et al. 2019	Yes	Yes	Yes	Yes	Yes
Vooren et al. 2019	Yes	Yes	Yes	Yes	No
Levi Yeyati, Montané and Sartorio 2021	Yes	Yes	Yes	Yes	Yes

Note: Based on authors' comparisons. None of the studies included the OECD categories more closely associated with passive labour market policies: sheltered and supported employment and rehabilitation, out-of-work income maintenance and support, and early retirement.

Considering the limited evaluation evidence from mentoring and volunteering programmes to foster youth employment (O'Higgins 2022) and the strong alignment between classifications across recent literature, the review maintained the 2017 classification of Kluve et al. but narrowed its focus to the ALMPs that were purposely designed for young people. Kluve et al. (2017 and 2019) had evaluated the ALMPs that were designed for wider populations but had measured impacts on youth. The rationale for such change was to factor in a policy or programme intention and corresponding mechanisms set in place for its implementation. In other words, although a programme might have pursued the improvement of labour market outcomes for all disadvantaged people in the labour market, this review only focused on programmes with interventions that targeted young people.

Thus, eligible studies must have evaluated an ALMP that provided at least one of the five categories of intervention described in Table 2.

▶ Table 2. Interventions of interest for the updated systematic review

Intervention	Description
Skills development	This includes skills training programmes offered outside the formal education system to enhance youth employability or facilitate the transition into employment. Employability skills range from job-specific technical skills to non-technical soft (or core) skills, such as self-management, teamwork and communication. Recent frameworks of core skills integrate basic digital skills and basic skills for green jobs (ILO 2021). The review documented the specific skills provided by an intervention, for example, technical skills, digital skills, business skills, literacy or numeracy skills and a composite of behavioural, life or soft skills. The novelty in relation to Kluve et al. (2019 and 2017) was the inclusion of digital skills.
Entrepreneurship promotion	Entrepreneurship promotion interventions aim to provide advisory services and to facilitate access to finance and markets for the development or growth of a youth-owned business. The review clustered interventions across the following services: business advisory and/or mentoring; business skills; access to markets and value chains; direct credit or facilitating access to credit; monetary or in-kind (start-up or growth) grants; and microfranchising initiatives or mechanisms.
Employment services	Employment services generally focus on facilitating matching and intermediation to support the transition of young people into employment through targeted services for jobseekers and employers. The offers to young jobseekers include job counselling, job-search assistance and/or mentoring services for activation (or reactivation) purposes, which are often complemented by job placement and technical or financial assistance. The basic idea for providing employment services to youth is that young workers have difficulty signalling their skills and credentials and/or lack the networks or knowledge to search effectively for vacancies and connect with employers. As a result, these programmes often focus on improving job-seeking skills and the efficiency of the matching process. The review distinguished between the services focused on counselling, on placement and on financial support for the job search.
Wage subsidies	Wage and hiring subsidies provide incentives to employers to hire first-time jobseekers for a given period by reducing labour costs. In return, they allow young workers to build up their work experience and acquire job-relevant skills. The review distinguished between the programmes that reduced employers' social security contributions, those that reduced the labour or wage cost and those that provided direct payments (subsidies) to young people.
Public works	Public work programmes and labour-intensive public employment programmes offer direct, short-term employment to young people in infrastructure, social development or community projects. They are often considered a solution in times of crises, when there is not enough labour demand and thus as a temporary safety net. In addition to delivering useful public assets and services, these programmes create employment for unemployed persons, provide supplementary employment to the underemployed persons and thus generated income and therefore smooth or even boosted consumption patterns.

Note: Kluve et al. 2019 and 2017 provide further information on the underlying theories of change and expected transmission mechanisms of these interventions.

For practical purposes and due to the limited number of studies on public works, the analysis clustered the category with wage subsidies, renaming it as "subsidized employment".

The review distinguished between programmes and interventions: A youth employment programme is a single entity that may consist of one or several interventions. This is increasingly the case because the evidence-based policy advocacy of the past decades has extensively recommended comprehensive and holistic employment programmes for youth to address their various vulnerabilities in the labour market (Kluve et al. 2019 and 2017; O'Higgins 2017; ILO 2010; Fares and Puerto 2009; Betcherman et al. 2007).

It is important to emphasize that if a programme had, for instance, a skills development track and an employment services track, and participants chose one over the other, they were considered to be two interventions within the same programme. In this case, participants were exclusive to one intervention type. If a programme always combined skills development and employment services, it was considered as a single intervention.

An additional consideration was made to identify primary intervention types under multipronged designs. The review defined "main category of intervention" as the largest and predominant intervention type within a programme and therefore eliminated the "unspecified" category that Kluve et al. (2019 and 2017) used.

2.3 Comparison group

This systematic review included studies that measured change in at least one outcome of interest among intervention participants and relative to non-intervention participants based on counterfactual analysis (comparing treated and control groups). Comparison and control groups must have included individuals who belong to the eligible population but *did not* receive any treatment or were due to receive the intervention at a secondary stage. The comparison group of some studies might have been exposed to interventions other than the evaluated intervention.

2.4 Outcome

Table 3 summarizes the primary outcomes of interest: employment, earnings and business performance. The systematic review captured the studies that had reported at least one selected outcome variable, even when they were measured conditional on other outcomes. It excluded studies that focused only on intermediary outcomes, such as the acquisition of knowledge and skills.

▶ Table 3. Outcomes of interest for the updated systematic review

Outcome category	Outcome measures investigated
Employment	Employment probability
	Unemployment probability
	Participation rate
	Hours worked
	Unemployment duration
	Quality of employment (contract, fixed term, benefits)
Earnings	Earnings or income
	Household income
	Consumption
	Salary or wage
	Profits
	Sales
Business performance	Number of employees or jobs created
	Capital and investment
	Business creation
	Business survival

2.5 Study design and methods

The systematic review covered experimental and quasi-experimental evaluations, including all research design categories that aimed to quantitatively estimate the causal effect of the intervention on the outcome of interest. In particular, it included: (i) randomized experiments, (ii) natural experiments, (iii) methods for causal inference under unconfoundedness (regression methods, statistical matching, propensity score matching) and (iv) selection of non-observables (instrumental variables, regression discontinuity design, difference-in-differences).

The review covered several types of publications and grey literature produced between 1990 and 2022: peer-reviewed journals, working papers, mimeos, books, policy or position papers, evaluation or technical reports and dissertations or theses. The review captured publications in several languages, as long as they met the inclusion criteria.

▶ 3 Search and data collection

3.1 Scoping, primary and complementary searching

The systematic review focused on expanding the sample of reports from Kluve et al. (2019 and 2017) with evaluations published after 2014. Subsequently, the review searched for new versions of reports and papers previously included by Kluve et al. as well as reports that may have been overlooked. A "report" was considered the actual paper or report and may have contained several studies. The search process followed guidelines of Kugley et al. (2017) and The Campbell Collaboration (2021). It included a primary search based on general and specialized databases and other electronic repositories and a complementary search that leveraged literature snowballing and direct correspondence with experts and authors.

The primary search began with the identification of controlled vocabulary terms based on the inclusion criteria. Keywords from the 113 studies included in Kluve et al. increased the accuracy of the terms for population, intervention and outcomes, as shown in Table 4. A frequency test of keywords in a group of 50 studies (of the 113 studies from Kluve et al.) led to the identification of 310 keywords.[4] In the absence of key terms, a quick skimming of the abstract or introduction of the report was done to identify keywords related to the controlled vocabulary sets.

▶ Table 4. Keywords and controlled vocabulary terms

Controlled vocabulary set	Keywords for studies from Kluve et al. (2017) and frequency (in parenthesis)	Controlled vocabulary terms
Population (condition of interest) terms	adolescent (3), young (6), trainees (2), youth (26), graduates (2), young individuals (3), young adults (3), unemployed (6), women (1), adults (1), young people (5), entrepreneurs (1), youths (4)	**English**: youth* (30), young* (17), unemployed (6). **Other languages**: jóvenes, jeunes, jovens
Intervention terms	program/me (16), skills (10), training (36), policies (1), services (1), skill (1), entrepreneurial (2), business (5), coaching (1), vocational (9), active labor/labour market programmes (2), payroll tax subsidy (1), subsidy (1), job-training (1), microenterprise (2), assistance (3), empowering (1), grants (3), search (1), labor/labour market program/me (2), job (2), subsidies (4), active labour market policy (3), literacy (1), interventions (1), active labor/labour market policies (1), job-search (2), apprenticeship (1), payroll subsidies (1), subsidized work (1), labor/labour market policies (1), welfare to work (1), payroll tax cuts (1), active labor/labour policy (1)	**English**: training (37), business (5), subsidy (7), search (5), labour market program/me (10) **Other languages**: capacitación, politique active de l'emploi, apoio, políticas activas de mercado de trabalho
Outcome terms	employment (26), income (7), self-employment (3), employability (1), earnings (14), formal employment (3), work experience (1), credit (1), empowerment (2), labor/labour (1), wages (5), business (5), survival (2), enterprise (1), incomes (1), work hours (1), labor/labour market outcomes (4), unemployment (2), hours of work (1)	**English**: employment (32), earnings (14), income (8), business (5) **Other languages**: empleo, ingresos, ocupado, emploi, emprego

Note: *= denotes a term that may include additional letters equally relevant to the search process. For instance, youth* allowed the search to capture instances in which the text used "youths".

Final controlled vocabulary terms were customized for each database (such as IDEAS/RePEc and EconLit). A search statement form captured search outcomes, including whether the database offered wildcards for truncated terms, alternate spelling or proximity connectors. The search included search terms in English, Spanish, French, German

[4] Those 50 studies were selected based on their order of appearance in the list of included studies reported in Kluve et al. 2017.

and Portuguese. In addition to the controlled vocabulary search, the process included key terms from the databases and thesaurus terms.

Thirty-four primary search sources were identified, starting with all sources with at least one of the 113 included studies in Kluve et al. An initial focus was placed on records with a publication date of 2013 or onwards, creating a small overlap with the base sample from Kluve et al.[5] Subsequently, the time range was expanded to 1990–2022.

To select preferred search statements while balancing the comprehensiveness and relevance of the review, sensitivity and precision indexes were computed using the Boolean connectors available in each database (AND OR) and whether the database allowed for key terms and thesaurus terms,[6] leading to a definite reference list. As a result, 58 of 175 search statements with higher sensitivity ratios were selected as preferred search statements, resulting in 79,292 references from the primary search process, which were subsequently exported to EPPI-Reviewer 4.

As noted, the primary search was complemented by literature snowballing and contacting authors and experts, with a focus on impact evaluations published since 2013. Six main sources supported the complementary search: (i) the 113 reports covered in Kluve et al. (2019 and 2017); (ii) 420 references corresponding to a list of included studies in Apunyo et al. (2022); (iii) 89 studies evaluating the impact of vocational training interventions on youth contained in an upcoming (at that time) meta-analyses by Stöterau, Kemper and Ghisletta (2022);[7] (iv) 99 references provided by 41 of the 94 experts and authors of studies included in Kluve et al. who were contacted by the review team; (v) 18 references identified by the research team during the search process; and (vi) 8,148 references emerging from citations or references.

Box 1
Quick note on the terminology

To enhance readability, this review report uses specific terms, as follows.

A youth employment **programme**, such as Jovenes en Accion in Colombia, may have included various **interventions**, such as job search assistance or training. The same intervention may have had multiple **cohorts**, either starting at different times or in different locations. For the analysis, each cohort was considered an independent group of **participants**.

Multiple **studies** could have evaluated these interventions, and each study may have had several **reports**, including working papers and journal publications. However, if two different reports were based on the same sample data, they were not considered independent and were considered part of the same study, even if different authors wrote them.

Each **outcome category** (such as employment outcome) may have been measured by different **outcome constructs** (employment probability, unemployment probability).

Finally, reports may have provided various **treatment effect estimates** for the same outcome category, including for different subgroups, statistical methods or outcome constructs. These treatment effect estimates were then converted into **effect sizes**, such as standardized mean difference, for analysis.

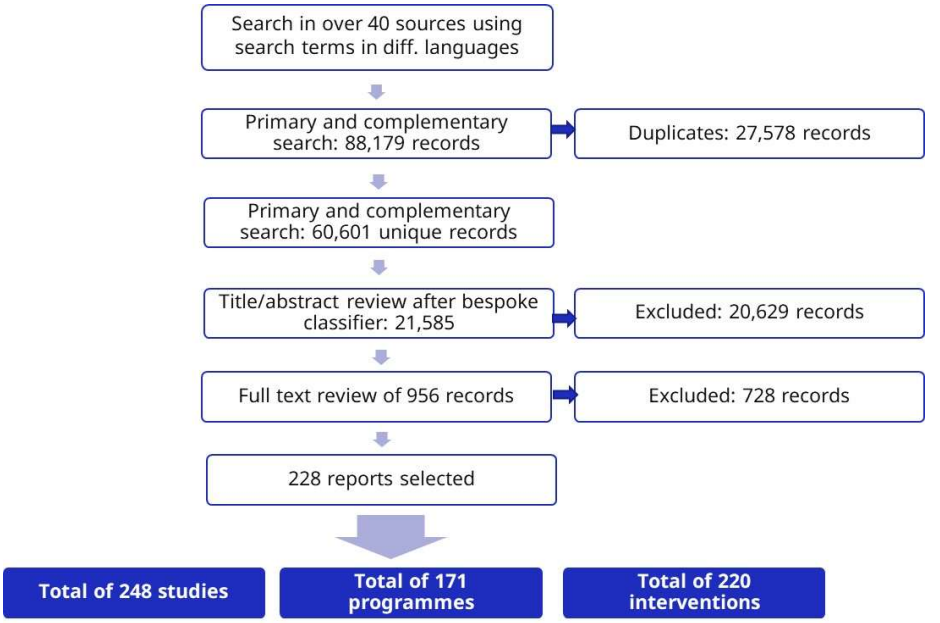
[5] Primary sources to search included subject databases (for example, Education Resources Information Centre, or ERIC), general databases, such as citation indexes (for example, Web of Science), dissertations and theses databases (for example, ProQuest Dissertations and Theses Database), grey literature sources (for example, the Abdul Latif Jameel Poverty Action Lab evaluation and publication database) or search engines (Google).

[6] The relevance of a search statement was based on both the sensitivity and precision within the same database and sensitivity and precision compared to the whole search statements records. "Sensitivity" was defined as the ratio between the number of relevant reports identified and the total number of relevant reports identified (overall and for database). "Precision" was defined as the ratio between the number of relevant reports identified and the total number of reports identified.

[7] The meta-analysis by Stöterau et al. (2022) leveraged the database from Kluve et al. (2017) and focused only on youth skills training interventions. It included new versions of impact evaluations as well as new evaluations, particularly from the period 2015–19. Their resulting database also benefited from important improvements in the coding manual, which were integrated into this review.

The primary and complementary searches resulted in **88,179 records**[8] (references) based on **40 sources** for this review. All records were uploaded into the EPPI-Reviewer and screened for duplicates. Around 31 per cent (27,578 records) of a total of 88,179 records were identified as duplicates (as shown in figure 1).

▶ Figure 1. Search results for updated systematic review



Using the EPPI-Reviewer's "build model" automatic functionality, a machine learning algorithm was developed to identify studies likely to match the inclusion criteria during the title and abstract screening (also known as "bespoke classifier"). The model worked with 95 per cent of the 60,601 non-duplicate records, or 57,677 records. A subsample of 2,925 records was used to train the model.[9] At the end, **18,650 records** yielded a probability greater than 50 per cent of being relevant to the systematic review. This batch of records along with the 2,925 subsample records were taken to the next stage and 21,585 records were double-screened by title and abstract.[10] Subsequently, **956 records** complied with the inclusion criteria at the title and abstract stage and were therefore fully screened, resulting in **228 records** (24 per cent) included and 728 records (76 per cent) excluded. The included reports represent a doubling of those captured by Kluve et al. (at 113 reports).

3.2 Data collection

The data extraction followed the same protocol that Kluve et al. used. Relevant information from included studies was systematically extracted using a coding tool and a coding manual, both respectively updated in view of new or revised variables. The coding tool included information on variables related to study methods, characteristics of the intervention and its implementation, characteristics of the subject samples of analysis, outcome variables, statistical findings and contextual features.

At the effect size level, as before, the coding tool captured subgroup analysis of employment, earnings and business performance outcomes and estimated treatment effects by age cohort, gender, education level, income level and location, among other dimensions. For some subgroups, such as those defined by education and income levels,

[8] In this section, the term "record" is used as synonymous for "report". See box 1 for more details on the terminology adopted in this review.

[9] A subsample of 2,925 references from the search statement "SSID_10_2021" (RePec/Ideas) and "SSID_96_2021" (Apunyo et al., 2022) were screened and coded with included or excluded to train the custom model. As a result of the title and abstract screening of these 2,925 records, 369 records (12.5 per cent) were coded as included and 2,556 (87.5 per cent) as excluded. The EPPI-Reviewer 4 used 90 per cent of this sample to train the model and 10 per cent to evaluate its accuracy (defined as the ratio between precision and recall).

[10] In this edition of the review, the screeners encountered various records with a title but no abstract. The search process also led to studies in languages beyond those in the search protocol, such as Japanese, Korean and Hungarian, but with an available title or abstract in English. After passing the title and abstract screening, they were sent for full screening to ILO colleagues native in those languages.

the review relied on the descriptions and classifications applied in the original studies. Types of outcomes were further disaggregated by occupation category (dependent versus self-employment), status of occupation (formal versus informal) and conditionality on other outcomes.

To describe the data and empirical methods, the coding tool included information about the research design, statistical methodology, type of significance test, type and method of measurement, date of data measurement and data source. The coding tool also captured the form and year of publication. For each category of intervention (skills training, entrepreneurship promotion, employment services, wage subsidies and public works), the coding tool extracted information about the type of intervention, targeting and delivery mechanism, payment system and provider, duration of specific interventions, selection of participants and conditionality of eligibility. This updated systematic review placed strong focus on avoiding what Kluve et al. defined as an "unspecified" category of interventions, which clustered reports for which it was not possible to identify a main category of intervention. The rationale was to reduce ambiguity in the interpretation of results. The method involved intensive discussions by the review team to understand the intensity of different programme components and a joint decision on the categorization. Most reports (and treatment effects) previously coded as "unspecified" intervention category moved into either skills training or employment services.

General programme characteristics recorded the target group by age, gender, education level, income level, location and employment status, as well as the type of organizations involved in designing, financing and implementing the programme. The coding tool kept record of region, country, scale and average duration of the programme. Any awareness-raising efforts and gender considerations integrated into the programme design and implementation were also captured (the annex covers the selected variable definitions, which included some of these design features).

The coding manual provided detailed instructions for coders to ensure consistency in extracting and interpreting relevant information, in particular regarding the selection of appropriate treatment effect estimates. Guidelines identified the treatment effect estimates with lowest risk of bias when studies reported multiple estimates for the same types of outcomes. The coders selected the preferred method of estimating the effect, which was then verified by the coding lead and/or discussed with the principal investigators. For example, estimates based on experimental designs were considered to provide the lowest risk of bias, followed by natural experiments and quasi-experimental designs. Other considerations outlined in the manual to mitigate the effects of potential bias included the use of covariates, type of data used and statistical methodology applied for the estimation.

When in doubt, the coder discussed information extracted from included studies with the review team, and coding decisions involving assumptions were documented in the coding sheets. For this updated systematic review, the authors of included studies were not contacted to provide missing information or clarify discrepancies, which deviated from the systematic review protocol of Kluve et al.

▶ 4 Study and intervention characteristics

4.1 Global overview

Table 5 provides an overview of the studies included in the 228 reports and the 220 interventions that the review documented. As shown in Figure 2, the sample includes a significant share of new literature published from 2015 onwards. The systematic search and selection approach allowed further refinements in the sample of studies published before 2015, and this resulted in the inclusion of 61 updated versions of reports already included in Kluve et al. (2019 and 2017), in addition to a total of 133 new reports for the entire period of analysis.

