



The impact of active labour market programmes on youth

An updated systematic review and meta-analysis



International
Labour
Organization



WORLD BANK GROUP



Ministry of Foreign Affairs

► **The impact of active labour market programmes on youth**

An updated systematic review and meta-analysis

2022

Copyright © International Labour Organization 2022
First published 2022

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

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

The impact of active labour market programmes on youth
An updated systematic review and meta-analysis
Geneva: International Labour Office, 2022

ISBN 9789220385210

► Acknowledgements

The report *The impact of active labour market programmes on youth* was authored by Susana Puerto, Chiara Curcio and Jonas Bausch from the International Labour Organization (ILO) and Jonathan Stöterau and Michael Weber from the World Bank Group. It was made possible through the substantial financial contribution from the Policy and Operations Evaluation Department (IOB) at the Ministry of Foreign Affairs of the Netherlands as well as complementary financial support from the ILO and the World Bank Group.

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 contribution 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 and Paulo Fernandes for his valuable contribution to the last stage of the study.

Special thanks go to Evangelia Bourmpoula and Yamila Simonovsky for their technical expertise and significant contribution to the coding, analysis and the review of programme costs; and to Felix Weidenkaff for the important feedback and advice in the early stages of the study. We thank Sukti Dasgupta, Chief of the ILO's Employment, Labour Markets and Youth Branch for her constant 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 and colleagues from the ILO Employment, Labour Markets and Youth Branch, including Kee Beom Kim and Anna-Karin Palm Olsson.

► Contents

Acknowledgements	4
Contents	5
Abbreviations	7
Executive summary	8
Background	8
Objective	8
Data and methodology	8
Results	9
Conclusion	11
1 Introduction	12
2 Inclusion criteria	16
2.1 Population and context	16
2.2 Interventions	16
2.3 Comparison group	18
2.4 Outcome	18
2.5 Study design and methods	18
3 Search and data collection	19
3.1 Scoping, primary and complementary search	19
3.2 Data collection	21
4 Study and intervention characteristics	23
4.1 Global overview	23
4.2 Regional differences	26
5 Effect size computation and meta-analysis methods	30
5.1 Standardizing effect size estimates	30
5.2 Dealing with dependent effect sizes	31
5.3 Summarizing effect size estimates	33
5.4 Assessing correlates of effect size magnitude	33
5.5 Assessing and accounting for reporting bias	34
6 Results	35
6.1 Summary effect sizes	35
6.2 Publication bias	39
6.3 Multivariate meta-regression results	40
6.3.1 Outcomes of interest	41
6.3.2 Country income level	41
6.3.3 Intervention characteristics	42
6.3.4 Participant characteristics	43
6.3.5 Evaluation and study characteristics	44
6.4 Costs	56
7 Conclusions	59
References	62
Annexes	65
A.1. Study and programme information	65
A.2. Primary search sources	66
A.3. Effect size information	67
A.4. Description of data used and empirical methods	70
A.5. Intervention and programme characteristics	70

A.6. Programme characteristics: Skills training	71
A.7. Programme characteristics: Entrepreneurship promotion	72
A.8. Programme characteristics: Employment services	73
A.9. Programme characteristics: Wage subsidies	74
A.10. Programme characteristics: Public Works	74
A.11. Targeting strategy	75
A.12. Cost–benefit information variables	77
A.13. Study-level variables	78
A.14. Risk of bias assessment variables	78

Figures

Figure 1. Search results for updated systematic review	20
Figure 2. Increasing body of evidence, by publication year, 1993–2021	24
Figure 3. Increasing body of evidence from low- and middle-income countries, 1993–2021	25
Figure 4. Increasing number of randomized control trial reports, 1993–2021	25
Figure 5. Evolution of evaluated youth employment interventions before (panel A) and after (panel B) 2015	26
Figure 6. Main categories of interventions, by region	27
Figure 7. Main outcome construct evaluated for each intervention, regional differences	28
Figure 8. Urban versus rural focus, regional differences	28
Figure 9. Summary forest plot for the full sample: Intervention category, country income level and programme scale	37
Figure 10. Summary forest plot for the full sample: Outcome categories, gender and youths' income level	38
Figure 11. Funnel plots	39
Figure 12. Increasing number of reports citing cost information, 1993–2021	57

Tables

Table 1. Comparison of active labour market programme categories in relevant and recent studies	16
Table 2. Interventions of interest for the updated systematic review	17
Table 3. Outcomes of interest for the updated systematic review	18
Table 4. Keyword and controlled vocabulary terms	19
Table 5. Characteristics of included studies and interventions in the updated systematic review	23
Table 6. Different components of interventions, by region	27
Table 7. Frequencies of coded treatment effects	35
Table 8. Test for publication bias	40
Table 9. Meta-regression results, full sample	45
Table 10. Meta-regression results, subsample outcome category	47
Table 11. Meta-regression results, subsample country income level	50
Table 12. Information on programme cost and cost–benefit analyses across country income level and main category of interventions	58

Boxes

Box 1. Africa at a glance	29
Box 2. Preselection and aggregation of effect size estimates for univariate random effects meta-analysis	32
Box 3. Reading a summary forest plot	36
Box 4. Impact evaluations of ALMPs in Africa	54

▶ Abbreviations

ALMP	active labour market programmes
ATE	average treatment effect
ATET	average treatment effect on the treated
ATT	average treatment effect on treated
CI	confidence interval
DiD	difference-in-differences
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-governmental organization
OECD	Organization for Economic Co-operation and Development
PEESE	precision-effect-estimate with standard error
PET	precision effect test
PSM	propensity score matching
RCT	randomized controlled trials
RDD	regression discontinuity design
SMD	standardized mean differences
WLS	weighted least squares

▶ Executive summary

Background

The COVID-19 pandemic severely affected decent job opportunities and labour market outcomes for youth.