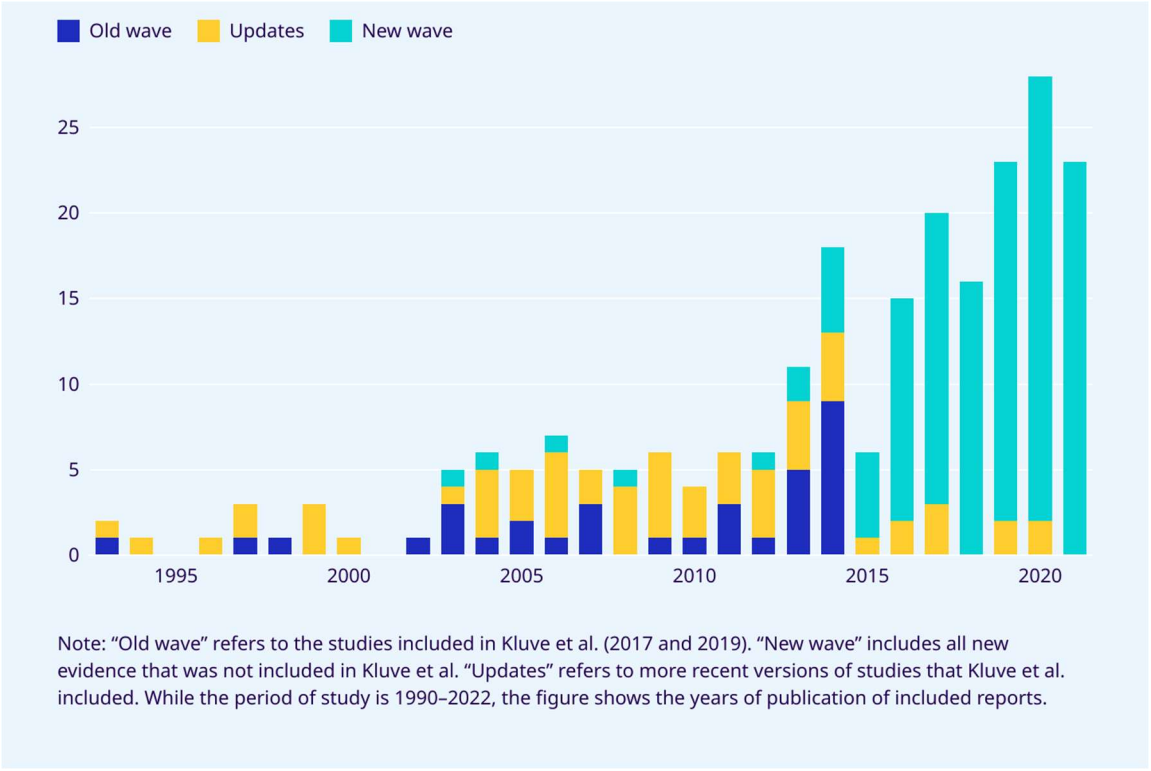
▶ **Table 5. Characteristics of included studies and interventions in the updated systematic review**

	Studies	%		Interventions	%
Publication period			**Country income level**		
1990–2000	10	4	High income	92	41.8
2001–05	15	6	Middle income	92	41.8
2006–10	42	16.9	Low income	36	16.4
2011–15	47	19	**Continent**		
2016–22	134	54	Africa	72	32.7
Publication status			Arab States	4	1.8
Peer-reviewed publication	93	37.5	Asia and the Pacific	19	8.6
Working paper	87	35.1	Europe and Central Asia	59	26.6
Technical report	57	23	Latin America and Caribbean	39	17.7
Other publication type	11	4.4	North America	27	12.3
Evaluation design			**Scale of intervention**		
Experimental (randomized controlled trial)	116	46.8	National	105	48.6
Quasi-experimental	97	39.1	Regional	48	22.3
Other	35	14.1	Local	60	26.8
Estimated parameter			**Actors involved in the design**		
Intention-to-treat effect	115	46.4	Private entity or NGO	132	60
Others (ATE, TOT, LATE, etc.)	101	40.7	Government	88	40
Timing of follow-up measurement			**Actors involved in the implementation**		
Short-term follow-up (less than 1 year)	141	56.9	Private entity or NGO	73	33.2
Medium-term follow-up (1–2 years)	47	19	Government	147	66.8
Long-term follow-up (more than 2 years)	30	12.1	**Main Intervention Category**		
Evaluation by subgroup			Skills training	126	57.3
Low-income and disadvantaged youth	137	55.2	Entrepreneurship promotion	31	14.1
Gender subsample	30	12.1	Employment services	32	14.5
Outcome category			Wage subsidies	29	13.2
Employment	153	61.1	Public works	2	0.9
Earnings or income	79	30.8	**Has intervention type**		
Business	16	8.1	Skills training	145	65.9
Outcome construct			Entrepreneurship promotion	43	19.5
Employment probability	107	43.1	Employment services	76	34.5
Unemployment probability	8	3.2	Wage subsidies	37	16.8
Participation rate	7	2.8	Public works	5	2.3
Hours worked	18	7.3	**Additional skill components**		
Unemployment duration	3	1.2	Business skills	35	15.9
Quality of employment	10	4	Soft skills	80	36.4
Earnings	42	16.9	Digital skills	6	2.7
Wage	2	0.8	**Duration of intervention**		
Analysis sample by sex			Less than 4 months	67	30.5
Male	22	8.9	4–8 months	45	20.5
Female	41	16.5	More than 8 months	54	24.5

	Studies	%		Interventions	%
Female and male together	184	74.2			
Analysis sample by age			**Additional design elements**		
Younger (than 25 years)	174	70.2	(Non-) monetary support	105	47.7
Older (than 25 years)	74	29.8	Certificate	52	23.6
Total studies	**248**		**Total interventions**	**220**	

Note: Studies may have been exclusive across the different typologies. For example, one study may have estimated multiple outcomes or looked into more than one intervention type. Shares reported in this table are computed out of non-missing values for the following variables: intention-to-treat effect (ITT), low-income or disadvantaged youth sample, male, female and female and male together. Similarly, interventions may not have been exclusive across the different typologies. For example, one intervention may have more than one characteristic reported in this table. The shares were computed out of non-missing values for the following variables: scale of intervention, intervention type and intervention duration.
ATE = average treatment effect; TOT = effect of treatment on the treated; LATE = local average treatment effect.

▶ Figure 2. Increasing body of evidence, by publication year, 1993–2021



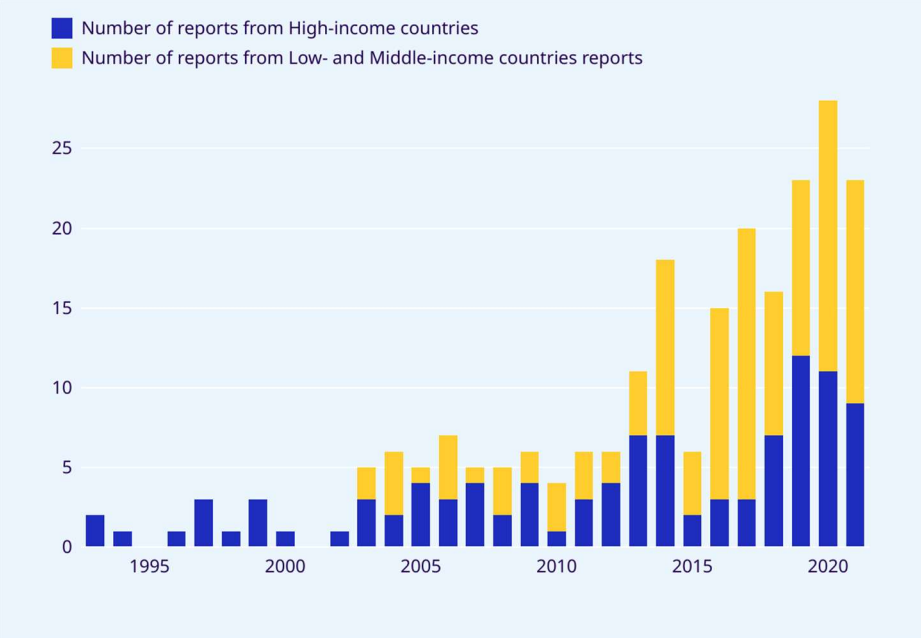
Note: "Old wave" refers to the studies included in Kluve et al. (2017 and 2019). "New wave" includes all new evidence that was not included in Kluve et al. "Updates" refers to more recent versions of studies that Kluve et al. included. While the period of study is 1990–2022, the figure shows the years of publication of included reports.

When looking at the sample's **geographical composition** (Figure 3), 42 per cent of the interventions included in this systematic review were implemented in high-income countries and 58 per cent in low- and middle-income countries, of which one third of the evaluations was carried out in Africa. The increased body of evidence from low- and middle-income countries (58 per cent of interventions), represents an important development in the impact evaluation field, compared with the 42 per cent share reported in Kluve et al. (2019 and 2017).

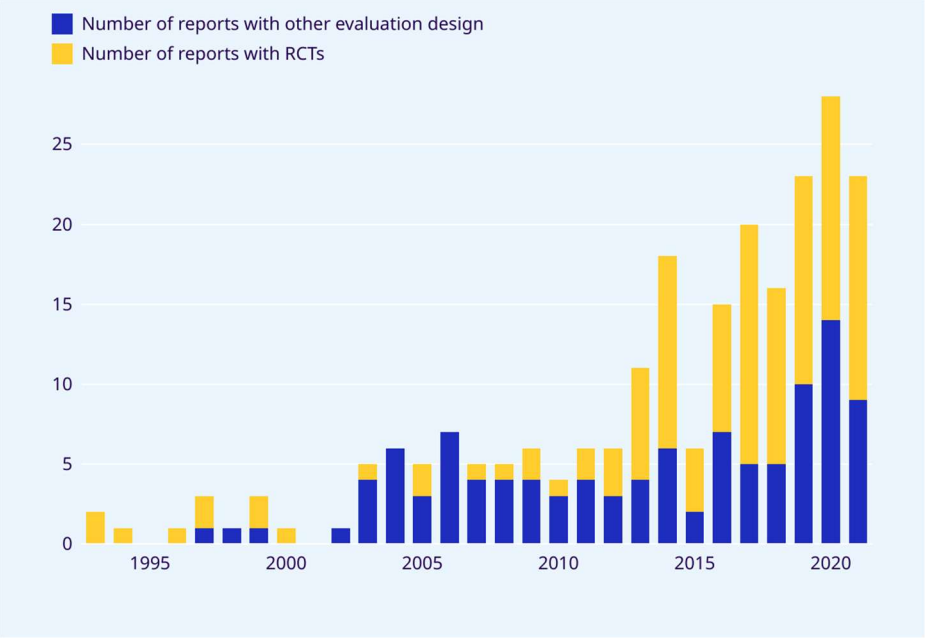
Regarding **publication status and study design**, the extensive search effort was also successful in identifying unpublished studies – the so-called "grey" literature. Around 38 per cent of the studies were from peer-reviewed journals, with the remainder covering technical reports from implementing organizations, working papers and other publication types (thesis and dissertations). The number of experimental studies or randomized control trials increased sharply, accounting for more than half of the total sample (Figure 4).

Table 5 also provides an overview of the **types of outcomes** evaluated by these studies. Most studies in the sample reported results for more than one outcome. Employment probability, earnings and hours worked were, by far, the most commonly measured and reported outcome construct within the sample, with 107, 42 and 18 studies, respectively.

▶ Figure 3. Increasing body of evidence from low- and middle-income countries, 1993–2021



▶ Figure 4. Increasing number of randomized control trial reports, 1993–2021

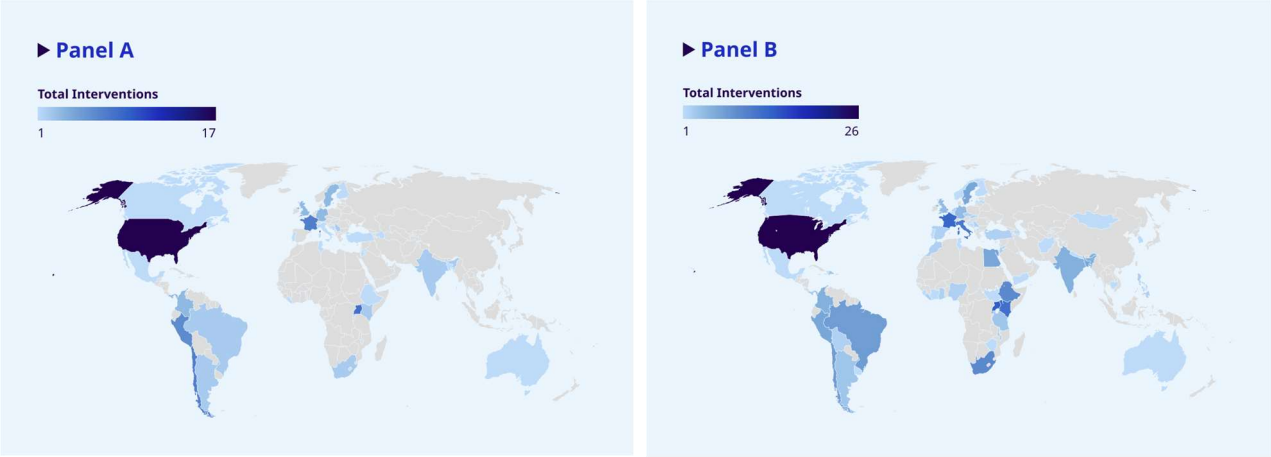


In relation to the **evaluation features**, 218 studies provided follow-up measurements at different points in time. In particular, 77 studies (31 per cent) measured changes in the outcomes of interest over 12 months after treatment exposure. These longer-term effects were estimated primarily across the skills training interventions. Some studies provided subgroup analysis in addition to the overall analysis. As shown in Table 5, more than half of the reports in the sample provided separate results for low-income and disadvantaged youth. On the other hand, one fourth of the studies reported impact estimates for a gender subsample. Finally, with respect to the Kluve et al. review, there was an increased number of studies (nearly 70 per cent of the sample) that evaluated impact estimates on persons younger than 25 years.

Panels A and B in Figure 5 show the evolution of youth employment **interventions** worldwide. The darker or coloured countries mark the increased number of interventions included in the review, particularly in low- and

middle-income countries. There has been an important increase in the number of evaluated interventions in Africa and in Asia and the Pacific since 2015.

▶ Figure 5. Evolution of evaluated youth employment interventions before (panel A) and after (panel B) 2015



Regarding **intervention categories**, this systematic review confirmed (as in Kluve et al. 2019 and 2017 and Card, Kluve and Weber 2018) that skills development interventions (as the main category) were the most commonly evaluated intervention type, constituting 57.3 per cent of the 220 included interventions. Sixty-six per cent of the interventions reported having a skills-related component, while business or soft skills were reported in 52.5 per cent of all interventions. Within the skills development interventions included in this systematic review, 86 per cent were delivered in a classroom and 55 per cent at a workplace. Only a small number of interventions presented more advanced and modern training approaches. In particular, only 2 per cent of the 126 skills training interventions included were delivered with a distance learning approach, while 5 per cent had a digital skills component. Globally, one third of the skills interventions in the sample targeted urban areas only, while almost one in every two interventions targeted both rural and urban areas. Partly due to the large number of skills training interventions in the sample, there was an important variation regarding the groups of young people these interventions aimed to serve, their scale, location and the engagement of private and public actors.

Although only 15 per cent of the interventions were mainly categorized as employment services, almost one in three of them included an employment service component, such as job placement, mentoring and coaching or job search assistance. And 20 per cent of the interventions included a component on entrepreneurship promotion, while 18.7 per cent had a component related to subsidized employment.

As for **implementation features**, two in three interventions involved the government or multilateral organizations in the delivery mechanisms, while one third involved civil society organizations or the private sector.

4.2 Regional differences

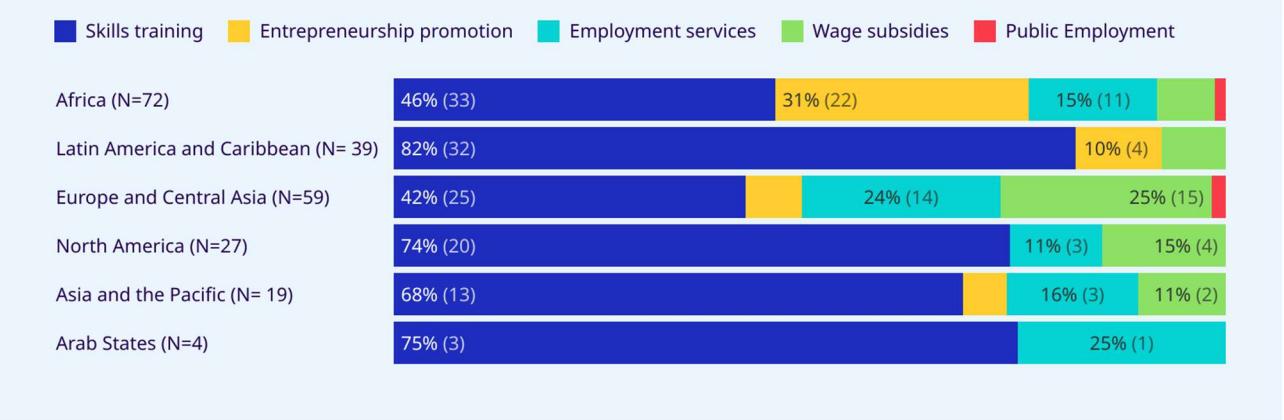
The search and selection process ended up with a rich and diverse sample of studies and interventions that provided interesting insights on how ALMPs for youth were designed and implemented worldwide. This section focuses on the major regional variations observed at the intervention level, highlighting those that emerged from the increasing body of evidence coming from developing countries.

Figure 6 shows the regional coverage across main category of interventions. As stated in section 4.1, skills training was the most common intervention implemented worldwide. Entrepreneurship promotion interventions were, on the other hand, mainly implemented in Africa. Europe and Central Asia presented a higher incidence of employment services. The number of evaluated interventions from public works programmes was relatively small.

Table 6 shows the frequency of intervention types as components within broader comprehensive designs. Skills interventions were often integrated into the design of programmes across all regions, with the largest regional share in North American countries, followed by Latin America and Caribbean. Africa had some level of diversity

across skills, entrepreneurship promotion and employment services, while in Europe and Central Asia, the presence of entrepreneurship-related interventions was rather limited.

▶ Figure 6. Main categories of interventions, by region



Note: The values in the bar segments represent the number of interventions, not the percentages.

▶ Table 6. Different components of interventions, by region

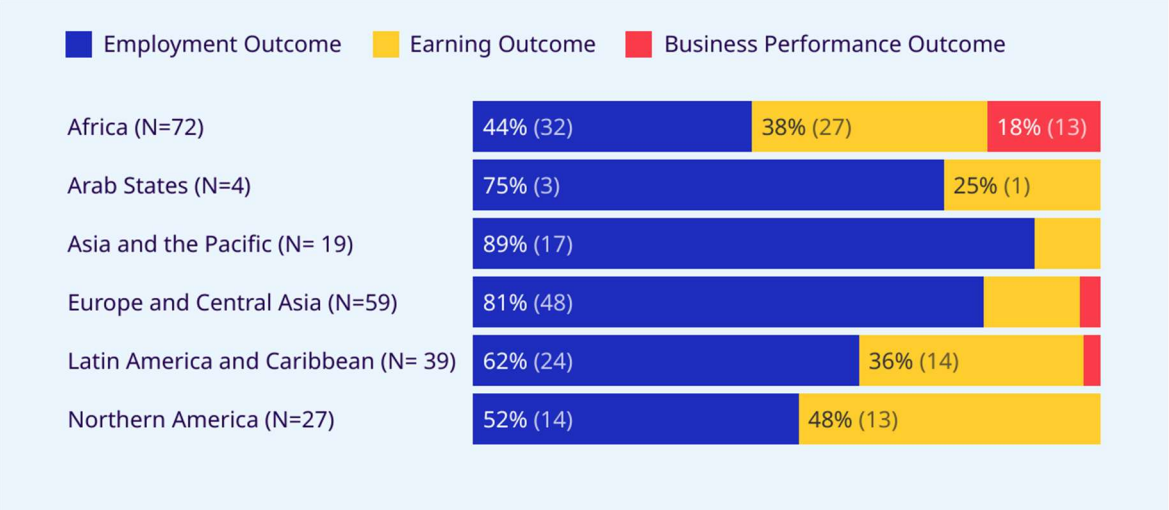
	Africa	Arab States	Asia & the Pacific	Europe & Central Asia	Latin America & Caribbean	North America	Total
Total interventions	**72**	**4**	**19**	**59**	**39**	**27**	**220**
Has skills training component	36	3	13	36	32	25	145
%	*50.0%*	*75.0%*	*68.4%*	*61.0%*	*82.1%*	*92.6%*	*65.9%*
Has entrepreneurship promotion component	32	0	2	4	4	1	43
%	*44.4%*	*0.0%*	*10.5%*	*6.8%*	*10.3%*	*3.7%*	*19.5%*
Has employment services component	19	2	9	21	7	18	76
%	*26.4%*	*50.0%*	*47.4%*	*35.6%*	*17.9%*	*66.7%*	*34.5%*
Has wage subsidies component	6	0	3	19	5	4	37
%	*8.3%*	*0.0%*	*15.8%*	*32.2%*	*12.8%*	*14.8%*	*16.8%*
Has public works component	1	0	0	3	0	1	5
%	*1.4%*	*0.0%*	*0.0%*	*5.1%*	*0.0%*	*3.7%*	*2.3%*

The largest share of interventions that had business and soft skills components (38 and 24 per cent, respectively), as well as five of the six digital skills interventions included in this systematic review, were implemented in Africa (the remaining one was in Europe). In addition to Africa, the other region where there was a large number of interventions with a business and soft skills component was Latin America and the Caribbean. Asia and the Pacific specifically accounted only for 10 per cent of the skills development interventions in the sample – 38 per cent included a soft skills component.

One third of the interventions that provided young people a certificate of participation was found in Africa, followed by North America and then Latin America and the Caribbean (22 and 20 per cent, respectively).

The review of outcomes measured across regions revealed the important presence of employment outcomes, with interventions in Asia and the Pacific reporting the largest share within a region. Yet, most cases with reported employment outcomes came from Europe and Central Asia (Figure 7). Earnings outcomes were particularly reported among evaluated interventions in Africa.

▶ Figure 7. Main outcome construct evaluated for each intervention, regional differences

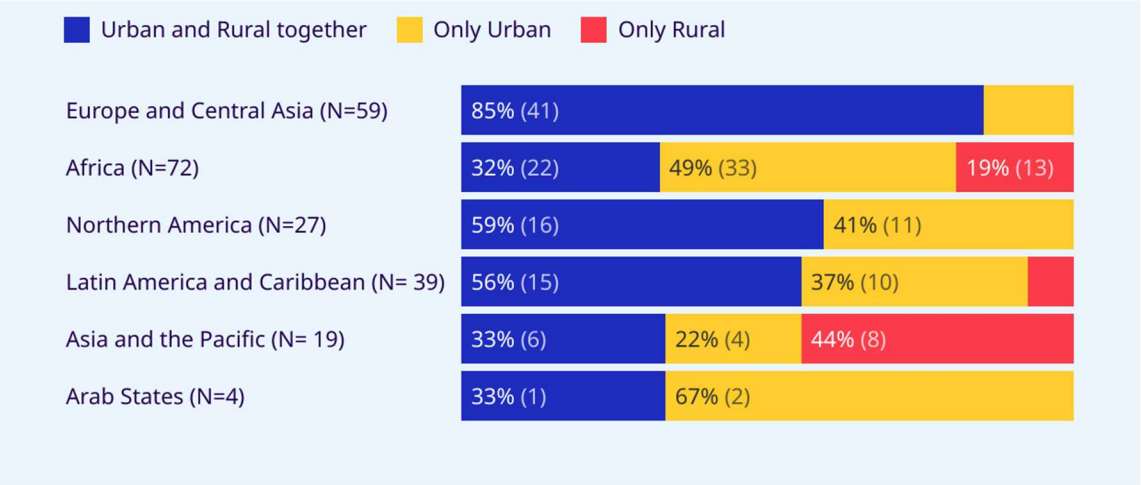


Note: The values in the bar segments represent the number of interventions, not the percentages.

Regional variation can be observed also when looking at programme financing. The interventions in Europe and Central Asia, as well as in Asia and the Pacific, were mainly financed entirely by public authorities (42 and 63 per cent, respectively). Yet, the interventions in Africa (41 per cent) and Latin America and the Caribbean (33 per cent) were mainly financed by non-government organizations or other national and international donors or by a combination of public and private sources.

Regarding targeting, most of the interventions aimed specifically at disadvantaged youth were in Africa (at 27 per cent), immediately followed by Latin America and the Caribbean (26 per cent). These interventions either targeted both rural and urban areas or only rural areas. Among the regions covered in this systematic review (Figure 8), the largest concentration of interventions targeting rural areas was found in Asia and the Pacific (at 44 per cent) and in Africa (at 19 per cent). Finally, Asia and the Pacific and North America had more interventions targeting youth younger than 25 years (at 84 per cent and 85 per cent, respectively).

▶ Figure 8. Urban versus rural focus, regional differences



Note: The values in the bar segments represent the number of interventions, not the percentages.

4.3 Programme costs

Cost considerations were an essential complement to the evidence on the effectiveness of interventions. Kluve et al. (2017) stressed their absence in impact evaluation studies as a factor limiting the practical influence of empirical evidence on programme design and policy decisions related to youth employment.

This systematic review collected data on the availability of information on intervention costs and the presence of cost–benefit analyses (which, for simplicity, also included cost-effectiveness analyses) in reports. Following the definitions in Fiala and Hempel (2011), cost–benefit analysis identified programme costs and weighted them against the (dollar) value of all programme benefits. The cost-effectiveness analysis, on the other hand, identified the full cost of a programme and related the cost to specific measures of outputs or outcomes. Examples included many of the evaluations of the Jovenes Programmes in Latin America.

In addition and comparing with Kluve et al. (2017), this systematic review expanded the coding exercise to integrate information on costs per beneficiary and the results of cost–benefit (and cost-effectiveness) analyses when available. The latter was distinguished between net benefit, net loss or neutral effect (cost equal to benefits). It also included a question on the time (in months) needed for an intervention to yield net benefits.

Figure 9 shows an increasing number of reports citing cost and presenting cost–benefit analyses. Up to 2012, there were typically three or fewer reports with cost information published each year. However, the number increased to approximately ten reports per year from 2013 up to 2021. Although much remains to be done to improve the reporting of costing and generate better evidence on the cost–benefit (effectiveness) and feasibility of youth employment interventions, the data suggest an encouraging trend.

▶ Figure 9. Increasing number of reports citing cost information, 1993–2021

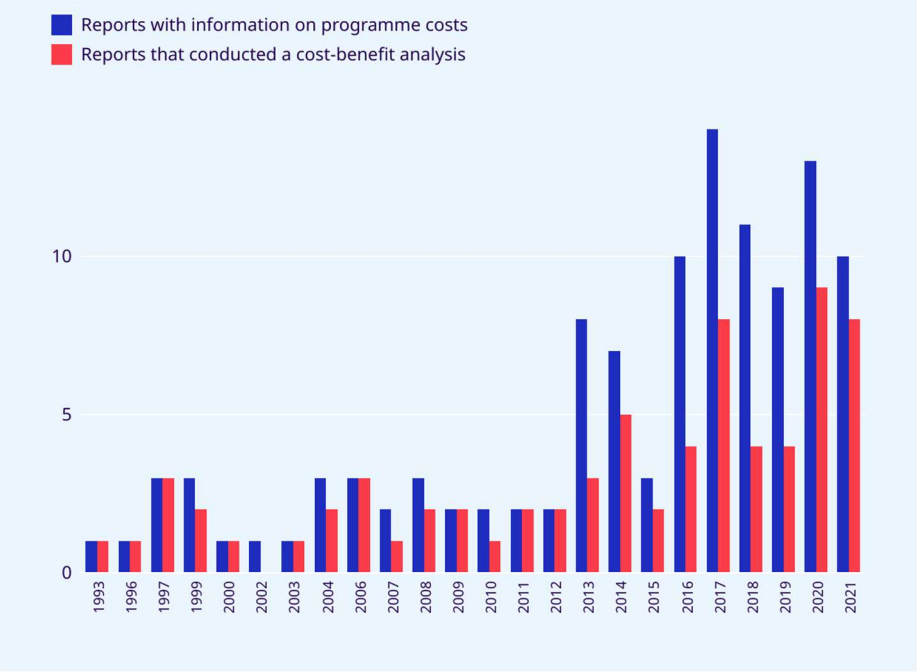


Table 7 reveals that among the 220 interventions analysed, 99 (45 per cent) provided cost information, 65 (30 per cent) presented cost–benefit analyses, and 79 (36 per cent) offered cost per beneficiary data. Most of the interventions with cost information came from middle- and high-income countries, while approximately 20 per cent of them were from low-income countries. However, a more nuanced picture emerged when examining the distribution of interventions by income level. Fewer interventions were identified in low-income countries, but they were more likely to report cost information, with a 56 per cent frequency compared with 42 per cent in high-income countries and 44 per cent in middle-income countries. The reporting of cost–benefit analysis and cost per beneficiary followed a similar pattern, but the likelihood of including this information is now more comparable across income levels.

In addition, the cost information sample aligned with the distribution of main intervention categories analysed in this review. As shown in Table 8, most of the interventions related to skills development (63 per cent for cost information, 65 per cent for cost–benefit analyses and 61 per cent for cost per beneficiary), followed by employment services (13 per cent, 15 per cent and 15 per cent, respectively), subsidized employment (13 per cent, 6 per cent and 13 per cent, respectively) and entrepreneurship promotion (11 per cent, 14 per cent and 11 per cent, respectively). Only one intervention in the public works category provided information on cost, cost–benefit and cost per beneficiary.

▶ Table 7. Cost information, cost–benefit analysis and unit cost across country income groups

	Country income level			
	High income	Middle income	Low income	Total
Total interventions	91	93	36	220
Cost				
No. of interventions with information on the cost	38	41	20	99
% of interventions with cost information by country income category	38.4	41.4	20.2	100
% of interventions with cost information as a share of total number of interventions	17.3	18.6	9.1	45
% of interventions with cost information within country income category	41.8	44.1	55.6	45
Cost–benefit analysis				
No. of interventions with information on cost and benefits	24	28	13	65
% of interventions with cost and benefits information by country income category	36.9	43.08	20	100
% of interventions with cost and benefits information as a share of total number of interventions	10.9	12.7	5.9	29.5
% of interventions with cost and benefits information within country income category	26.1	30.4	36.1	29.5
Cost per beneficiary				
No. of interventions with information on the cost and benefits	25	35	19	79
% of interventions with cost and benefits information by country income category	31.6	44.3	24.1	100
% of interventions with cost and benefits information as a share of total number of interventions	11.4	15.9	8.6	35.9
% of interventions with cost and benefits information within country income category	27.5	37.6	52.8	35.9

▶ Table 8. Cost information, cost–benefit analysis and unit cost across main categories of intervention

	Main category of intervention				
	Skills training	Entrepreneurship promotion	Employment services	Subsidized employment*	Total
Total interventions	126	31	32	31	220
Cost					
No. of interventions with information on the cost	62	11	13	13	99
% of interventions with cost information by country income category	62.3	11.1	13.1	13.1	100
% of interventions with cost information as a share of total number of interventions	28.1	5	5.9	5.9	45
% of interventions with cost information within country income category	49.2	35.9	40.6	42	45
Cost–benefit analysis					
No. of interventions with information on costs and benefits	42	9	10	4	65
% of interventions with cost and benefits information by main intervention category	64.6	13.9	15.4	6.2	100
% of interventions with cost and benefits information as a share of total number of interventions	19.1	4.1	4.6	1.8	29.5
% of interventions with cost and benefits information within main intervention category	33.3	29	31.3	12.9	29.5
Cost per beneficiary					
No. of interventions with information on the cost and benefits	48	9	12	10	79
% of interventions with cost and benefits information by country income category	60.8	11.4	15.2	12.7	100
% of interventions with cost and benefits information as a share of total number of interventions	21.8	4.1	5.5	4.5	35.9
% of interventions with cost and benefits information within country income category	38.1	29.0	37.5	32.3	35.9

Note: * = Including public works.

Table 8 shows that 79 per cent of the interventions reporting cost information provided further details on the cost per beneficiary. To facilitate cross-country comparisons of cost information, the GDP deflator index was applied to convert all cost values to 2020 US dollars.

Table 9 presents descriptive statistics on cost per beneficiary, stratified by country income level, the main category of intervention and the net impact of the intervention.

▶ Table 9. Cost per beneficiary of youth employment interventions covered in this review (in 2020 US$)

	No. of interventions	Mean	Median	Standard deviation	Minimum	Maximum
Aggregate sample	79	5 702.84	840.20	20 896.09	23.59	140 156.40
Country income level						
High-income countries	25	16 103.37	3 249.80	35 373.50	122.00	140 156.40
Middle-income countries	35	957.94	418.35	1 354.79	43.11	7 403.32
Low-income countries	19	758.53	489.87	685.04	23.59	2 620.09
Main category of intervention						
Skills training	48	3 102.82	715.94	6 832.90	39.65	31 017.63
Entrepreneurship promotion	9	1 667.31	743.92	2 272.07	237.56	7 403.32
Employment services	12	952.73	234.68	1 265.82	23.59	3 529.27
Subsidized employment*	10	27 515.04	1 701.76	54 063.34	77.98	140 156.40
Net impact						
Net benefit –yes	29	1 590.65	489.87	4 716.58	41.70	25 863.91
Net benefit – no	19	11 245.46	1 521.67	31 839.14	225.13	140 156.40
No cost–benefit analysis	31	6 152.63	1 148.43	21 685.64	23.59	119 072.40

Note: *= Including public works.

The cost-level data in this sample demonstrate significant variation, with an average cost per beneficiary of approximately $5,702.84 and a median cost of $840.20.[11] This large variation can be partially explained by the different methodologies used by various authors to calculate the cost per beneficiary. For instance, Egebark and Kaunitz (2018) and Saez, Schoefer and Seim (2021) arrived at their cost estimates by comparing the expected number of jobs generated by the programme to the foregone payroll taxes. Other authors used alternative approaches, resulting in challenges when comparing results across studies. To complement table 9, a list with the cost per beneficiary for each of the interventions evaluated is available in Annex A2.

After controlling for other intervention characteristics, exploratory analysis of the cost-level data conducted in this review revealed that the cost per beneficiary negatively correlated with the income level of the country, while a positive correlation was observed with the duration and scale of the programme. Skills training interventions were found to be relatively more costly when compared to other categories of interventions, but the preliminary analysis suggested that investing more in this type of intervention is likely to result in net benefits.

It is important to acknowledge that the descriptive analysis reflected in this section was based on the interventions identified and coded from the empirical evaluations covered in this systematic review. This sample may not be entirely representative and, therefore, may not fully reflect the variety in design of youth employment interventions worldwide or offer a sufficient representation across various geographical locations, income groups and other pertinent factors.

In particular, reporting of cost information for youth-targeted interventions became more consistent in the past decade. However, there were challenges in harmonizing this information to allow for cross-country and time

[11] Note that the presence of two outliers, both related to the Swedish employer-paid payroll tax cuts programme, had a strong impact on this variation. These two interventions reported the highest cost per beneficiary in the entire sample, at $140,156.40, compared to other interventions.

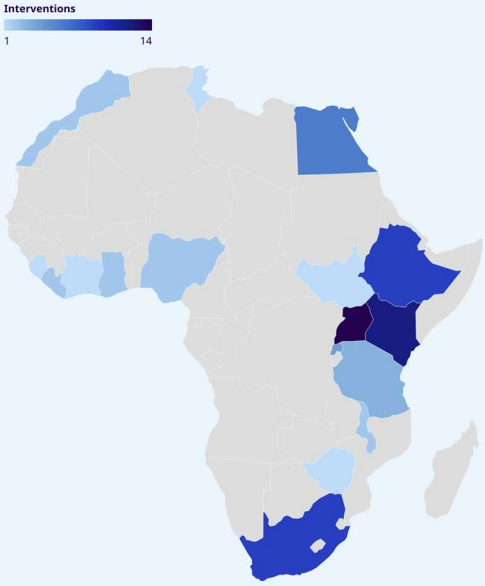
comparisons, especially due to the various types of interventions. A primary challenge was the different methodologies used by authors to report costs. For instance, some authors included "opportunity costs" in skills training interventions, while others classified beneficiaries differently, which directly impacted the reported cost per beneficiary information.

Thus, it is crucial to exercise caution when interpreting the results and applying them to a larger population. It is recommended that further research be conducted with a more diverse and representative sample to obtain a greater understanding of the design characteristics of youth employment interventions worldwide.

Box 2
Africa at a glance

As shown in Table 5, Africa accounted for 72 interventions, equivalent to 30 per cent of all interventions covered by the review. The increasing body of evidence sheds light on the design and implementation of youth active labour market programmes across the region. Most interventions (60 per cent) were located mainly in East and Southern Africa, followed by some West Africa countries. Rigorous empirical evidence on youth employment was absent in Central Africa, pointing to a relevant area for future research. While 45 per cent of the interventions in this systematic review related to skills development, entrepreneurship was a prominent intervention type in Africa. The interventions in Africa had a strong focus on the most vulnerable youth (at 43.1 per cent) and youth younger than 25 years (65.3 per cent).

▶ **Geographic distribution of youth employment interventions in Africa evaluated in this review**

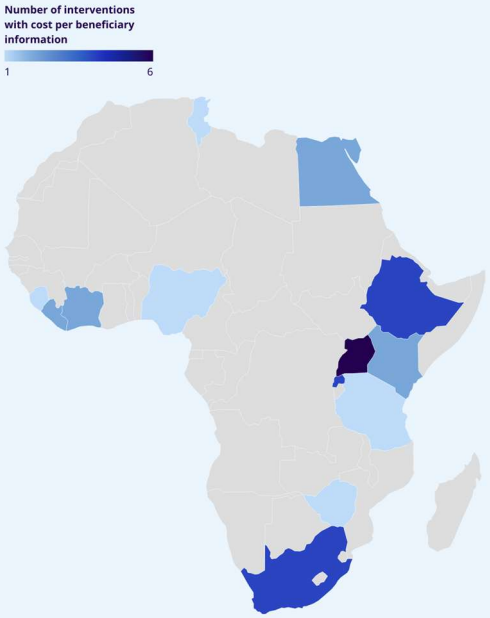


▶ **Characteristics of interventions from Africa**

	No. of interventions	%		No. of interventions	%
Scale of intervention			**Has intervention type**		
National	27	37.5	Skills training	36	50.0
Regional	21	29.2	Entrepreneurship promotion	32	44.4
Local	23	31.9	Employment services	19	26.4
Actors involved in the design			Wage subsidies	6	8.3
Private entity or NGO	32	44.4	Public works	1	1.4
Government	21	29.2	**Additional skill components**		
Both	19	26.4	Business skills	13	18.1
Actors involved in implementation			Soft skills	19	26.4
Private entity or NGO	35	48.6	Digital skills	5	6.9

Government	20	27.8	**Additional design elements**		
Both	14	19.4	Younger than 25 years	47	65.3
Main intervention category			Availability of certificate	16	22.2
Skills training	33	45.8	Target disadvantaged youth	31	43.1
Entrepreneurship promotion	22	30.6	Rural focus	13	18.1
Employment services	11	15.3	Urban focus	33	45.8
Wage subsidies	5	6.9	Rural and urban focus	26	36.1
Public works	1	1.4			

▶ Interventions with cost per beneficiary information in Africa



The following table presents cost per beneficiary information for 32 interventions in Africa, showing an average cost of $953.80 and a median cost of $609.45. Entrepreneurship promotion interventions had higher average and median costs, while employment services interventions had lower costs, although data are limited for this category. Four subsidized employment interventions had cost per beneficiary information in Africa, with the Employment Tax Incentive in South Africa having the lowest cost, at $852.60, and the PEJEDEC–Apprenticeship Scheme in Ivory Coast having the highest cost, at $1,836.47.

In terms of skills training interventions, this category had the most information on cost per beneficiary in Africa. The average and median costs are $652.49 and $488.51, respectively, with the NGO Landmine Action Programme in Liberia having the highest cost, at $1,629.79. Most interventions in Africa with cost per beneficiary information had a positive net impact and lower average and median costs compared to interventions without cost–benefit analysis or with a negative or neutral effect. One intervention with a net benefit was the BRAC Training Programme in Uganda, which had a national scale, medium/long duration and provided a soft skills component.

▶ Data on cost per beneficiary, Africa

	OBS	Mean	Median	Standard deviation	Minimum	Maximum
Africa sample	32	953.80	609.45	1 328.95	23.59	7 403.32
Country income level						
Middle-income countries	15	1 183.59	822.31	1 802.58	43.11	7 403.32
Low-income countries	17	751.05	489.87	696.80	23.59	2 620.09

Main category of intervention						
Skills training	16	649.53	484.10	489.56	39.65	1 629.79
Entrepreneurship promotion	8	1 846.03	783.12	2 360.35	394.39	7 403.32
Employment services	4	42.84	43.76	15.02	23.59	60.26
Subsidized employment*	4	1 297.43	1 250.33	413.78	852.60	1 836.47
Net impact						
Net benefit – yes	14	652.49	488.51	536.63	41.70	1 629.79
Net benefit – no	6	799.97	783.12	324.64	352.46	1 352.23
No cost–benefit analysis	12	1 382.26	578.74	2 064.04	23.59	7 403.32

Note: *= Including public works.

5 Effect size computation and meta-analysis methods

5.1 Standardizing effect size estimates

To compare estimated treatment effects across studies, the SMD was computed for both continuous outcome variables (such as income) and dichotomous outcome variables (such as employment probability) reported in the primary studies. Additionally, an indicator variable was coded if a treatment effect estimate was positive and statistically significant. The SMD captures the relative magnitude of the treatment effect in a way that is dimensionless and hence comparable across outcomes and studies. SMDs capture the ratio of the treatment effect for a specific outcome relative to the standard deviation of that outcome within the evaluation sample used to estimate the treatment effect. Following Hedges and Olkin (1985), adjustments for potential bias in studies based on small sample were made. The resulting SMD effect size is commonly referred to as "Hedges' g". Hedges' g and its standard error SE_g were computed as:

$$g = \frac{\bar{Y}_t - \bar{Y}_c}{S_p} * \left[1 - \frac{3}{4 * (n_t + n_c - 2) - 1}\right]$$

$$SE_g = \sqrt{\left[\frac{n_t+n_c}{n_c*n_t} + \frac{g^2}{2*(n_c+n_t)}\right]}.$$

where $\bar{Y}_t$ and $\bar{Y}_c$ are the mean outcome in the treatment group and comparison group, respectively. Similarly, n_t and n_c are the respective sample sizes. The numerator of g represents the causal raw impact of the programme on the outcome. Many studies reported either matching- or regression-based estimates of the treatment effect (even for randomized controlled trial-based designs).[12] Hence, the SMDs in most cases were computed using the formulae used by Waddington et al. (2012). In matching-based studies, $\bar{Y}_t - \bar{Y}_c$ is reflected by the average treatment effect on the treated (ATET). S_p is the pooled standard deviation of the outcome after treatment and was computed as:

$$S_p = \sqrt{\frac{(n_c - 1) * S_c^2 + (n_t - 1) * S_t^2}{n_t + n_c - 2}}$$

With S_t and S_c as the standard deviation in the treatment and comparison group, respectively. If either the comparison or treatment group's standard deviation was not reported, the standard deviation of the total sample S_T or the comparison group's standard deviation was used to compute g. In the case of dichotomous outcome variables, the S_t and S_c were computed based on the number of observations and the proportion in the respective group, if available.

For partial effect sizes estimated using multivariate analysis, g and its standard error were estimated based on the formula described in Keef and Roberts (2004):

$$g = \frac{\hat{\beta}}{\hat{\sigma}}$$

$$SE_g = \sqrt{\frac{g^2}{v-2} * \left(\frac{v}{t^2} + v * [c(v)]^2 - v + 2\right)} \quad \text{, where } \frac{1}{c(v)} = \sqrt{\frac{v}{2}} * \frac{\Gamma\left(\frac{v}{2} - \frac{1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}$$

where $\hat{\beta}$ refers to the coefficient of the treatment variable in the regression, $\hat{\sigma}$ is the pooled standard deviation of the outcome, *v* is *n-k* degrees of freedom and *Γ()* is the gamma function.[13]

There are two approaches for the calculation of the pooled standard deviation from regression-based studies. In Hedges' approach, $\hat{\sigma}$ is the standard deviation of the error term in the regression. Because this was rarely reported, the review team followed Cohen's approach and computed $\hat{\sigma}$ from the standard deviation of the dependent variable across all observations (S_T) (see Lipsey and Wilson 2001):

[12] This is in line with experiences documented by previous systematic reviews in related fields, such as Baird et al. (2013) or Tripney et al. (2013).
[13] For studies with large *n*, *c(v)* was considered equal to 1.

$$\hat{\sigma} = \sqrt{\frac{S_T * (n_t - 1) - (\beta^2 * (n_c * n_t)/(n_c + n_t))}{n_t - 1}}.$$

If information for calculating SE_g was not available, it was approximated by:

$$SE_g = \frac{g}{t}$$

where t is the *t*-value associated with a *t*-test on the treatment effect of a regression.

If none of the values for S_p, S_T or S_c could be obtained from the report (or by contacting the authors), the standard deviation of the outcome variable was approximated using the formula from Borenstein et al. (2009):

$$S_p = SE * \sqrt{\frac{n_c * n_t}{n_c + n_t}}$$

where SE is the standard error of a means test (such as regression coefficient). Because this formula is technically only correct for bivariate effect sizes, sensitivity analysis was performed on the sample without these imputations. For some studies, the review team transformed reported effect size statistics (often *t*-, *F*-, *p*- or *z*-values) prior to calculating effect sizes, following the procedures suggested in Lipsey and Wilson (2001).

Prior to synthesizing computed effect sizes, checks were made for outliers that could be a result of erroneous coding or misleading assumption in the computation of SMD. In cases in which the SMDs or their standard errors seemed implausibly large, the original reports were revisited to check whether they were in accordance with the findings stated by the authors. Because it was not possible to solve all outlier issues following this approach, the data were censored by removing outliers from the sample (trimming). In the main analysis, outliers with a Hedges' g below -1 or above 1 or an inverse standard error below 1 or above 100 were excluded, following common procedure in meta-analyses, such as Askarov and Doucouliagos (2020).

5.2 Dealing with dependent effect sizes

Meta-analyses commonly involve a multilevel, nested data structure – with effect size estimates at the bottom level and studies at the highest level. Specifically, a single *programme* can include more than one *intervention* that differ in their intervention design. Each of these interventions can combine several *components* (such as skills training and career counselling services). For example, a programme can include two interventions: one providing only skills training and another combining skills training with employment services. Interventions were therefore regarded as the systematic review's primary unit of interest. Each intervention could be delivered to multiple *cohorts* that represent mutually exclusive sets of beneficiaries (for example, subsequent intake years or separate regions). Each intervention may have been evaluated by more than one *study* that may have been published in multiple *reports* (working papers, technical reports or journal publications).

The coding and analysis took additional care to account for this multilevel data structure. Against this background, a *study* was defined as an evaluation of one intervention-cohort based on a specific dataset for a specific sample of beneficiary and comparison groups. Therefore, a study population (all individuals in the underlying dataset) could be a subsample of the intervention population. The study population might itself differ from the sample population for a specific treatment effect estimate on a specific outcome construct, such as in the case of gender subsample analysis. Two reports were treated as part of the same study if they were based on the same dataset and hence could not be treated as independent, even if they were written by different authors.

Conventional meta-analysis methods assume that all effect sizes in the sample are independent of each other. When effect sizes are not independent, conclusions based on these conventional procedures can be misleading or even wrong because the core assumption of most statistical techniques may be violated. Treatment effects reported in primary studies may be regarded as independent from each other when the underlying data derive from different sample populations. One challenge is that each report might present different treatment effect estimates for the same outcome construct and the same sample population. For example, a treatment effect might be reported in a study for the entire (pooled) sample and subsequently reported for subgroups of the same sample,

such as males and females.[14] In such instances, a multitude of treatment effects could be reported for the same group where there was no a priori reason to give preference to one measure over another.

Two approaches were followed to account for the multilevel data structure in the analysis, depending on the meta-analytical model. For the univariate random effects meta-analysis (to estimate summary effect sizes in section 6.1), an algorithm for preselection and aggregation of effect sizes was developed (**Error! Reference source not found.**). For the multivariate meta-analysis (to estimate correlates of effect size magnitude in section 6.3), all effect sizes coded from primary studies were included separately, and potential biases from effect size dependency were accounted for by clustering standard errors on the study level (following Stanley and Doucouliagos 2012).

Box 3
Preselection and aggregation of effect size estimates for univariate random effects meta-analysis

Conventional methods of meta-analysis, such as traditional random effects models, cannot easily deal with potential biases stemming from the multilevel data structure commonly encountered in social sciences. For these methods, it is important that only one effect size per outcome construct and study (as defined here) is included in the meta-analysis (Borenstein et al. 2009). Creating such construct-specific study-level effect size aggregates (at the intervention or study level or across different subgroups as part of the moderator analysis) required careful estimation to avoid the situation whereby a single group of participants influenced the summary effect size disproportionately. Specifically, prior to estimating summary of the effect size estimates, study-level effects were generated by combining estimates within each study into one effect size per subgroup. This approach typically involved two steps.

First, by identifying a set of effect sizes that derived from the same independent group of participants and then, where applicable, selecting the effect sizes for this group where it was possible to establish a preference (for example, keeping only pooled estimates and discarding subgroup estimates except when needed in the analysis). By dropping some of the effect sizes derived from the sample, this redundancy was removed from the analysis as far as possible.* This method provided a better approach to the data than averaging effect sizes across all overlapping subgroups.**

Second, in cases in which multiple effect sizes were reported for each independent group without clear justification for dropping some rather than others (such as where the same outcomes were reported at several points in time for the same group), the aggregate ("synthetic") effect sizes were estimated for each independent group. Based on the method for combining effect sizes from the same independent population suggested by Borenstein et al. (2009), the approach was as follows: Let g_{ij} and SE_{ij} be the i^{th} effect size, where $i = (1, \ldots, m)$ and its standard error, respectively, for the sample population identified by j. To arrive at a single combined (aggregate) effect size for group j, the review team took the simple average:

$$g_j = \frac{1}{m}\sum_{i=1}^{m} g_{ij}$$

and the standard error of g_j given by

$$SE_j = \sqrt{\left(\frac{1}{m}\right)^2 \left(\sum_{i=1}^{m} g_{ij}^2 + \sum_{i \neq k} \rho_{ikj}\sigma_{ij}\sigma_{kj}\right)}$$

where ρ_{ikj} is the correlation coefficient between g_{ij} and g_{kj} in study j.***
Hence, the independent group aggregates were assembled at the relevant unit of analysis, such as at the intervention or study level (depending on the assumed correlation addressed in the procedure). Then the random effects meta-analysis was applied to the aggregated data and estimated summary effect sizes.

Notes:
* = Here, redundancy indicates providing additional information about a group that is not needed for the desired level of aggregation. For example, if the goal is to create programme aggregates for all participants, then male and female subgroup estimates may be dropped. On the other hand, if the goal is to create an aggregate for females for each programme, then pooled estimates would be dropped.
** = For the purpose of brevity, the guidelines used to drop effect sizes within each group are not included here. This information is available upon request.
*** = The first best option is to attempt to estimate ρ_{ik} from the data. However, in cases where there was an insufficient number of observations, then some assumption about ρ_{ik} had to be made. Assuming that $\rho_{ik} = 0$ is likely to overestimate precision and assuming that

[14] No studies were encountered in the sample that assessed different treatments using the same group of individuals as the comparison group (multi-arm studies with pooled comparison).

$\rho_{ik} = 1$ is likely to underestimate precision, the more conservative assumption was adopted, that $\rho_{ik} = 1 \; \forall \; (i, j)$ where $i \neq k$.

5.3 Summarizing effect size estimates

Different meta-analysis models have been developed to summarize effect size estimates from primary studies. Each model makes different assumptions and estimates different parameters of interest. The classic common effect model assumes that all study effect sizes are the same and equal to the true effect size, which is often violated in practice.[15] A fixed-effects model assumes that the study effect sizes are different and "fixed" for each study. This implies that included studies define the entire population of interest, and inference concerns only the specific studies included in the meta-analysis. A random-effects model also assumes that the study effect sizes are different for each study but that they are “random”: Each study has its own population effect, and differences between effect sizes either arise due to sampling variation or because of differences between studies.

Because studies in the meta-regression sample varied widely in terms of intervention characteristics, geographic contexts, methods and outcomes, the random effects model was deemed the most adequate for analysis. The summary effect sizes based on these models are presented alongside the 95 per cent confidence intervals in the forest plots in Section 6. The number of effect size estimates and individual interventions for each subgroup are displayed in the respective forest plots to provide an indication of the size of the evidence base. This estimated univariate meta-analysis model is therefore:

$$g_{\mathrm{ij}} = \beta_0 + u_{\mathrm{j}} + \varepsilon_{\mathrm{ij}}$$

where g_{ij} is the i^{th} effect size estimate (SMD) from study j (for different cohorts or different outcome measures, for example). And uj ~ (0, τ_{j}^2) is the study-level error term, with τ_{j}^2 as the between-study variance in true effects, assumed equal across estimates within each study. β_0 represents the estimated summary of effect size in the respective sample of studies.

Heterogeneity tests were used to examine whether the variation in effect size estimates within outcome categories was larger than expected from sampling error alone (Deeks et al. 2001). To test for heterogeneity, the team employed I-squared statistics and Q-statistics. These statistics tested whether the percentage of variability in effect estimates was estimated due to heterogeneity rather than by chance. A significant Q (p-value <0.05) and an I-squared value of at least 50 per cent were considered to be indicators of heterogeneity.

Subsample analyses were performed to test hypotheses on whether variations in (average) effect sizes reported were associated with study, participant or intervention characteristics (moderator variables). This subsample provided a first – descriptive evidence on how effect size magnitudes differed across certain dimensions. Ideally, moderator analysis should be conducted with a minimum of ten studies for each individual moderator variable (Borenstein et al. 2009). The results from these subsample random effects meta-analyses are presented in Section 6.

5.4 Assessing correlates of effect size magnitude

To assess the drivers of effect size magnitudes in more detail, multivariate meta-regression models were estimated. This allowed the review team to assess which factors correlate with the magnitude of effect sizes while controlling for other potentially moderating factors, such as research design. In contrast to the random effects meta-analysis underlying the subsample analysis, these multivariate meta-regressions do not aggregate effect sizes within studies prior to analysis. This not only retains a larger amount of information but also enables an assessment of the drivers of effect-size magnitude that differ *within* studies (such as follow-up duration).

[15] The term “common effect” is used rather than "fixed effect" model because it is more descriptive of the underlying model assumption and less easy to confuse with the (plural) “fixed effects model”.

In the main analysis, an unrestricted weighted least squares (WLS) model was estimated, with weights as the inverse variance of each effect size estimate as recommended by Stanley and Doucouliagos (2012). The analysis accounts for effect size dependency by clustering the standard errors at the study level. The estimated WLS model thus took the following form:

$$g_{\mathrm{ij}} = \beta_0 + \sum \beta_k Z_{\mathrm{jk}} + \varepsilon_{\mathrm{ij}}$$

where Z_{ik} represents $\mathrm{K}\,(\mathrm{k} = 1, \ldots, \mathrm{K})$ moderator variables representing study-level or intervention-level heterogeneity. $\varepsilon_{\mathrm{ij}} \sim N(0, \sigma_{\mathrm{ij}}^2)$ is the error term and σ_{ij}^2 the standard error corresponding to effect size estimate g_{ij}.

Because unrestricted WLS models place more weight on larger studies, they may serve as a conservative lower-bound of summary effect size estimates. Moreover, the WLS model assumes that each estimate relates to a single true effect and thus are generally considered as a version of fixed effect models (Kaiser et al. 2022). To provide a better idea of correlates of effect size magnitude, results from the multivariate random effects meta-regression model are presented side by side with the WLS results in section 6.5. However, since the random-effects multivariate meta-regression involves all effect sizes separately and does not account for effect size dependency it may underestimate standard errors of coefficient estimates. The random effects meta-regression model was estimated using a restricted maximum likelihood method:

$$g_{\mathrm{ij}} = \beta_0 + \sum \beta_k Z_{\mathrm{jk}} + u_{\mathrm{j}} + \varepsilon_{\mathrm{ij}}$$

Finally, to assess whether correlates of effect size magnitude differ across some core dimensions, the review team conducted subsample analysis for the multivariate meta-regression. Two dimensions were considered as core moderators of effect size heterogeneity according to the theory of change outlined in Kluve et al. (2017): First, across the key two categories of outcome measures (employment or earnings outcomes) and, second, across the two country-income levels (high- or low- and middle-income countries).

5.5 Assessing and accounting for reporting bias

Selective reporting of findings (or “file drawer effect”) is a primary challenge for meta-analyses. Publication bias refers to the selection of results, for example, by authors or journal editors based on the direction of the estimated effect, its statistical significance or combination of both (Card, Kluve and Weber 2018). A common method to inspect reporting bias is a funnel plot. Funnel plots show the relationship between the effect size and the precision of the effect size estimate (the inverse standard error). In these plots, less precise estimates – plotted lower down the y axis – are typically scattered more widely around the true effect. In the absence of reporting bias, the standard error of an estimate should be orthogonal to the reported effect sizes, and the plot should be symmetric around the true effect size. The idea underlying funnel plot assessment to detect publication bias is that the magnitude of the reported estimate will depend on its standard error. In the presence of a positive reporting bias, one would expect a skew towards the right for less precise estimates (funnel asymmetry) (Stanley and Doucouliagos 2012).

The review also tested for reporting bias by means of regression, thereby following Stanley and Doucouliagos (2012). In a first step, the WLS meta-regression model was estimated with the estimates’ standard error SE_{ij} as explanatory variable. This is commonly called the "funnel asymmetry test", or FAT, to assess the presence of reporting bias (null hypothesis $\gamma_{FAT} = 0$). The WLS model thus becomes

$$g_{\mathrm{ij}} = \beta_{PET} + \gamma_{FAT} SE_{ij} + \varepsilon_{\mathrm{ij}}$$

In this regression, β_{PET} represents the “precision effect test", or PET, testing for the presence of a genuine effect beyond reporting bias (null hypothesis $\beta_{PET} = 0$). Again, following Stanley and Doucouliagos (2012), if the PET provides evidence for a genuine effect beyond publication bias, the precision-effect-estimate with standard error, or PEESE, model is estimated. The PEESE includes the estimates’ variance SE_{ij}^2 instead of the standard error SE_{ij} in the univariate robust variance estimation meta-regression model and is a more efficient estimate of the summary effect size in the presence of reporting bias:

$$g_{\mathrm{ij}} = \beta_{PET} + \gamma_{FAT} SE_{ij}^2 + \varepsilon_{\mathrm{ij}}$$

In the multivariate meta-regression, the potential reporting bias is assessed along other dimensions by including indicator variables for the evaluation design (such as experimental methods) and publication status (such as peer-reviewed reports).

5.6 Limitations of the study

To date, meta-analysis is the most widely accepted method to synthesize quantitative findings from a selected body of literature (McKenzie et al. 2022). As any statistical method, meta-analysis has some limitations that need to be taken into account. Most importantly, the quality of the results relies substantially on the underlying data – such as the quality of underlying primary studies. The three most common criticism regard (i) the presence of publication bias, (ii) quality of underlying studies and (iii) comparability of studies, interventions and measures. We discuss these challenges in turn and how they can be addressed by meta-analytic methods.

- *Publication bias*: Some authors argue that meta-analyses can exacerbate the consequences of p-hacking and publication bias in the underlying body of literature (Vosgerau et al. 2019). The study addresses this challenge in two ways: First, we conducted a systematic, rigorous search of all available literature. Second, we assessed and accounted for publication bias by using common methods used in meta-analysis (PET-PEESE). A large variety of approaches to correct for publication bias exists and researcher disagree on the most reliable method. Given the ongoing disagreement, recent guidance is to test the sensitivity of results against various methods to adjust for publication bias (Bartoš et al. 2023)
- *Quality of underlying studies*: The validity of meta-analysis results depends crucially on the methodological quality of underlying primary studies (such as their internal validity). Some researchers suggest to only select primary studies of vetted high-quality, such as randomized controlled trials (Simonsohn, Simmons and Nelson 2022). However, this may lead to a bias by indirectly restricting the sample also to high-quality interventions because these are more often accompanied by sophisticated evaluations. One alternative is to assess the quality of studies on a range of dimensions (risk-of-bias assessment) and assess the measure as part of the meta-analysis. However, such detailed risk-of-bias assessments are often challenging to implement and have their own limitations (Kluve et al. 2017). The updated review analysis included both experimental and quasi-experimental studies. To address differences in study quality, we tested for differences by study design in the multivariate meta-analysis framework (see section **Error! Reference source not found.** for results).
- *Heterogeneity ("averaging of incommensurable results")*: Meta-analysis requires primary studies that are similar enough to each other to make meaningful comparisons. This regards both the underlying interventions (treatments, sample populations, etc.) as well as evaluation designs (outcome measures, etc.). The review analysis addressed this by using statistical techniques to account for differences between studies. However, these techniques may not always be effective, especially if the studies differ in fundamental ways that cannot be accounted for statistically.
- *Effect size dependency*: Common meta-analysis methods assume that each observation in the sample (effect size) is independent of each other. This challenge and how this study managed it are discussed in depth in section 5.2. However, researchers disagree on the most appropriate method to address potential effect size dependency in a multi-level structure often observed in social sciences (Gucciardi, Lines and Ntoumanis 2022). Other methods than those proposed here (such as the clustering of standard errors) may be more appropriate in our setting and may need to be explored.

6 Results

6.1 Sample description

The search process identified 5,051 reported treatment effects. Note that an individual study may have contributed to multiple outcome categories and hence the aggregate number of studies may be larger than the number of studies in the sample. In other words, the same participants may have provided an estimate for earnings and employment outcomes.

Most treatment effect estimates were based on employment outcomes (59 per cent of treatment effects), as shown in Table 10, which includes the employment probability and other employment-related measures, such as hours worked or quality of employment. One third (35 per cent) of treatment effects related to earnings outcomes and 6 per cent of treatment effects were calculated based on business performance outcomes. Most treatment effects (at 53 per cent) stemmed from studies that reported on randomized controlled trials, followed by quasi-experimental studies (41 per cent).

In terms of the main intervention category, treatment effects largely linked to skills trainings (at 74 per cent), while entrepreneurship promotion programmes, employment services and wage subsidy interventions each represented between 7 per cent and 11 per cent of the treatment effects.

Table 10. Frequencies of coded treatment effects

	No. of treatment effects	%	No. of studies
Outcome category			
Employment	2 978	58.9	153
Earnings	1 795	35.2	79
Business performance	295	5.8	16
Selected outcome constructs			
Employment probability	2 023	40	107
Earnings	961	19.1	42
Hours worked	510	10.1	18
Quality of employment	209	4.1	10
Selected evaluation designs			
Experimental (randomized controlled trial)	2 659	52.6	116
Quasi-experimental	2 073	41.04	97
Main intervention category			
Skills training	3 734	73.9	151
Entrepreneurship promotion	361	7.1	31
Employment services	364	7.2	32
Subsidized employment	592	11.7	31
Total	**5 051**		

Note: (i) N = 5,051 treatment effect estimates. When depicting selected outcome constructs or evaluation designs, the percentages do not add to 100 because categories were not mutually exclusive. (ii) Two studies on public works were merged into the category "subsidized employment".

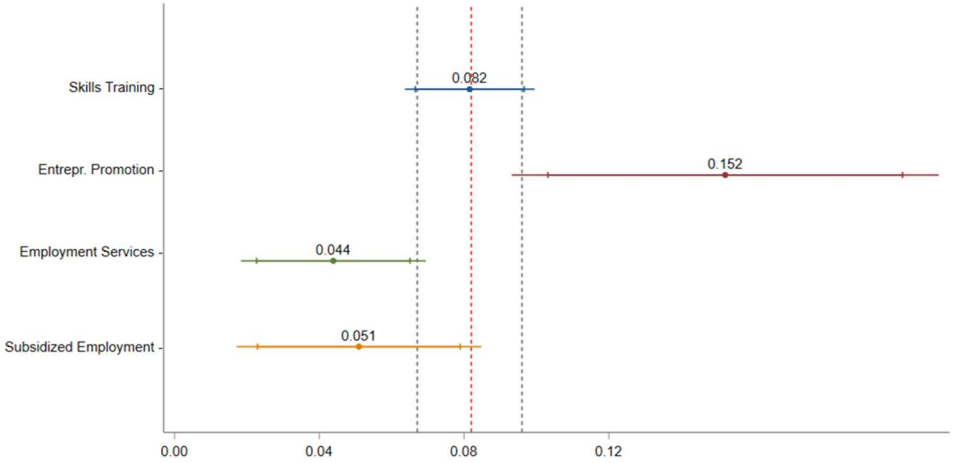
From the full sample of 5,051 coded treatment effect estimates, 4,242 SMD effect sizes could be computed using various imputation methods to address information missing in the primary study (see section 5.1). Subsequently, outliers were excluded to account for potential errors in coding or effect size computation. This reduced the sample from 4,242 SMDs to 4,060 effect size estimates. As expected, the censoring procedure resulted in slightly smaller average effect sizes and narrower confidence intervals. The robustness test of the results with respect to the censoring ranges indicated that the results held irrespectively. The subsequent analysis consisted of 210 reports, reflecting 227 independent studies and the evaluation of 204 interventions across 157 programmes.

6.2 Summary effect sizes

The results in this section relied on effect size estimates across studies based on the random-effects meta-analysis model. It should be considered as descriptive evidence of the underlying data and not as policy messages because the analysis did not account for publication bias or take into account other factors that may be collinear with the respective effect size moderator.[16]

Figure 10 shows results by **main intervention category**. The overall average effect of the youth employment programmes was positive, at 0.08 SMD, and statistically significant at the 1 per cent level (CI = 0.07 SMD, 0.09 SMD). The magnitude of this average effect size estimate is considered "medium" by contemporary benchmarks in social sciences. For instance, several recent meta-analyses of educational interventions globally (Kraft 2023 and 2020; Evans and Yuan 2022) established effect sizes below 0.05 SMD as small, those in the 0.05–0.20 SMD range as medium and those above 0.20 SMD as large. Moreover, the average effect sizes were substantially larger than the average effect size of 0.04 SMD reported in Kluve et al. (2019 and 2017). However, average effect sizes differed significantly between intervention types.

▶ **Figure 10. Average effect sizes across intervention category**



Note: CI at 90 per cent and 95 per cent. The analysis covered 227 independent samples (studies) based on the full sample of 4,060 effect sizes that remained after censoring (SMD limit = 1, inverse standard error limits = 1–100). Clustering excluded 1,909 effect sizes because a higher-level effect size was available (see section 5.2 for details). The red horizontal line represents the average effect in the full sample (CI = 0.07 SMD, 0.09 SMD). For skills training, 142 independent studies based on 1,884 effect sizes remained of 3,144. For entrepreneurship promotion, 30 independent studies based on 64 effect sizes remained of 352. For employment services, 31 independent studies based on 89 effect sizes remained of 294. For subsidized employment (combining wage subsidies and public works), 24 independent studies based on 114 effect sizes remained of 270.

- For programmes with **skills training** as the main category of intervention, the average effect size was 0.08 SMD (CI = 0.07 SMD, 0.010 SMD). This effect was slightly lower in magnitude than related studies, particularly Stöterau, Kemper and Ghisletta (2022), who reported 0.105 SMD, albeit using a different estimation procedure (robust variance estimation). But the impact was substantially larger than the 0.05 SMD for skills training in the predecessor study by Kluve et al. (2019, table 4).
- The average effect size of **entrepreneurship promotion** interventions was significantly larger, at 0.15 SMD, although with substantial heterogeneity (CI = 0.10 SMD, 0.21 SMD). The SMD was slightly larger than the 0.12 SMD reported in Kluve et al. (2019). Other meta-analyses of entrepreneurship-related interventions not restricted

[16] For example, observing a higher average effect size in low-income countries may be confounded by a large share of skills training programmes relative to other country income groups, rather than factors related to their specific context.

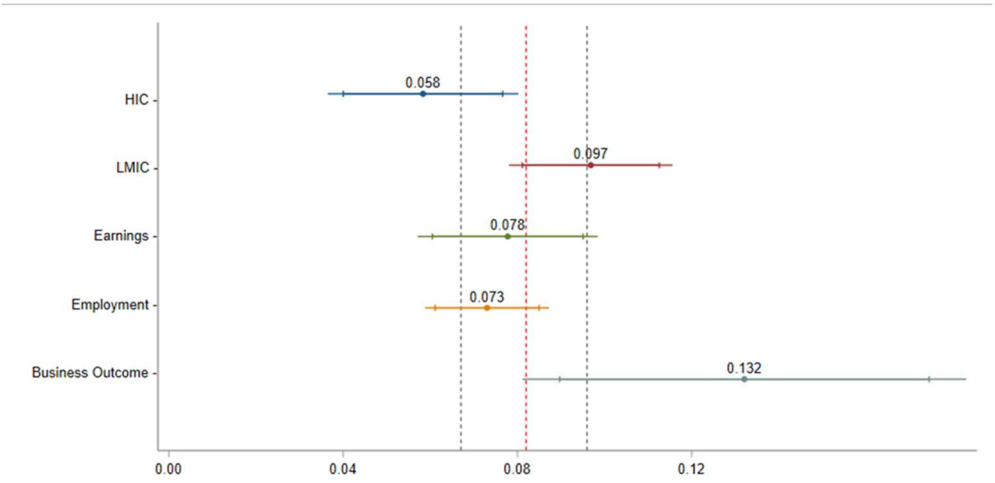
to youth described positive effects of varying magnitude.[17] However, as described below, this large average effect size was partially driven by impacts on business outcomes (see **Error! Reference source not found.**).

- The studies of **employment services** and **subsidized employment** interventions reported smaller effects on average, which were nonetheless statistically significant, with 0.04 SMD (CI = 0.02 SMD, 0.06 SMD) and 0.05 SMD (CI = 0.02 SMD, 0.08 SMD), respectively. These results reflected a marked increase from the findings in Kluve et al. (2019), which documented almost zero impact of employment services and subsidized employment interventions (SMD = 0 and 0.02, respectively).[18]

Figure 11 divides effect sizes across **country income groups** and **outcome categories**.[19] Again, for all subsamples, the average effect sizes were positive and statistically significant.

- Average effect sizes reported from interventions in the **low- and middle-income countries** at 0.097 SMD were substantially larger than for interventions in the **high-income countries**, at 0.058 SMD (CI = 0.04 SMD, 0.08 SMD).
- The overall effect of youth ALMPs on **earnings** and **employment** outcomes were similar, with 0.078 SMD and 0.073 SMD, respectively (compared to 0.05 SMD and 0.04 SMD in Kluve et al. (2019), table 4). The impacts of youth-targeted ALMPs on measures of **business performance** (such as profits, sales or business survival) were substantially larger, at 0.13 SMD, yet also exhibited much greater heterogeneity, with a 95 per cent CI ranging from 0.07 SMD to 0.17 SMD.

Figure 11. Average effect sizes across country income groups and outcomes



Note: CI at 90 per cent and 95 per cent. The analysis covered 227 independent samples (studies) based on the full sample of 4,060 effect sizes that remained after censoring (SMD limit = 1, inverse standard error limits = 1 to 100). Clustering excluded 1,909 effect sizes because a higher-level effect size was available (see section 5.2 for details). For high-income countries, 86 independent studies based on 765 effect sizes remained of 1,380. For low- and middle-income countries, 141 independent studies based on 1,386 effect sizes remained of 2,680. For earning outcomes, 151 independent studies based on 816 effect sizes remained of 1,382. For employment outcomes, 205 independent studies based on 1,255 effect sizes remained of 2,391. For business outcomes, 39 independent studies based on 80 effect sizes remained of 287.

Note that the majority of business performance outcome measures here relate to entrepreneurship interventions conducted in low- and middle-income countries. Nonetheless, the large magnitude and heterogeneity of effect

[17] For a sample of evaluations of business training interventions for beneficiaries of all ages, McKenzie (2020) obtained an average impact of 10 per cent on business profits (not directly comparable to SMDs), with a 95 per cent CI of 4–16 per cent and 5 per cent on sales (95 per cent CI of 0.2–9.2 per cent). An earlier meta-analysis by Grimm and Paffhausen (2015) reported small effects at best, also for a non-age specific sample. Other studies reviewing impacts from entrepreneurship interventions include Piza et al. (2016) and Valerio, Parton and Robb (2014).

[18] Kluve et al. (2019) conjectured that the small effects observed on average from employment services and subsidized employment interventions emerged due to within-programme or within-study heterogeneity (programmes were not well targeted or did not offer the right combination of interventions to address the challenges that youth experienced) or from poor implementation, even if the design was adequate.

[19] Note that studies often report effect size estimates for several outcome measures. The process described in section 5.2 ensured that each study only provided one effect size estimate per outcome construct. Yet, some studies may be presented in all three subsamples of outcome constructs.

sizes on business performance highlights that these outcomes are inherently different from those measured on employment or earnings of youth. Indeed, business outcomes could be considered an intermediate outcome on the pathway to changes in employment and earnings: Only if young entrepreneurs can generate sales and profit from their business will there be changes in their earnings or ability to remain self-employed in the longer term.

At the same time, the relationship between business outcome and youth employment or earnings is not always linear or direct. Successful business performance does not necessarily guarantee immediate improvements in earnings. There may be time lags or other factors that influence the translation of business success into tangible earnings benefits for the individuals targeted by an intervention. Additionally, some interventions may prioritize the creation of sustainable and scalable businesses over immediate profit and earnings for youth, further complicating the interpretation of business outcomes in relation to direct employment or earnings impacts.

These reasons thus suggest that impacts on business outcomes are likely very distinct from impacts on employment or earnings and hence should be treated separately. Therefore, to not confound the main results, effect size estimates on business performance outcomes are excluded from the remainder of the empirical analysis discussion that follows this section. However, the analysis retained the sample of effect sizes from the studies that reported the impacts of entrepreneurship promotion interventions on employment and earnings of youth. In effect, this reduced the full sample from 4,060 to 3,773 effect size estimates and the number of independent studies from 227 to 217. The number of interventions and studies that reported impacts on business outcomes unfortunately were not sufficient to conduct a separate empirical meta-analysis. **Error! Reference source not found.** therefore qualitatively reviews the evidence on the impacts of (entrepreneurship promotion) programmes on business performance outcomes.

Box 4
Impacts of active labour market programmes on youth business performance outcomes

Enterprise promotion interventions are critical for empowering youth to establish and sustain a business. This box summarizes the design features and business outcomes from 20 reports evaluating such interventions across different contexts.

Common design features

The analysed enterprise promotion interventions exhibited diverse designs, each incorporating distinct elements aimed at fostering entrepreneurship and enhancing business outcomes. One design feature was capital support, which had a central role in promoting business creation and profitability. The interventions often provided financial resources in various forms, including cash grants, in-kind grants and subsidized loans. These resources aimed to alleviate capital constraints and thus better enable beneficiaries to start or expand a business. For instance, in some cases, the provision of an in-kind grant proved particularly effective because it directed recipients' investments towards productive assets (Crépon, El Komi and Osman 2020; Blattman et al. 2016). Similarly, business plan competitions, which offered small financial prizes, were instrumental in fostering entrepreneurship by reducing barriers to entry and addressing credit constraints, thereby increasing the likelihood of self-employment (Fafchamps and Quinn 2016).

Training and mentorship also emerged as critical components of effective enterprise promotion programmes. Many interventions paired training sessions with mentorship opportunities after recognizing that skill development alone is insufficient to guarantee success. This combination proved more effective than standalone training, particularly for enhancing profitability and business sustainability (Brooks, Donovan and Johnson 2018). Localized mentorship, wherein experienced entrepreneurs guided less-experienced counterparts within the same community or sector, had a particularly strong impact. By providing tailored, market-specific knowledge, mentors significantly boosted profitability, whereas formal, classroom-style business training often yielded limited improvements in business performance (Mariani et al. 2019).

Integrated approaches that simultaneously addressed multiple barriers, such as skill gaps, capital constraints and market access, demonstrated the most substantial and sustained improvements in business outcomes. Multifaceted programmes combined financial support with technical training, market linkages and advisory services to create comprehensive ecosystems for entrepreneurial success. These interventions enabled beneficiaries to overcome a broader range of challenges and achieve long-term business viability (Brudevold-Newman et al. 2017).

Another notable feature of effective enterprise promotion interventions was their focus on targeted populations. Many programmes were tailored to address the specific needs of marginalized groups, including women, youth and individuals in post-conflict settings. Such targeted approaches acknowledged the unique barriers these populations face, such as gender-specific constraints, limited access to networks or the challenge of rebuilding a livelihood in a fragile context. Tailored interventions not only enhanced the effectiveness of support but also contributed to greater inclusivity and equity in entrepreneurship (Blattman et al. 2020).

Main findings on business outcomes

Capital support was consistently associated with increased business creation, even in fragile settings, because it enabled participants to overcome financial barriers to establish an enterprise (Blattman et al. 2020; Jaramillo and Parodi, 2003). Business plan competitions further demonstrated the power of financial incentives, increasing self-employment rates by up to 33 percentage points among winners (Fafchamps and Quinn, 2016). Profitability and growth were also notable outcomes of well-designed interventions. Programmes that integrated grants with tailored support mechanisms achieved considerable profit increases, ranging from 20 per cent to 40 per cent (Crépon et al. 2020; Brooks et al. 2018). Localized mentorship contributed to operational efficiency and profitability by equipping entrepreneurs with practical, context-specific insights (Brooks, Donovan and Johnson 2018).

However, the sustainability of these outcomes varied across programmes. Although some interventions demonstrated long-lasting benefits, others experienced diminished impact within one to two years, often due to such challenges as inadequate savings mechanisms or market constraints (Brudevold-Newman et al. 2017). Gender-specific impacts revealed that women typically benefited tremendously from these programmes, gaining access to resources and opportunities that enabled them to succeed. Nonetheless, unique challenges, including societal norms and limited scalability, persisted as barriers to achieving their full potential (Blattman et al. 2016).

The analysis revealed that the most effective enterprise promotion interventions combined financial, human capital and market-based support that resulted in robust business outcomes. Programmes tailored to address specific barriers that youth and marginalized groups experience had greater impact because they considered the unique challenges confronting these populations. Moreover, sustained engagement through follow-up mentorship and advisory support was critical to ensuring long-term business success.

To maximize the effectiveness of such interventions, policies should adopt a comprehensive approach that integrates skill development, access to capital and mentorship. Improving programme targeting and enhancing accessibility for underrepresented groups, such as women and rural youth, is vital to expanding reach and equity. Additionally, ongoing monitoring and adaptive strategies are necessary to maintain and scale up the positive impacts over time.

Overall, the findings underscored the importance of holistic and well-targeted enterprise promotion interventions in fostering sustainable youth entrepreneurship. Such efforts are essential not only for individual empowerment but also for contributing to broader economic development and societal progress.

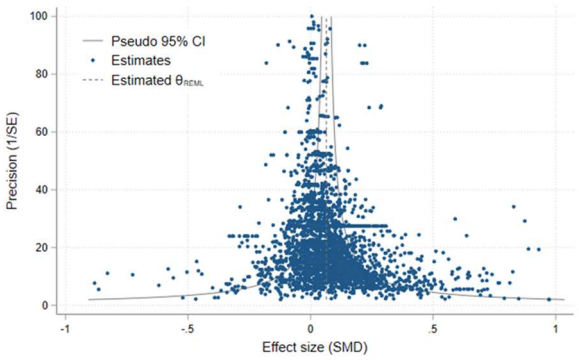
6.3 Publication bias

Figure 12 illustrates the funnel plots used to assess for publication bias in the pooled sample of effect size estimates, after excluding business outcome measures.[20] The dashed line refers to the unweighted mean effect size from the random effects model. For reference, the two figures include the boundaries of the 5 per cent statistical significance level: Panel A displays the funnel plot with disaggregated effect sizes and panel B with pre-aggregated effect sizes at the study level. The results were similar in quality and direction. As expected, the more precise estimates generally centred more closely around the estimated true mean effect size. The distribution of reported effect size estimates was fairly symmetrical, with the cloud of estimates being slightly skewed towards the lower right area of the plot. This suggests modest positive reporting bias, mostly due to low-precision studies reporting a positive effect of interventions.

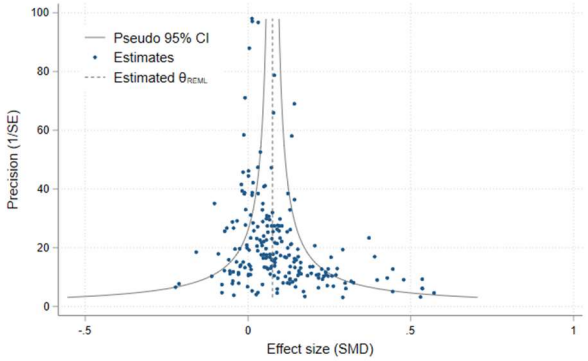
[20] See Annexes for the funnel plots that include business outcome measures.

▶ **Figure 12. Funnel plots**

A. Fully disaggregated effect sizes



B. Aggregated effect sizes



Note: Both plots exclude effect sizes measured for business outcomes. Plot A: For the fully disaggregated effect size plot, the full sample was 3,773 (of 3,944) non-missing effect sizes remaining after censoring (SMD limit = 1, inverse standard error limits = 1–100). Plot B: For the aggregated effect size plot, 2,071 of the 3,773 effect size estimates were selected (1,702 effect sizes were excluded due to a higher level available) and then aggregated to 804 groups and 217 independent clusters (by study ID level).

Building on the graphical evidence, two common approaches to test for reporting bias were implemented: first, an estimation of Egger's test for small sample analysis. This test confirms failure to reject the null hypothesis of no small-sample effects (theta = 0.7, z = 13.31, probability >|z|= 0.0000). Second, a test for publication bias followed the rationale of Stanley and Doucouliagos (2012), which includes the standard error (PET) or the variance (PEESE) as a predictor in an unrestricted WLS model estimate using inverse variance weights.

Even without accounting for publication bias, the summary effect size estimated using the WLS procedure in Table 11 was smaller than the random effects estimates shown in section 6.2 (0.075 average effect from WLS, Table 11, column 1, versus 0.08 SMD in Figure 10).[21] The FAT–PET test on the presence of publication bias in the pooled sample (Table 11, column 2) suggested there was publication bias but gave little indication of a genuine effect beyond the reporting bias. The PEESE model therefore accounted for reporting bias using the standard error (Table 11, column 3). This further reduced the summary effect size, compared to estimates from the model that did not account for publication bias (from 0.075 standardized difference to 0.055 standardized difference in the pooled sample). These results already excluded business outcomes. Annex **Error! Reference source not found.** shows the test for publication bias, including business outcome measures.

▶ **Table 11. Test for publication bias**

	(1) No publication bias adjustment	(2) FAT–PET test	(3) PEESE model
Average effect	0.075	0.014	0.055
	0.000	0.241	0.000
Standard error		1.120	
		0.000	
Variance			4.833
			0.000
Estimates	3773	3773	3773
Interventions	194	194	194
Reports	164	166	165
Studies	217	217	217

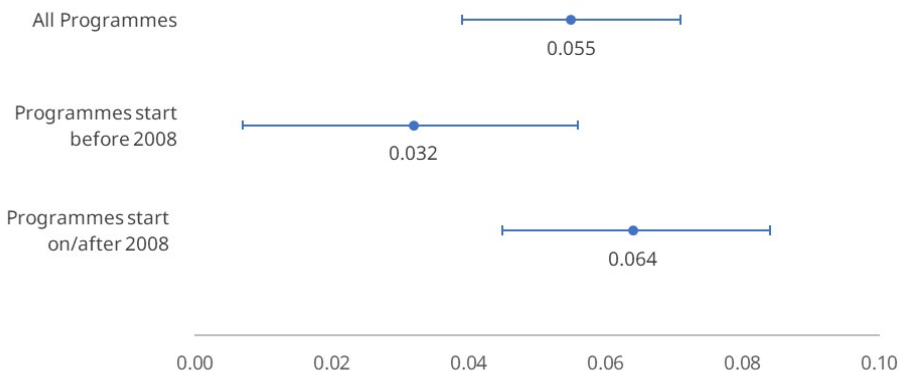
Note: The model estimated using the random effects model (REML). The sample excluded business performance outcomes, censoring SMD at +/- 1 and inverse standard error at 1–100, with Rho = 0.8 and publication bias: clustered standard errors at study ID level.

[21] As outlined in section 5.4, the reason is that WLS models place more weight on larger studies and assume that each estimate relates to a single true effect. Hence, they should be considered a conservative lower bound.

6.4 Subsample meta-regression results

This section explores variations in effect sizes that were associated with moderator variables after correcting by publication bias. On average, ALMPs improved employment and earnings of young participants by 0.055 SMDs over those of the study's comparison group. Figure 13 shows that the impact increased considerably over time. To illustrate this, the analysis split the total sample to compare programmes that started before and on or after 2008. The results suggested that, on average, impacts reported for newer programmes were twice as large as those of earlier programmes (0.032 SMD versus 0.064 SMD).

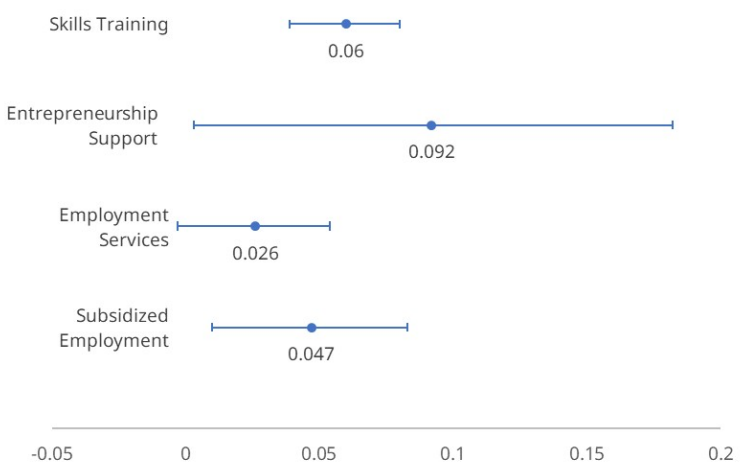
▶ Figure 13. Average effect sizes across subsamples of programmes that started before and on or after 2008



Note: The figure shows estimated standardized mean difference (Hedges' g) and 95 per cent confidence interval. The analysis used the random-effects model (REML) to account for publication bias and used the PEESE model proposed by Stanley and Doucouliagos (2012). The sample excluded business outcomes. Effect sizes were censored at g = +/-1 SMD and above an inverse standard error of 1–100.

Across intervention types (Figure 14), entrepreneurship support reported larger impacts on average, at 0.092 SMD. However, the impacts varied widely, with the 95 per cent CI almost spanning the zero line. Skills training programmes had slightly smaller impact on average, but these were clearly positive and statistically significant at the 1 per cent level. Employment services and subsidized employment programmes that supported youth in their job search or subsidized employment in the public or private sector had smaller impact on average (at 0.026 SMD and 0.047 SMD, respectively).

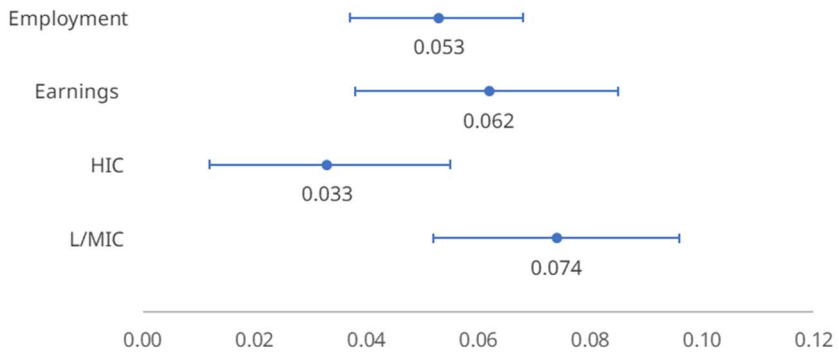
▶ Figure 14. Average effect sizes across intervention types



Note: The figure shows estimated standardized mean difference (Hedges' g) and 95 per cent CI. The analysis used the random-effects model (REML) to account for publication bias and used the PEESE model proposed by Stanley and Doucouliagos (2012). The sample excluded business outcomes. Effect sizes were censored at g = +/-1 SMD and above an inverse standard error of 1–100.

The impacts of ALMPs on youth employment and earnings appeared of similar magnitude (at 0.053 SD versus 0.062 SD in Figure 15). The subsample analysis across country income groups found greater impact in the low- and middle- income countries than in the high-income countries, on average.

▶ Figure 15. Average effect sizes across outcomes of interest and country income groups



Note: The figure shows estimated standardized mean difference (Hedges' g) and 95 per cent CI. The analysis used the random-effects model (REML) to account for publication bias and used the PEESE model proposed by Stanley and Doucouliagos (2012). The sample excluded business outcomes. Effect sizes were censored at g = +/-1 SMD and above an inverse standard error of 1–100.

6.5 Multivariate meta-regression results

The multivariate meta-regression introduces study- and intervention-level characteristics as covariates to better understand the factors influencing the effectiveness of youth labour market interventions. The meta-analysis regresses the SMDs (see Hedges' g in Section 5) on a set of explanatory variables selected based on data availability and the theory of change outlined in Kluve et al. (2017). To structure the discussion, the covariates were clustered into seven groups:

i. **Intervention category**: This group included a single variable for the main intervention category. For interventions combining multiple categories, the main category was determined by the review team based on the largest and predominant intervention type within a programme (see Section 2).

ii. **Characteristics of the study**: This group included variables indicating whether the underlying report was peer reviewed; whether the evaluation design reflected ITT, ATT, ATET or LATE estimates; whether the evaluation design was experimental (randomized controlled trial); and whether the follow-up assessment occurred at least one year after the baseline.

iii. **Outcome measures**: This group included a variable depicting whether the effect was measured as an employment or earnings outcome. It also included an indicator for whether follow-up data were collected more than one year after youth exited the programme. And the group included a variable denoting whether the outcome was measured conditionally or unconditionally on other labour market outcomes. For example, earning outcomes may have been measured only for those in employment (conditional) or for all individuals regardless of employment status (unconditional). Unconditional earnings included zero values for those not in employment, while conditional measures excluded these individuals from the sample, typically resulting in higher average incomes. Conditional earnings often represented such measures as wages rather than overall incomes.

iv. **Programme context**: The group included one indicator of whether the programme was implemented in a high-income country or in a low- or middle-income country. It also included a variable for whether the programme began before or after 2008.

v. **Sample characteristics**: This group captured whether studies that conducted subgroup analyses found different effects for specific socio-economic groups (such as female versus male). It also included studies of programmes that, by design, targeted specific socio-economic groups and hence the sample only included these groups. The variables captured effect sizes measured only for disadvantaged youth (constructed as youth from low-income households or with low educational attainment); female-only or male-only subsamples; and youth younger than 25 or those aged 25 or older.

vi. **Intervention characteristics**. This group included variables that indicated whether the intervention combined various intervention types; provided soft skills or incentives for programme participation (such as childcare, catering, transport, allowances or stipends or a salary); implemented compliance mechanisms or participant monitoring; offered certification upon programme participation or completion; and the duration of the intervention.

vii. **Programme characteristics**: The group included variables that indicated whether the programme was implemented at the national, regional or local level and whether public or private actors[22] were involved in implementation.

The analysis tested five model specifications using both WLS regressions on the Hedges' g and random effects models. The specifications increasingly added explanatory variables:

- **Specification I**: Main category of intervention
- **Specification II**: *Specification I*, plus study characteristics, country income group and follow-up duration.
- **Specification III:** *Specification II*, plus outcome measure characteristics, programme start date and sample characteristics.
- **Specification IV:** *Specification III*, plus intervention characteristics.
- **Specification V:** *Specification IV*, plus additional programme characteristics.

Results are presented for the entire sample (the "pooled sample" in Table 12) as well as for the subsamples measuring only employment or only earnings outcomes (Table 13). The analysis also disaggregated results by country income groups, with separate models for the high-income countries and the low- and middle-income countries (

[22] Public sector included government agencies and multilateral organizations, and the private sector included companies and civil society organizations.

Table 14).

The tables show that the number of observations reduced as more covariates were introduced due to missing information in some evaluations regarding programme design and participant characteristics. For instance, for the pooled sample model shown in Table 12, the number of effect size estimates dropped from 3,773 in specification I to 3,110 in specification V (a reduction of 18 per cent).

The meta-regression results suggested that several characteristics of interventions correlated with the reported effect sizes. Both the "what" (the main category of intervention) and the "how" (such as design and implementation features) appeared to determine an intervention's effectiveness. These relationships varied between the high-income and the low- and middle-income countries. Notably, the high-income country sample lacked a sufficient number of entrepreneurship interventions that measured impacts on labour market outcomes beyond business outcomes: Of the 31 entrepreneurship-focused interventions, only three were implemented in high-income countries.[23] As a result, the pooled estimates should be interpreted with caution. The respective subsections here on intervention characteristics therefore discuss the results from the high- and low- and middle-income countries separately (see

Table 14).

When interpreting the relevance of different correlates of effect size magnitude, it is important to keep in mind that they may be correlated with one another and may reflect the overall quality of an intervention. For example, programmes offering certification might have required validated curricula and/or maintained a rigorous level of quality control. Certification may also have been linked to "payment by results" models, whereby training providers were compensated (for instance, by the government) based on participant outcomes.

6.5.1 Intervention category (the "what")

In the high-income countries, skills training appeared to slightly outperform other intervention types. This was particularly evident when compared with the subsidized employment interventions – wage subsidies and public works programmes, which appeared less effective than skills training.

In the low- and middle-income countries, entrepreneurship promotion interventions demonstrated larger effect sizes than skills training. This difference may have stemmed from the focus of entrepreneurship interventions on self-employment outcomes, which often materialize more immediately as a direct result of programme participation. This result introduced a challenge when comparing their effectiveness with other intervention types, such as skills training, whose employment outcomes may take longer to emerge. Nonetheless, the finding was statistically significant and consistent with previous research. Moreover, the evidence suggested that employment services interventions yielded greater effect sizes than skills training when controlling for other factors influencing effect size magnitude (specification IV). In contrast, the evidence on subsidized employment interventions showed smaller unconditional (specifications I and II) effect sizes compared to skills training.

6.5.2 Study characteristics

The studies based on experimental evaluation design (randomized controlled trials) reported smaller effect sizes that were independent of outcome measure or country income group. This finding pointed to the greater validity of results from studies using these methods. It was also consistent with findings from most other meta-analyses across a range of fields.

Other study characteristics did not show consistent association with effect size magnitude. The peer-reviewed publications did not systematically report larger effect sizes than the non-peer-reviewed studies. Interestingly, the intention-to-treat (ITT) estimates were of similar magnitude to the ATET estimates across specifications and subgroup analyses. Because the latter adjusted impact estimates for programme non-compliance (such as drop-outs), they usually yielded larger effects.

[23] The subsample analysis for the high-income countries therefore also excluded entrepreneurship promotion programmes.

6.5.3 Outcome measure

The pooled regression analysis found similar effect sizes of ALMPs on employment and earnings outcomes (see Table 12) that were consistent with the univariate analysis in section 6.2. However, in the low- and middle-income countries, effect sizes for earnings outcomes were significantly larger than those for employment outcomes. In contrast, the employment outcomes in the high-income countries showed slightly larger effect sizes than the earnings outcomes, although these differences were not consistent across all specifications and were not statistically significant at the 5 per cent level (see

Table 14, specification IV, random-effects model).

Evaluations measuring impact one year or more after the end of the intervention revealed distinct patterns across country income groups. In the high-income countries, the effect sizes increased considerably over time, consistent with findings from previous studies, such as Stöterau, Kemper and Ghisletta (2022); Kluve et al. (2019); and Card, Kluve and Weber (2018).[24] This suggested a pronounced temporal pattern in which impacts grew after programme completion. However,

Table 14 indicates that this trend did not hold in the low- and middle-income countries, where effects sizes reported a year after programme completion were smaller than those captured immediately after programme exit.

In the high-income countries, outcomes measured conditionally on another labour market indicator showed considerably smaller effect sizes, with coefficients implying a reduction of between 0.04 SMD and 0.05 SMD. Because the multivariate model simultaneously, accounted for the outcome construct, this coefficient could be interpreted as a difference in average effect sizes between overall earnings and (hourly) wages conditional on being employed.[25] This finding suggested that the effects of ALMPs on youth earnings in high-income countries were primarily driven by the first-stage impact on employment probability rather than by the increases in wages or job quality once a young person was employed. In contrast, ALMPs in the low- and middle-income countries appeared to affect both employment and (conditional) earnings similarly. This was in line with the observation that the effect on employment (relative to earnings) was larger in the high-income countries than in the low- and middle-income countries (see Table 13).

6.5.4 Programme context

The multivariate regression analysis confirmed that ALMPs in the low- and middle-income countries reported larger effect sizes than in the high-income countries, with the difference ranging from 0.012 SMD to 0.04 SMD (see Table 12). This finding was consistent with the previous research by Stöterau, Kemper and Ghisletta (2022); Kluve et al. (2019); and Betcherman et al. (2007). Notably, the difference persisted after accounting for a broad set of study and intervention characteristics. However, there was some evidence that the differences may have stemmed from variations in programme settings between the high- and the low- and middle-income countries because the coefficient estimates were close to zero in the most comprehensive specification (specification V, WLS). One possible explanation was that programmes in the high-income countries were more often implemented at a national scale, which may have presented greater challenges in implementation and may have involved more heterogeneous participant groups.

The subsample analysis for earnings outcomes provided clear evidence of the stronger effect of youth-targeted ALMPs in the low- and middle-income countries, in comparison to the high-income countries. In this case, the effect sizes were both large and statistically significant, with a coefficient ranging from 0.04 SMD to 0.07 SMD (see Table 13). For employment outcomes, the differences were smaller (below 0.02 SMD) and were not consistently statistically significant across specifications.

The sample that was split between programmes that started before 2008 and in or after 2008 showed that, on average, the impacts reported for newer programmes were larger than those of earlier ones. This result was

[24] The Card, Kluve and Weber (2018) sample mostly included ALMPs not targeting youth.

[25] In some cases, this also reflected the difference between employment probability and job quality (conditional on being employed). However, the sample size for this comparison was rather small and hence the coefficient more likely reflected the difference between conditional and unconditional earnings.

particularly evident in the pool sample (see Table 12), the employment outcomes subsample (see Table 13) and the high-income countries subsample (see

Table 14).

6.5.5 Sample characteristics

The characteristics of the participants can also influence the effectiveness of ALMPs. In the analysis, some groups of young people appeared better placed to benefit from the interventions or, put differently, some ALMPs seemed to be designed in ways that offered greater opportunity to certain types of young people than to others. This subsection thus examines whether ALMPs had different effects depending on the characteristics of the population for which the effect was estimated.

The estimates related to the socio-economic background of youth showed that, on average, ALMPs had larger effect on the earnings of youth from disadvantaged backgrounds. This effect was largely driven by the interventions in the low- and middle-income countries and was consistent with findings from the previous studies, including Kluve et al. (2019).

Across gender, the pooled models (see Table 12) showed large and statistically significant positive effects for samples that included only young women, compared with those that combined both sexes. This pattern was particularly strong in the subsample of earnings outcomes and in the low- and middle-income countries. In the latter, the estimates capturing results only for young men were considerably lower than those for the combined-gender sample. There was no evidence of gender-differentiated effects among programmes in the high-income countries. This result indicates that ALMPs can help address the specific barriers young women face in labour markets in developing countries.

Last, there was some evidence that effects were larger for participants younger than 25 years than for those aged 25 or older (see Table 12). This result held across country income groups (see

Table 14), even though the respective subgroup samples were likely too small to retain statistical significance across all specifications.

6.5.6 Intervention characteristics (the “how”)

Two intervention characteristics showed similar patterns of effectiveness across country income groups. One, the inclusion of soft skills training in programme content was associated with larger effect sizes, particularly when additional factors were accounted for (specifications IV and V) and in relation to earnings outcomes. Although the previous analysis by Kluve et al. (2019) did not find a significant difference between programmes with and without soft skills components, the updated dataset in this study suggests that integrating soft skills into training curricula can enhance youth labour market outcomes.[26]

Two, interventions that provided certification of participation or completion, or otherwise validated the skills acquired by participants, also exhibited larger effect sizes. This result was especially pronounced for earnings outcomes, where the estimated effects were roughly twice as large as those for employment outcomes (Table 13). These findings support the argument by Stöterau, Kemper and Ghisletta (2022) that ALMPs help jobseekers signal their competencies and employability to potential employers.

In the low- and middle-income countries, there was strong evidence to suggest that comprehensive, multipronged interventions lead to larger positive effects on youth labour market outcomes. In particular, programmes that combined multiple intervention types (skills training and employment services) were associated with significantly higher effect sizes. Additionally, longer-duration interventions (at least four months) tended to lead to better outcomes for youth, while interventions offering in-kind or monetary incentives to participants were associated with smaller effect sizes. This negative relationship may partly reflect the targeting criteria of participants: Programmes that provided additional incentives for participants might reach youth facing more substantial labour

[26] The result does not represent an estimate for programmes that exclusively focused on soft skills.

market constraints. Further research is needed to disentangle the incentive package and clarify the underlying mechanisms, including how the targeting and selection of youth beneficiaries influences outcomes.

In the high-income countries, by contrast, the comprehensive interventions did not seem to be more effective than those focused on a single intervention type. Moreover, longer-duration programmes (eight months or more) were associated with smaller effects than shorter interventions (four months or less).

6.5.7 Programme characteristics

- **Geographical scope:** The evidence indicated that interventions implemented at the local or regional level tended to have greater impact on youth labour market outcomes than those at the national level. This pattern held consistently across the high-, low- and middle-income countries. A possible explanation is that local and regional interventions may be more responsive to specific labour market conditions, better aligned with the needs of youth and logistically easier to deliver and monitor.
- **Implementing actors:** In the high-income countries, programmes designed or implemented solely by public sector actors showed greater effectiveness than those involving public–private collaboration or led exclusively by private sector or civil society organizations.[27] In contrast, in the low- and middle-income countries, public–private partnerships tended to yield stronger impacts on youth labour market outcomes than interventions implemented by either the public or private sector alone.

▶ Table 12. Meta-regression results, full sample

	Pooled sample									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V
SMD sampling variance	5.838***	4.310***	4.301***	4.101***	4.085***	3.606***	3.509***	3.558***	3.295***	3.582***
	0	0	0	0	0	0	0	0	0	0
Mainly skills training	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Mainly entrepreneurship promotion	0.002	0	0.024	0.03	0.026	0	0.008	0.025*	0.039***	0.037
	0.939	0.996	0.264	0.164	0.47	0.992	0.498	0.076	0.006	0.105
Mainly employment services	-0.02	-0.01	-0.014	-0.006	0.003	-0.015**	0.001	0	0.015*	0.016
	0.165	0.39	0.212	0.602	0.827	0.035	0.911	0.982	0.094	0.206
Mainly subsidized employment	-0.02	-0.024	-0.043***	-0.030**	-0.022	-0.035***	-0.027***	-0.029***	-0.015	-0.015
	0.268	0.2	0.006	0.047	0.173	0	0.002	0.004	0.153	0.246
Publication not peer reviewed		*ref.*	*ref.*	*ref.*	*ref.*		ref.	ref.	ref.	ref.
Peer-reviewed publication		-0.004	-0.004	0.002	0.004		-0.004	0	0.001	0.007
		0.643	0.633	0.878	0.718		0.406	0.95	0.753	0.199
Non-intention-to-treat estimate		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Intention-to-treat estimate		-0.001	0.004	0.002	0.004		-0.001	-0.007	-0.010**	-0.006
		0.923	0.562	0.674	0.321		0.879	0.195	0.045	0.25
Non-experimental design (IV, regression discontinuity design)		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Experimental (randomized controlled trial)		-0.023**	-0.027***	-0.032***	-0.033***		-0.030***	-0.037***	-0.040***	-0.040***
		0.025	0.003	0	0.004		0	0	0	0

[27] Public sector included government agencies and multilateral organizations, and private sector included companies and civil society organizations.

	Pooled sample									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V
Measure before one year after exit from the programme		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Measured over one year after exit from programme		0.006	0.001	0.002	0.001		-0.006	-0.006	-0.004	-0.007
		0.677	0.944	0.894	0.947		0.153	0.197	0.332	0.142
Low- and middle-income country		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
High-income country		-0.039***	-0.029***	-0.028**	-0.013		-0.031***	-0.024***	-0.035***	-0.012*
		0	0.003	0.036	0.4		0	0	0	0.069
Earnings outcome			ref.	ref.	ref.			ref.	ref.	ref.
Employment outcome			-0.002	-0.002	-0.006			-0.001	-0.002	-0.007
			0.669	0.712	0.407			0.821	0.674	0.122
Unconditional outcome			ref.	ref.	ref.			ref.	ref.	ref.
Conditional outcome			0.008	0.006	0			-0.005	-0.004	-0.007
			0.321	0.481	0.975			0.391	0.49	0.207
Programme started before 2008			ref.	ref.	ref.			ref.	ref.	ref.
Programme started in or after 2008			0.012	0.013	0.023			0.029***	0.025***	0.037***
			0.184	0.177	0.109			0	0	0
Pooled sample			ref.	ref.	ref.			ref.	ref.	ref.
Sample of disadvantaged youth only			0.008	0.015	0.006			0.012*	0.01	0.011
			0.452	0.243	0.66			0.089	0.146	0.199
Both male and female participants			ref.	ref.	ref.			ref.	ref.	ref.
Sample of male participants only			-0.001	0.001	-0.002			-0.006	-0.006	-0.011*
			0.93	0.95	0.844			0.278	0.247	0.07
Sample of female participants only			0.011*	0.012**	0.011*			0.016***	0.015***	0.016***
			0.098	0.04	0.08			0.001	0.002	0.003
Participants aged 25 or older			ref.	ref.	ref.			ref.	ref.	ref.
Sample of participants younger than 25			0.013**	0.01	0.002			0.011**	0.010*	0.005
			0.042	0.172	0.847			0.041	0.057	0.428
Intervention without extra services				ref.	ref.				ref.	ref.
With extra services				-0.015	-0.007				-0.010*	-0.009
				0.21	0.574				0.072	0.17
Intervention without soft skills component				ref.	ref.				ref.	ref.
With soft skills component				0.011	0.020*				0.024***	0.034***
				0.158	0.051				0	0
Intervention did not provide incentives for participation				ref.	ref.				ref.	ref.
Participation incentives				-0.004	0.001				-0.006	-0.005
				0.675	0.884				0.264	0.389
Intervention did not monitor participation				ref.	ref.				ref.	ref.

	Pooled sample									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V
Monitoring of participants				0.002	-0.002				0.006	0.002
				0.884	0.827				0.275	0.773
Intervention did not provide certification				ref.	ref.				ref.	ref.
Certification				0.017*	0.026***				0.027***	0.032***
				0.063	0.004				0	0
Intervention shorter than 4 months					ref.					ref.
4–8 months					0.015					0.026***
					0.292					0
Intervention longer than 8 months					-0.007					-0.007
					0.585					0.346
Intervention with regional or local scale					ref.					ref.
National scale					-0.027**					-0.018***
					0.025					0.005
Implemented by private and public combined					ref.					ref.
Public only					-0.011					-0.009
					0.416					0.232
Private only					-0.013					-0.006
					0.392					0.378
Constant	0.040***	0.077***	0.055***	0.049***	0.060***	0.056***	0.089***	0.063***	0.060***	0.049***
	0.005	0	0.001	0.007	0.003	0	0	0	0	0
Estimates	3 773	3 650	3 436	3 436	3 110	3 773	3 650	3 436	3 436	3 110
Reports	194	173	153	153	119	194	173	153	153	119
Interventions	203	180	161	161	128	203	180	161	161	128
Adjusted R-squared	0.041	0.115	0.153	0.16	0.175					

Note: *= p<0.10, **= p<0.05, ***= p<0.01.

▶ Table 13. Meta-regression results, subsample outcome category

	Subsample: Employment outcomes										Subsample: Earnings outcomes									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
SMDs sampling variance	5.900***	5.407***	5.711***	5.598***	5.487***	4.046***	4.093***	4.154***	3.947***	4.115***	5.819**	1.478	1.056	0.589	0.773	2.377***	1.756***	1.860***	1.312**	1.592**
	0	0	0	0	0	0	0	0	0	0	0.045	0.315	0.505	0.679	0.643	0	0.002	0.002	0.029	0.017
Mainly skills training	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Mainly entrepreneurship promotion	0.011	0.015	0.029	0.036	0.038	0.019	0.03	0.035	0.047**	0.063	-0.006	-0.02	0.027	0.026	0.014	-0.011	-0.011	0.028*	0.044**	0.025
	0.521	0.372	0.159	0.111	0.459	0.297	0.124	0.113	0.039	0.1	0.85	0.548	0.488	0.519	0.732	0.433	0.443	0.096	0.011	0.318
Mainly employment services	-0.019	-0.011	-0.017	-0.013	0.003	-0.01	0.002	-0.002	0.004	0.02	-0.023	-0.011	-0.015	-0.004	-0.042	-0.030**	-0.01	-0.008	0.011	-0.03
	0.144	0.402	0.153	0.35	0.833	0.247	0.871	0.866	0.726	0.241	0.312	0.421	0.305	0.858	0.151	0.026	0.431	0.542	0.45	0.135
Mainly subsidized employment	-0.012	-0.007	-0.032*	-0.031*	-0.021	-0.018**	-0.009	-0.015	-0.012	-0.012	-0.058***	-0.070***	-0.072**	-0.022	0.003	-0.093***	-0.078***	-0.060***	-0.025	-0.007
	0.474	0.685	0.058	0.076	0.273	0.042	0.413	0.255	0.361	0.505	0.001	0.005	0.013	0.573	0.957	0	0	0	0.123	0.721
Publication not peer reviewed		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Peer-reviewed publication		0.003	0	-0.001	0.003		0.002	0.002	0.002	0.008		-0.011	-0.003	0.014	0.019		-0.009	-0.003	0.004	0.01
		0.757	0.98	0.925	0.798		0.734	0.721	0.788	0.277		0.451	0.827	0.371	0.294		0.162	0.631	0.495	0.169
Non-intention-to-treat estimate		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Intention-to-treat estimate		-0.005	0.002	0	0.005		-0.006	-0.012*	-0.014**	-0.006		0.012	0.005	0.003	-0.002		0.011	0.003	-0.003	-0.009
		0.656	0.799	0.952	0.267		0.361	0.086	0.042	0.437		0.164	0.406	0.59	0.802		0.117	0.638	0.664	0.176
Non-experimental design (IV, regression discontinuity design)		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Experimental (randomized control trial)		-0.014	-0.018*	-0.020*	-0.013		-0.020***	-0.024***	-0.025***	-0.019**		-0.036	-0.042	-0.048***	-0.067***		-0.042***	-0.051***	-0.056***	-0.067***
		0.219	0.053	0.059	0.139		0.003	0.001	0.001	0.039		0.152	0.116	0.037	0.01		0	0	0	0
Measure before one year after		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.

	Subsample: Employment outcomes										Subsample: Earnings outcomes									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
exit from the programme																				
	-0.002	-0.01	-0.009	-0.01		-0.006	-0.008	-0.007	-0.009		0.016	0.018	0.019	0.022		-0.004	-0.001	0	-0.001	
Measured over one year after exit from programme																				
	0.868	0.279	0.35	0.333		0.254	0.159	0.268	0.159		0.483	0.452	0.439	0.435		0.519	0.908	0.975	0.869	
Low- and middle-income country	ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.	
	-0.023**	-0.012	-0.005	0.01		-0.016***	-0.011*	-0.015**	0.006		-0.063***	-0.051***	-0.071***	-0.057		-0.053***	-0.044***	-0.065***	-0.039***	
High-income country																				
	0.022	0.281	0.757	0.486		0.006	0.065	0.048	0.468		0	0.007	0.001	0.118		0	0	0	0.001	
Unconditional outcome		ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.	
		0.015**	0.012	0.006			0.003	0.003	0.002			-0.002	-0.001	-0.004			-0.013**	-0.007	-0.008	
Conditional outcome																				
		0.038	0.112	0.442			0.772	0.705	0.858			0.878	0.927	0.678			0.039	0.234	0.183	
Programme started before 2008		ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.	
		0.017*	0.016	0.034**			0.039***	0.033***	0.047***			0.01	0.01	0.005			0.012	0.009	0.01	
Programme started in or after 2008																				
		0.071	0.143	0.04			0	0	0			0.371	0.402	0.797			0.106	0.259	0.317	
Pooled sample		ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.	
		0.004	0.005	-0.01			0.002	-0.001	-0.006			0.018	0.030***	0.048**			0.029***	0.032***	0.037***	
Sample of disadvantaged youth only																				
		0.785	0.725	0.449			0.837	0.947	0.622			0.226	0.038	0.033			0.005	0.002	0.003	
Both male and female participants		ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.	
		-0.001	-0.002	-0.005			-0.006	-0.008	-0.012			-0.004	0.002	-0.002			-0.007	-0.003	-0.006	
Sample of male participants only																				
		0.933	0.903	0.656			0.437	0.301	0.14			0.691	0.901	0.901			0.38	0.66	0.427	
Sample of female participants only		0.008	0.008	0.007			0.013*	0.011*	0.011			0.015*	0.023***	0.024**			0.020***	0.024***	0.029***	
		0.249	0.281	0.326			0.051	0.091	0.122			0.073	0.037	0.04			0.005	0	0	
Participants aged 25 or older		ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.	
		0.007	0.004	-0.007			0.006	0.005	-0.01			0.022**	0.013	0.006			0.011*	0.01	0.012	
Sample of participants younger than 25																				
		0.336	0.594	0.394			0.39	0.531	0.281			0.021	0.157	0.523			0.1	0.201	0.153	

	Subsample: Employment outcomes										Subsample: Earnings outcomes									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
Intervention without extra services				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
With extra services				-0.013	-0.002				-0.003	0.004				-0.013	-0.032				-0.017*	-0.034***
				0.3	0.847				0.713	0.686				0.422	0.151				0.057	0.001
Intervention without soft skills component				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
With soft skills component				0.004	0.01				0.006	0.01				0.019	0.021				0.042***	0.052***
				0.698	0.37				0.469	0.295				0.336	0.274				0	0
Intervention did not provide incentives for participation				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
Participation incentives				0.01	0.017				0.011	0.015*				-0.034	-0.041				-0.032***	-0.037***
				0.324	0.114				0.154	0.088				0.179	0.136				0	0
Intervention did not monitor participation				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
Monitoring of participants				-0.009	-0.009				-0.001	-0.006				0.028	0.023				0.024***	0.019**
				0.534	0.52				0.905	0.466				0.122	0.223				0.002	0.017
Intervention did not provide certification				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
Certification				0.004	0.008				0.013*	0.017**				0.053***	0.070***				0.055***	0.064***
				0.721	0.46				0.085	0.047				0.018	0.004				0	0
Intervention shorter than 4 months					ref.					ref.					ref.					ref.
4–8 months					0.021					0.031***					-0.03					-0.01
					0.199					0.001					0.412					0.367
Intervention longer than 8 months					-0.009					-0.013					-0.002					-0.005
					0.453					0.22					0.943					0.648
Intervention with regional or local scale					ref.					ref.					ref.					ref.
National scale					-0.030**					-0.016*					-0.026					-0.024***
					0.031					0.068					0.128					0.01
Implemented by private					ref.					ref.					ref.					ref.

	Subsample: Employment outcomes										Subsample: Earnings outcomes									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
and public combined																				
Public only					-0.029					-0.026**					0.008					0.007
					0.103					0.013					0.513					0.439
Private only					-0.030**					-0.024**					0.028					0.023**
					0.049					0.011					0.145					0.024
Constant	0.039***	0.061***	0.047***	0.034*	0.035*	0.053***	0.074***	0.059***	0.058***	0.053***	0.041**	0.101***	0.060**	0.052**	0.077**	0.061***	0.109***	0.071***	0.061***	0.070***
	0	0	0.004	0.057	0.052	0	0	0	0	0.001	0.035	0	0.018	0.011	0.034	0	0	0	0	0
Estimates	2391	2296	2146	2146	1925	2391	2296	2146	2146	1925	1382	1354	1290	1290	1185	1382	1354	1290	1290	1185
Reports	182	162	142	142	114	182	162	142	142	114	133	123	108	108	85	133	123	108	108	85
Interventions	185	163	145	145	116	185	163	145	145	116	140	129	114	114	93	140	129	114	114	93
Adjusted R-squared	0.042	0.062	0.098	0.105	0.135						0.055	0.275	0.308	0.342	0.377					

Note: *= p<0.10, **= p<0.05, ***= p<0.01.

▶ Table 14. Meta-regression results, subsample country income level

	Subsample: High-income countries										Subsample: Low- and middle-income countries									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
SMD sampling variance	6.332***	6.096***	5.610***	4.869***	4.296**	4.803***	4.116***	4.337***	4.192***	4.628***	3.246***	4.113***	4.924***	4.829***	4.747***	2.823***	3.414***	3.873***	3.629***	3.905***
	0.001	0.001	0.004	0.007	0.03	0	0	0	0	0	0.001	0	0	0	0	0	0	0	0	0
Mainly skills training	ref.	ref.	ref.	ref.	ref.						ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Mainly entrepreneurship promotion	0	0	0	0	0						-0.011	0.004	0.033	0.076**	0.045	0.002	0.014	0.039***	0.079***	0.063***
	.	.	.	.	.						0.598	0.852	0.151	0.013	0.273	0.874	0.227	0.003	0	0.003
Mainly employment services	-0.005	-0.001	-0.021	-0.015	-0.016	0.001	0.009	-0.002	0.009	-0.025	-0.019	-0.004	-0.001	0.03	0.022	-0.015	0.001	-0.004	0.030**	0.022
	0.612	0.927	0.113	0.357	0.543	0.949	0.442	0.864	0.575	0.337	0.164	0.825	0.968	0.282	0.429	0.135	0.934	0.738	0.034	0.182
Mainly subsidized employment	0.005	-0.015	-0.060***	-0.039**	-0.023	-0.008	-0.025*	-0.058***	-0.039**	-0.025	-0.055***	-0.048**	-0.007	0.012	-0.008	-0.049**	-0.027*	0.002	0.017	-0.02

	Subsample: High-income countries										Subsample: Low- and middle-income countries									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
	0.779	0.529	0.001	0.018	0.418	0.472	0.052	0	0.021	0.357	0.003	0.039	0.804	0.683	0.776	0	0.056	0.913	0.27	0.274
Publication not peer reviewed		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Peer-reviewed publication		-0.014	-0.01	0.003	0.005		-0.021***	-0.016**	-0.003	-0.003		0.01	0.015	-0.002	0.005		0.011**	0.017***	0.002	0.007
		0.124	0.164	0.742	0.759		0.005	0.038	0.734	0.815		0.457	0.264	0.93	0.769		0.035	0.002	0.77	0.344
Non-intention-to-treat estimate		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Intention-to-treat estimate		-0.005	0.009	0.008*	0.005		-0.01	0.002	0	0		0.005	0.004	0.005	0.02		0	-0.013*	-0.011	-0.003
		0.749	0.11	0.085	0.279		0.21	0.844	0.993	0.963		0.686	0.741	0.684	0.114		0.977	0.064	0.105	0.702
Non-experimental design (IV, regression discontinuity design)		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Experimental (randomized control trial)		-0.018	-0.009	-0.007	-0.011		-0.026***	-0.019*	-0.026**	-0.017		-0.037***	-0.044***	-0.041**	-0.044**		-0.036***	-0.040***	-0.034***	-0.032***
		0.172	0.382	0.521	0.509		0.004	0.069	0.032	0.34		0.007	0.002	0.01	0.015		0	0	0	0
Measure before one year after exit from the programme		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Measured over one year after exit from programme		0.031	0.024	0.029	0.037		0.017**	0.015*	0.023***	0.026**		-0.015*	-0.017*	-0.014	-0.020*		-0.016***	-0.016***	-0.015***	-0.023***
		0.142	0.275	0.149	0.15		0.047	0.082	0.008	0.014		0.077	0.062	0.145	0.072		0.001	0.001	0.002	0
Earnings outcome			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.
Employment outcome			0.01	0.012	0.009			0.015**	0.014*	0.01			-0.026***	-0.028***	-0.030***			-0.018***	-0.020***	-0.023***
			0.294	0.129	0.163			0.046	0.061	0.265			0.001	0	0			0	0	0
Unconditional outcome			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.
Conditional outcome			-0.019	-0.016	-0.028			-0.049***	-0.040***	-0.051***			0.011	0.011	0.01			0.005	0.007	0.006
			0.357	0.46	0.302			0	0.001	0.001			0.169	0.158	0.109			0.342	0.159	0.26
Programme started before 2008			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.
Programme started in or after 2008			0.022**	0.020**	0.007			0.038***	0.031***	0.003			0.008	-0.027	0.017			0.030***	-0.01	0.025**
			0.041	0.035	0.65			0	0.002	0.852			0.609	0.139	0.499			0	0.261	0.026
Pooled sample			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.
Sample of disadvantaged youth only			-0.01	0.017	0.01			-0.015	-0.012	-0.046*			0.029	0.029	0.034*			0.026***	0.033***	0.045***

	Subsample: High-income countries										Subsample: Low- and middle-income countries									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
			0.34	0.299	0.663			0.246	0.441	0.058			0.16	0.119	0.088			0.002	0	0
Both male and female participants			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.
Sample of male participants only			0.015	0.017	0.019			0.012	0.017	0.014			-0.020**	-0.023**	-0.023**			-0.023***	-0.027***	-0.029***
			0.31	0.184	0.234			0.246	0.111	0.251			0.035	0.014	0.011			0	0	0
Sample of female participants only			-0.005	-0.002	-0.007			-0.008	0.001	-0.002			0.016**	0.015**	0.019***			0.024***	0.020***	0.022***
			0.689	0.88	0.527			0.434	0.928	0.895			0.038	0.034	0.003			0	0	0
Participants aged 25 or older			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.
Sample of participants younger than 25			0.014**	0.009*	0.003			0.006	0	-0.003			0.004	0.015	0.006			0.013*	0.015**	0.009
			0.023	0.088	0.62			0.497	0.966	0.763			0.833	0.315	0.688			0.059	0.047	0.299
Intervention without extra services				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
With extra services				0.035**	-0.024				0.030***	-0.016				0.046**	0.025				0.044***	0.027**
				0.025	0.294				0.004	0.388				0.027	0.291				0	0.011
Intervention without soft skills component				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
With soft skills component				-0.004	0.018				0.033***	0.063***				0.013	0.011				0.014*	0.023***
				0.667	0.438				0.003	0				0.434	0.524				0.057	0.004
Intervention did not provide incentives for participation				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
Participation incentives				0.003	-0.001				-0.004	-0.005				0.006	0.023				0.007	0.015*
				0.845	0.931				0.729	0.752				0.697	0.13				0.268	0.074
Intervention did not monitor participation				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
Monitoring of participants				0.007	0.005				0.008	-0.018				-0.01	-0.006				-0.006	-0.001
				0.626	0.826				0.44	0.207				0.469	0.688				0.41	0.952
Intervention did not provide certification				ref.	ref.				ref.	ref.				ref.	ref.				ref.	ref.
Certification				0.012	0.021				0.018*	0.026*				0.023*	0.026*				0.035***	0.035***
				0.329	0.133				0.095	0.055				0.094	0.062				0	0
Intervention shorter than 4 months					ref.					ref.					ref.					ref.
4–8 months					0.012					-0.005					0.01					0.018**
					0.667					0.8					0.515					0.013
Intervention longer than 8 months					-0.021					-0.048**					0.003					0.008

	Subsample: High-income countries										Subsample: Low- and middle-income countries									
	Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)					Weighted least squares Hedges' g regressions					Random effects SMD regressions (restricted maximum likelihood)				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
					0.52					0.028					0.891					0.493
Intervention with regional or local scale					ref.					ref.					ref.					ref.
National scale					-0.024					-0.028*					-0.021					-0.019**
					0.213					0.079					0.21					0.028
Implemented by private and public combined					ref.					ref.					ref.					ref.
Public only					0.025***					0.024*					-0.042					-0.038***
					0.002					0.079					0.117					0
Private only					-0.02					0.016					-0.037**					-0.026***
					0.309					0.463					0.01					0.002
Constant	0.017***	0.019	0.004	-0.006	0.007	0.026***	0.050**	0.034**	0.017	0.073**	0.071***	0.085***	0.066**	0.053**	0.050**	0.074**	0.091**	0.058***	0.047**	0.029**
	0.007	0.196	0.83	0.718	0.82	0	0	0.026	0.329	0.029	0	0	0.006	0.027	0.041	0	0	0	0	0.02
Estimates	134277	132268	129663	129663	112950	134277	132268	129663	129663	112950	241115	232104	213489	213489	198169	241115	232104	213489	213489	198169
Interventions	88	79	74	74	59	88	79	74	74	59	113	100	86	86	69	113	100	86	86	69
Adjusted R-squared	0.018	0.04	0.087	0.103	0.122						0.069	0.058	0.115	0.136	0.175					

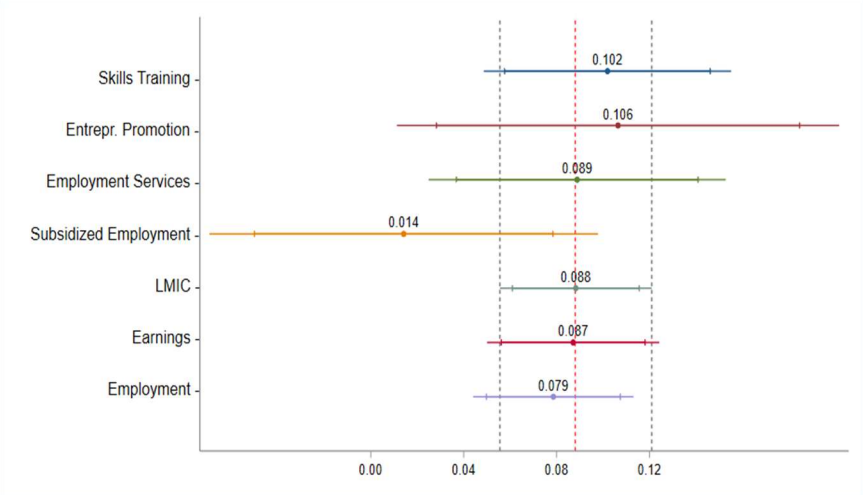
Note: *= p<0.10, **= p<0.05, ***= p<0.01.

Box 5
Impact evaluations of active labour market programmes in Africa

Almost one in three (30 per cent, or 72 in total) of the interventions covered in this review were implemented across Africa (see Box 2), in either low- or middle-income countries. The analysis of the subsample of active labour market programmes (ALMPs) in Africa found similarities with the findings that emerged in the analysis of evaluations in the low- and middle-income countries as well as some important deviations.

- **The ALMPs evaluated in Africa successfully improved labour market outcomes of youth.** Across all interventions and outcome measures, the average estimated effect size was 0.09 SMD, as shown in the figure below. This finding was in line with the average effects found for the low- and middle-income countries globally (see Figure 15).
- **Overall, entrepreneurship interventions, skills trainings and employment services were about equally effective in improving labour market outcomes for young people.** Point estimates across these three categories of interventions varied between 0.09 SMD and 0.10 SMD, and the differences were not statistically significant. Subsidized employment interventions appeared less effective in improving labour market outcomes, although the estimate was accompanied by a large confidence interval.
- **Accounting for sample and intervention characteristics, entrepreneurship promotion interventions appeared more impactful than other types of ALMPs.** The multivariate regression analysis revealed that when controlling for sample and intervention characteristics, the effect of entrepreneurship interventions improved by between 0.04 SMD and 0.11 SMD, relative to skills training. This was particularly relevant in the African context, where one in three included interventions delivered mainly entrepreneurship promotion (compared to 14 per cent globally).

▶ **Summary forest plot for the sample of interventions implemented in Africa: intervention category, country income level and programme scale**



Note: The analysis looked at 51 independent samples (studies) based on the full sample of 623 effect sizes that remained after censoring (SMD limit = 1, inverse standard error limits = 1–100). Clustering excluded 422 effect sizes because a higher-level effect size was available (see section 5.2 for details).

- **Multipronged interventions appeared more effective, yet the evidence on more comprehensive programmes was considerably more nuanced than for the low- and middle-income countries globally.** The studies evaluating interventions with multiple components reported considerably higher effect sizes (with an additional average effect between 0.08 SMD and 0.15 SMD, see Annex A22). However, other factors associated with higher effect sizes in the low- and middle-income countries globally were not correlated with better labour market outcomes in the Africa subsample. For instance, there was no evidence that adding a soft skills training component to an ALMP enhanced its impact, nor that providing certification of participation or attendance associated with increased effectiveness. Regarding the duration of programmes, which is a proxy of programme intensity, the evidence was inconclusive: Programmes lasting between four and eight months reported smaller effects (when compared with shorter than four months), while ALMPs longer than eight months showed larger impacts, on average. After confirming the findings from the low- and middle-income countries at the global level, no (positive or negative) differential effect for interventions providing incentives to programme participants was observed.

- **Young people from economically disadvantaged backgrounds who participated in ALMPs in Africa reported better labour market outcomes.** The differential impact of 0.05–0.08 SMD was consistently significant across specifications in the multivariate regressions (see Annex A22). This finding represented an encouraging trend for a region in which 43 per cent of evaluated interventions targeted low-income or economically disadvantaged young people (see Box 2).
- **Interventions implemented at the regional or local level reported considerably larger effect sizes compared to ALMPs rolled out at the national level.** The magnitude of the impact was large (up to 0.10 SMD) and contrasted with the findings from the low- and middle-income countries at the global level, where no such difference was observed. This may point to the importance of targeting interventions to local labour market contexts. It may also suggest that effectively implementing interventions at national scale may be constrained by uneven capacities of implementing partners, public or private.
- **When measuring labour market outcomes more than a year after the end of the interventions, the average effect sizes diminished considerably across the sample of interventions implemented in Africa.** This finding suggests that maintaining the initially significant and positive impact of ALMPs in many cases is still challenging. It merits further research to better understand how and under which conditions effects do materialize and are maintained in the medium to long run.

▶ 7 Conclusion

In 2024, around 20.4 per cent of the global youth population was categorized as NEET – not in employment, education or training. Although youth labour market outcomes have improved since the COVID-19 crisis, young people continue to encounter significant challenges in accessing productive and decent employment. Weak job creation, pervasive informality and limited access to social protection systems constrain their opportunities, often pushing them into inactivity or precarious work. In low-income countries, three out of four employed youth remain self-employed or in temporary positions. The pandemic crisis also generated severe learning losses, disrupted education and skills development pathways and reduced opportunities for work-based learning. These setbacks have further eroded youth employability and deepened existing skills mismatches.

The time to act and foster effective transitions into decent jobs for youth is now. This calls for substantially scaling up interventions and investments in youth, particularly those experiencing slower, more uneven labour market recoveries and facing potentially long-term scarring effects on their career trajectories. Such efforts should be guided by rigorous evidence on what works, for whom and under which conditions.

To inform these investments, this report presents the results of the most comprehensive systematic review of youth employment interventions to date. The analysis represented here examined the effectiveness of youth-targeted ALMPs in improving employment, earnings and business performance outcomes. And it assessed how context, youth characteristics, evaluation features and programme design and implementation moderate the impact of these interventions. In doing so, it identified the design features and delivery mechanisms associated with greater effectiveness, thus offering actionable lessons for future programming.

This systematic review synthesized more than three decades of impact evaluation evidence, considerably expanding on earlier literature, such as Kluve et al. (2019 and 2017). It incorporated new or updated reports, refined inclusion criteria and applied enhanced search methodologies, including machine learning techniques, to capture the most relevant and recent information.

The evidence base on youth employment programmes increased substantially in the past decade. This review doubled the information gathered and computed by Kluve et al. (2019 and 2017) across all key metrics: from 113 to 228 reports, from 107 to 220 interventions, from 87 to 171 ALMPs and from evidence originating in 31 countries to 62 countries. Most of the new evidence originated in low- and middle-income economies, thus enabling a more balanced global picture and richer analysis of programme heterogeneity across regions and income levels.

Today's evidence is geographically broader and methodologically stronger. Africa represented the largest share of studies (33 per cent), followed by Europe (27 per cent) and Latin America and the Caribbean (18 per cent). Experimental designs accounted for 47 per cent of the 248 studies, reflecting the field's growing methodological rigor. Roughly three quarters of studies were either peer reviewed (38 per cent) or published as working papers (35 per cent). The majority of impact estimates concerned employment outcomes (60 per cent), followed by earnings (35 per cent) and business performance (6 per cent). Nonetheless, evidence gaps persist, notably for entrepreneurship support in high-income countries and for outcomes related to business performance and self-employment in developing regions.

Main findings

1. Overall effectiveness

Youth-focused ALMPs were effective in improving labour market outcomes overall. The random-effects meta-analysis yielded an average effect size of 0.08 SMD on the employment and earnings of youth. This corresponded to a medium effect size, according to recent benchmarks in the social sciences (0.05–0.20 SMD). Importantly, results remained positive and statistically significant after correcting for publication bias (0.06 SMD) and when restricting the sample to experimental evaluations only (0.05 SMD).

Effectiveness improved over time. Programmes initiated after 2008 were associated with larger effect sizes (ceteris paribus), suggesting learning and refinement in programme design and delivery. These findings represented clear

progress since the studies of Kluve et al. (2019 and 2017), which found an average impact of 0.04 SMD based on reports published before 2015.

2. Differences by country income group

ALMPs tended to be more effective in the low- and middle-income countries than in the high-income countries, on average. The average effect was 0.09 SMD in the low-income countries and 0.10 SMD in the middle-income countries versus an average effect of 0.06 SMD in the high-income countries. This finding held after controlling for difference in study quality and intervention characteristics.

This finding likely reflects the greater marginal returns to investment in youth employment in developing contexts, where young people face deeper structural barriers, such as limited formal job opportunities, a weak skills base and pervasive informality. Even modest investments in skills or job placement can therefore yield relatively large short-term impacts. However, the effects in the low- and middle-income countries tended to fade within a year, underscoring the need for sustained follow-up, such as career guidance, apprenticeships and business mentoring.

In contrast, ALMPs in the high-income countries tended to yield larger long-term effects and operated mainly through the employment rather than earnings channel, reflecting a stronger focus on unemployed youth rather than those in informal or low-paid jobs.

3. The "what": Type of intervention

Programme type was a major determinant of impact.

- **In the high-income countries, the studies of skills training interventions reported the largest impacts on employment and earnings of youth**. This aligned with Card, Kluve and Weber (2018), who found that human capital-oriented ALMPs tended to have stronger long-term effects for adult populations.[28]
- **In the low- and middle-income countries, entrepreneurship interventions reported the largest impacts**, **followed by employment services and skills training interventions.** This may have been linked to the scarcity of private sector jobs or wage employment in general, prompting the need to support youth in creating their own business ventures. In these countries, young people are often exposed to low-quality education, widespread informality and labour market information asymmetries, which limits their ability to look for jobs or signal work-relevant skills and competencies to employers. In these contexts, often small-scale, low-cost employment services can help overcome the barriers, facilitating entry into the labour market.

These differentiated findings highlight the importance of aligning ALMP portfolios with national labour market structures. They also mark a significant advance over earlier analyses (such as Kluve et al. 2019) that found no systematic differences across intervention types, thus suggesting improvements in both the evidence base and programme design quality over time.

4. The "how": Design and implementation

Programme design and delivery mechanisms were critical to success.

- **In the high-income countries, interventions led by public institutions alone yielded larger effects than those involving public–private partnerships.**
- **In the low- and middle-income countries, the opposite held.** Although less consistent across specifications, public–private partnerships were associated with higher effectiveness, likely due to their ability to mobilize complementary capacities, reach specific groups and address multiple labour market constraints.

[28] There was little evidence from youth entrepreneurship programmes in the high-income countries, which suggests room for further testing of their impact in such context.

- **In the low- and middle-income countries, programmes tended to have larger impacts if they lasted for at least four months and integrated multiple interventions**. This suggests that ALMPs that offer various services with sufficiently long duration are better able to address the many constraints young people face in these countries. Such programmes often offer skills training and job matching or subsidized employment.
- **Larger effects were also found among programmes that integrated soft skills in training content and those that certified participation or the skills acquired by young participants.** Hence, an important factor in low- and middle-income countries seems to be that ALMPs enable young jobseekers to better signal their (non-technical) competencies to employers.

Overall, the evidence points to the added value of comprehensive, skills-plus approaches over single, short-duration interventions.

5. The "who": Young people benefiting from ALMPs

Many evaluation studies reported impact estimates across different groups of participants, for example, by gender or age. The results hinted at larger impacts for participants younger than 25 years. Moreover, in the low- and middle-income countries, the ALMPs had larger impact on youth from economically disadvantaged backgrounds and young women. This suggested a strong equity and inclusion dimension: **Programmes that intervened early in the life cycle and prioritized young women and disadvantaged youth not only enhance fairness but also maximize aggregate programme impact**.

6. Programme set-up and delivery scale

The regional or local programmes often had larger impacts than the national ones. Possibly, these localized approaches facilitated direct engagement with local organizations and employers. Local programmes also may have benefited from more homogenous participant groups and therefore were better tailored to their needs. Furthermore, the capacity of implementing partners may have been particularly strong in some areas of the low- and middle-income countries and more uneven at the national level. This finding points to the value of decentralized implementation and investment in local delivery ecosystems.

7. Cost-effectiveness and evidence gaps

The systematic review revealed an encouraging positive trend in the inclusion of cost information and cost–benefit analyses in impact evaluations since the mid-2010s. Around three quarters of the available cost–benefit analyses reported that programme benefits exceeded cost, with particularly high ratios among skills training interventions (at 95 per cent). However, reporting practices remained inconsistent. Disclosing a minimum set of information, including the total cost of the programme, the unit cost per programme participant, the currency and its corresponding year, will be critical to strengthen comparability and support prioritization of scalable, high-impact interventions under fiscal constraints.

Persistent gaps remain in the evidence base, particularly regarding long-term effects, business performance outcomes, entrepreneurship programmes in high-income contexts and overall evidence from Arab States. Moreover, further comparative analyses (including multi-arm experimental designs) are needed to assess the relative effectiveness of different intervention components and delivery models.

8. Methodological and research priorities

Leveraging the extensive dataset from the systematic review, future research should substantiate the findings and assess robustness of results by:

- applying alternative techniques to address publication bias and dependency of effect sizes;

- incorporating risk-of-bias assessments to account for study quality; and
- leveraging advanced meta-analytic methods, such as robust variance estimation, network meta-analysis or Bayesian approaches (see Bartoš et al. 2023; Wilson et al. 2016).

Although meta-analysis provides rigorous quantitative synthesis, it cannot capture the full complexity of programme mechanisms. Further work may extract and synthesize some of the more qualitative findings in the underlying primary studies to gain a deeper understanding of the drivers and mechanisms of successful ALMPs for youth. Combining quantitative and qualitative synthesis will enhance the policy relevance and interpretability of future evidence reviews.

▶ References

Apunyo, Robert, Howard White, Caroline Otike, Thomas Katairo, Susana Puerto, Drew Gardiner, Alison Annet Kinengyere, John Eyers, Ashrita Saran, and Ekwaro A. Obuku. 2022. "Interventions to Increase Youth Employment: An Evidence and Gap Map", *Campbell Systematic Reviews* 18: e1216.

Asenjo, Antonia, Verónica Escudero, and Hannah Liepmann. 2022. "Why Should We Integrate Income and Employment Support? A Conceptual and Empirical Investigation", ILO Working Paper No. 72. Geneva: ILO.

Askarov, Zohid, and Hristos Doucouliagos. 2020. "A Meta-Analysis of the Effects of Remittances on Household Education Expenditure", *World Development* 129: 104860.

Baird, Sarah, Francisco H. G. Ferreira, Berk Özler, and Michael Woolcock. 2013. "Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review", *Campbell Systematic Reviews* 8.

Barford, Anna, Adam Coutts, and Garima Sahai. 2022. *Youth Employment in Times Of COVID. A Global Review of COVID-19 Policy Responses to Tackle (Un)Employment and Disadvantage Among Young People*. Geneva: ILO.

Bartoš, František, Maximilian Maier, Eric-Jan Wagenmakers, Hristos Doucouliagos, and T. D. Stanley. 2023. "Robust Bayesian Meta-analysis: Model-averaging across Complementary Publication Bias Adjustment Methods", *Research Synthesis Methods* 14 (1): 99-116.

Betcherman, Gordon, Martin Godfrey, Susana Puerto, Friederike Rother, and Antoneta Stavreska. 2007. "A Review of Interventions to Support Young Workers: Findings of the Youth Employment Inventory." Social Protection Discussion Paper Series, No. 715. Washington, DC: World Bank.

Blattman, Christopher, Eric P. Green, Julian Jamison, Christian Lehmann, and Jeannie Annan. 2016. "The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda", *American Economic Journal: Applied Economics* 8 (2): 35-64.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2020. "The Long-Term Impacts of Grants on Poverty: Nine-Year Evidence from Uganda's Youth Opportunities Program", *American Economic Review: Insights* 2 (3): 287-304.

Borenstein, Michael, Harris Cooper, Larry V. Hedges, and Jeffrey C. Valentine. 2009. "Effect Sizes for Continuous Data." In *The Handbook of Research Synthesis and Meta-Analysis,* edited by H. Cooper, V. Hedges and J. C. Valentine. 221–235. New York: Russell Sage Foundation.

Borenstein, Michael, Larry V. Hedges, J.P.T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-analysis.* New York: John Wiley & Sons Ltd.

Brooks, Wyatt, Kevin Donovan, and Terence R. Johnson. 2018. "Mentors or Teachers? Microenterprise Training in Kenya." *American Economic Journal: Applied Economics* 10 (4): 196–221.

Brudevold-Newman, Andrew, Maddalena Honorati, Pamela Jakiela, and Owen Ozier. 2017. *A Firm of One's Own: Experimental Evidence on Credit Constraints and Occupational Choice.* IZA Discussion Papers 10583; also World Bank Policy Research Working Paper 7977.

Caliendo, Marco, and Ricarda Schmidl. 2016. "Youth Unemployment and Active Labor Market Policies in Europe", *IZA Journal of Labor Policy* 5: 1.

Card, David, Jochen Kluve, and Andrea Weber. 2010. "Active Labour Market Policy Evaluations: A Meta-analysis", *The Economic Journal* 120 (548): F452–F477.

—. 2018. "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations", *Journal of the European Economic Association* 16 (3): 894–931.

Crépon, Bruno, Mohamed El Komi, and Adam Osman. 2024. "Is It Who You Are or What You Get? Comparing the Impacts of Loans and Grants for Microenterprise Development", *American Economic Journal: Applied Economics* 16 (1): 286-313.

Deeks, Jonathan J., Douglas G. Altman, and Michael J. Bradburn. 2001. “Statistical Methods for Examining Heterogeneity and Combining Results from Several Studies in Meta-analysis.” In *Systematic Reviews in Health Care: Meta-analysis in Context*, Second edition, edited by M. Egger, G.D. Smith and D.G. Altman. London: BMJ Publishing Group.

Egebark, Johan and Niklas Kaunitz. 2018. "Payroll Taxes and Youth Labor Demand," *Labour Economics* 55 (C): 163-177.

Eichhorst, Werner, and Ulf Rinne. 2015. “An Assessment of the Youth Employment Inventory and Implications for Germany’s Development Policy.” IZA Research Report No. 67.

Elder, Sara, and Niall O’Higgins. 2023. “Has Youth Employment Recovered?” ILO Brief. Geneva: ILO.

Escudero, Verónica, Jochen Kluve, Elva López Mourelo, Clemente Pignatti. 2018. “Active Labour Market Programmes in Latin America and the Caribbean: Evidence from a Meta-Analysis”, *Journal of Development Studies* 55 (12): 2644–2661.

Evans, David K., and Fei Yuan. 2022. “How Big Are Effect Sizes in International Education Studies?” *Educational Evaluation and Policy Analysis* 44 (3), 532-540.

Fares, Jean and Susana Puerto. 2009. “Towards Comprehensive Training”, Social Protection Discussion Paper Series No. 0924. Washington, DC: World Bank.

Fiala, Nathan V. and Kevin Marcus Hempel. 2011. *Measuring Success of Youth Livelihood Interventions: A Practical Guide to Monitoring and Evaluation* (English). Washington, DC: World Bank Group.

Grimm, Michael and Anna Luisa Paffhausen. 2015. “Do Interventions Targeted at Micro-Entrepreneurs and Small and Medium-Sized Firms Create Jobs? A Systematic Review of the Evidence for low- and middle-Income Countries”, *Labour Economics* 32: 67–85.

Gucciardi, Daniel F., Robin L. J. Lines, and Nikos Ntoumanis. 2022. “Handling Effect Size Dependency in Meta-analysis”, *International Review of Sport and Exercise Psychology* 15 (1), 152-178.

Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith. 1999. “The Economics and Econometrics of Active Labour Market Programs.” In *Handbook of Labour Economics* 3, edited by O. Ashenfelter and D. Card. Amsterdam: Elsevier.

Hedges, Larry and Ingram Olkin. 1985. *Statistical Models for Meta-analysis*. New York: Academic Press.

ILO. 2010. *Global Employment Trends for Youth: Special Issue on the Impact of the Global Economic Crisis on Youth*.

—. 2020a. *Youth and COVID-19: Impacts on Jobs, Education, Rights and Mental Well-Being*.

—. 2020b. “*Delivering Income and Employment Support in Times of COVID-19: Integrating Cash Transfers with Active Labour Market Policies*.” Policy Brief.

—. 2021. *Global Framework on Core Skills for Life and Work in the 21st Century*.

—. 2022. *Global Employment Trends for Youth 2022: Investing in Transforming Futures for Young People*.

—. 2024. *Global Employment Trends for Youth 2024: Decent Work, Brighter Futures.*

—. 2025. *World Employment and Social Outlook: Trends 2025.*

Jaramillo, Miguel and Sandro Parodi. 2003. *Jóvenes emprendedores: evaluación de programas de promoción*. Lima: Instituto APOYO.

Kaiser, Tim, Annamaria Lusardi, Lukas Menkhoff, and Carly Urban. 2022. “Financial Education Affects Financial Knowledge and Downstream Behaviors”, *Journal of Financial Economics* 145 (2): 255–272.

Keef, Stephen P., and Leigh A. Roberts. 2004. “The Meta-analysis of Partial Effect Sizes”, *British Journal of Mathematical and Statistical Psychology* 57 (1): 97–129.

Kluve, Jochen, and Christoph M. Schmidt. 2002. ”Can Training and Employment Subsidies Combat European Unemployment?”, *Economic Policy* 17 (35): 409–448

Kluve, Jochen, Susana Puerto, David Robalino, Jose M. Romero, Friederike Rother, Jonathan Stöterau, Felix Weidenkaff, and Marc Witte. 2017. "Interventions to Improve the Labour Market Outcomes of Youth: A Systematic Review", *A Campbell Systematic Review* 12.

—. 2019. "Do Youth Employment Programs Improve Labor Market Outcomes? A Quantitative Review", *World Development* 114.

Kraft, Matthew A. 2020. "Interpreting Effect Sizes of Education Interventions", *Educational Researcher* 49 (4): 241-253.

—. 2023. "The Effect Size Benchmark that Matters Most: Education Interventions Often Fail", *Educational Researcher* 52 (3), 183-187.

Kugley, Shannon, Anne Wade, James Thomas, Quenby Mahood, Anne-Marie K. Jørgensen, Karianne Hammerstrøm, and Nila Sathe. 2017. "Searching for Studies: A Guide to Information Retrieval for Campbell Systematic Reviews", *Campbell Systematic Reviews* 13 (1): 1–73.

LaLonde, Robert J. 2003. *Employment and Training Programs. Means-Tested Transfer Programs in the United States*. Chicago, ILL: University of Chicago Press: 517–586.

Levi Yeyati, Eduardo, Martín Montané, and Luca Sartorio. 2021. "What Works for Active Labor Market Policies?," Working Papers No. 43, Red Nacional de Investigadores en Economía.

Lipsey, Mark W., and David B. Wilson. 2001. "Practical Meta-analysis", Applied Social Research Methods Series No. 49. SAGE Publications.

Mariani, Marco, Alessandra Mattei, Lorenzo Storchi, and Daniele Vignoli. 2019. "The Ambiguous Effects of Public Assistance to Youth and Female Start-Ups between Job Creation and Entrepreneurship Enhancement", *Scienze Regionali* 18 (2): 237-260.

McKenzie, David. 2020. "*Small Business Training to Improve Management Practices in Developing Countries Reassessing the Evidence for 'Training Doesn't Work'*", Policy Research Working Paper No. 9408. Washington, DC: World Bank.

McKenzie Joanne E and Sue E Brennan. 2022. "Synthesizing and presenting findings using other methods". Chapter 12 in Higgins, Julian, James Thomas, Jacqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew Page, and Vivian Welch*. Cochrane Handbook for Systematic Reviews of Interventions (current version).*

O'Higgins, Niall. 2017. *Rising to the Youth Employment Challenge: New Evidence on Key Policy Issues.* Geneva: ILO.

—. 2022. *On the Design of Volunteer Programmes to Facilitate the Entry and Re-entry of Young People Into Work.* Geneva: ILO and United Nations Volunteers Programme.

OECD. 2022. "OECDilibrary".

Piza, Caio, Tulio A. Cravo, Linnet Taylor, Lauro Gonzalez, Isabel Musse, Isabela Furtado, Ana C. Sierra, and Samer Abdelnour. 2016. "The Impact of Business Support Services for Small and Medium Enterprises on Firm Performance in low- and middle-Income Countries: A Systematic Review." *Campbell Systematic Reviews* 1.

Saez, Emmanuel, Benjamin Schoefer, and David Seim. 2021. "Hysteresis From Employer Subsidies", *Journal of Public Economics* 200 (104459): ISSN 0047-2727.

Simonsohn, Uri, Simmons, Joseph and Leif D. Nelson. 2022. "Above Averaging in Literature Reviews", *Nature Reviews Psychology* 1: 551–552.

Stanley, Tom and Hristos Doucouliagos. 2012. *Meta-regression Analysis in Economic and Business.* London: Routledge.

Stöterau, Jonathan, Johanna Kemper, and Andrea Ghisletta. 2022. *The Impact of Vocational Training Interventions on Youth Labor Market Outcomes: A Meta-Analysis.*

The Campbell Collaboration. 2021. *Campbell Collaboration Systematic Reviews: Policies and Guidelines Version 1.8*.

Trine, Janice, Jorge Hombrados, Mark Newman, Kimberly Hovish, Chris Brown, Katarzyna Steinka-Fry, and Eric Wilkey. 2013. "Technical and Vocational Education and Training (TVET) Interventions to Improve the Employability and Employment of Young People in low- and middle-Income Countries: A Systematic Review", *Campbell Systematic Reviews* 9

Valerio, Alexandra, Brent Parton, and Alicia Robb. 2014. *Entrepreneurship Education and Training Programs Around the World: Dimensions for Success*. Washington, DC: World Bank.

Vooren, Melvin, Carla Haelermans, Wim Groot, and Henriëtte Maassen van den Brink. 2019. "The Effectiveness of Active Labor Market Policies: A Meta-Analysis", *Journal of Economic Surveys* 33 (1): 125–149.

Voceru, Joachim, Uri Simonsohn, Leif D. Nelson, and Joseph Simmons. 2019. "99% Impossible: A Valid, or Falsifiable, Internal Meta-analysis", *Journal of Experimental Psychology: General* 148 (9): 1628–1639.

Waddington, Hugh, Howard White, Birte Snilstveit, Jorge Garcia Hombrados, Martina Vojtkova, Philip Davies, Ami Bhavsar, John Eyers, Tracey Perez Koehlmoos, Mark Petticrew, Jeffrey C. Valentine, and Peter Tugwell. 2012. "How to Do a Good Systematic Review of Effects in International Development: A Tool Kit", *Journal of Development Effectiveness* 4 (3): 359–387.

Wilson, David B., Emily Tanner-Smith and Dimitris Mavridis. 2016. "Network Meta-analysis", Campbell Systematic Reviews, Methods Research Paper, Vol. 12, No 1.


International
Labour
Organization


WORLD BANK GROUP


Co-funded by
the European Union


Ministry of Foreign Affairs