The recession induced by the pandemic has impacted young people through job and income loss, greater barriers in the transition to work, and disruptions in learning, education, and training. Between 2019 and 2020, 34 million young workers lost their job, equivalent to a decline in global youth employment of 8.2 per cent, more than twice the employment loss of 3.6 per cent among adults. The recession did not only translate into greater unemployment but also massive labour force exit. In 2020, the number of youth not in employment, education, or training (NEET) reached a record high of 282 million. Young women shouldered the worst of the youth employment crisis, being 1.5 times less likely than young men to be employed and twice as likely as young men to have NEET status (ILO 2022).

The jobs and learning losses resulting from the pandemic and the emerging effects of new and ongoing crises can have long-lasting negative repercussions for youth. Such economic shocks not only disrupt the labour market trajectory of young workers who lose their job early in their career, but also hamper the transition to employment for new labour force entrants. Past economic crises have shown that such short-term shocks often lead to long-term scarring effects among youth.

Active labour market programmes (ALMPs) can support recovery efforts and improve labour market prospects of youth. ALMPs were promoted as an essential component of the COVID-19 recovery plans, with both new and existing programmes launched and repurposed. However, not many programmes targeted young people or were agile enough to address the uneven challenge felt by youth. Nonetheless, well-designed and well-implemented ALMPs can help young people access jobs and thus prevent long-term scarring. Earlier evidence highlighted their positive effect on employment and earnings outcomes of young beneficiaries, and the important role of multipronged interventions in addressing the various needs and constraints faced by youth (Kluve et al. 2017 and 2019).

Further research about the impact of youth-targeted ALMPs is therefore crucial to enhance their effectiveness and facilitate a sustainable and resilient recovery with and for young people. In a global context of tight fiscal space, evidence on what works in youth employment must underpin decision-making for immediate and targeted investments in youth. Understanding “what” types of ALMPs are effective, for “whom”, in which contexts, and “how” to increase their cost-effectiveness is central to an economic recovery that benefits young people and promotes a better future of work for all.

Objective

This systematic review and meta-analysis examine the effectiveness of youth-targeted ALMPs to improve labour market outcomes of young people. To this end, the review synthesizes the empirical evidence of the past three decades from impact evaluations of youth-targeted ALMPs. The analysis seeks to assess the role that context, youth characteristics, evaluation features, programme design and implementation had in moderating the impact on employment, earnings and business performance outcomes.

Data and methodology

The systematic review builds on a significantly expanded database, benefitting from a surge in impact evaluations of youth employment programmes in recent years. The updated review also employs a refined and more comprehensive literature search. To be included, studies must have been conducted between 1990 and 2022 and meet the following inclusion criteria:

1. Investigate an ALMP that was primarily targeted at young people (broadly applying the age range of 15 to 35 years) and designed to address their labour market constraints.

2. Assess an ALMP that includes 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, namely: 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 includes studies from 62 countries. Nearly 60 per cent of the 220 included interventions were implemented in low- or in middle-income countries, with Africa accounting for 33 per cent of all interventions. Skills training is the most commonly evaluated intervention type, constituting 57 per cent of the included interventions. Moreover, the analysis builds on a large share of high-quality evaluation designs, with 47 per cent of studies conducted as a randomized controlled trial (RCT) and 38 per cent published in peer-reviewed journals.

To synthesize the evidence and assess what determines 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 included 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. Finally, sub-sample analysis was conducted to assess whether such correlates differed by country-income level and outcome category.

Finally, the report offers a deep-dive on Africa, leveraging the growing evidence on youth employment from the region. The analysis provides a detailed descriptive overview of the studies and intervention characteristics specific to Africa and assesses the role that key design and implementation features plays in moderating the impact of ALMPs in the continent.

Results

The report highlights seven key findings:

First, ALMPs are effective in improving the labour market outcomes of youth. The overall effect size across all studies is 0.08 SMD with a 95 per cent confidence interval (CI) of 0.07–0.09 SMD. The result suggests some progress vis-à-vis the study of Kluge et al. (2017 and 2019), which built on reports published before 2015 and found an average impact of 0.04 SMD (CI = 0.02–0.06 SMD). Recent meta-analyses in social sciences define “small” effects as those below a 0.05 SMD and “medium” as those in the 0.05–0.20 SMD range, supporting the case for increased effectiveness of ALMPs. Moreover, the new set of results documents a statistically significant positive impact of ALMPs across all country income groups and outcome categories.

Second, youth-focused ALMPs are more effective in low- and middle-income countries than in high-income countries on average. The study finds an effect size of 0.09 SMD (95 per cent CI: 0.06–0.11 SMD) for low-income countries, 0.10 SMD (95 per cent CI: 0.08–0.12 SMD) for middle-income countries and 0.06 SMD (95 per cent CI: 0.04–0.08 SMD) for high-income countries. Looking at the sub-sample of ALMPs from Africa, the study documents an overall effect size of 0.10 SMD (95 per cent CI: 0.08–0.13 SMD). To explain why ALMPs achieve larger impacts in

developing countries, Betcherman et al. (2007) and Kluge et al. (2017 and 2019) argued that since youth in low- and middle-income countries face more intense labour market challenges, a marginal investment to improve their situation pays off more there than in high-income countries. Moreover, the impacts of ALMPs in low- and middle-income countries are relatively larger on earnings than on employment outcomes. The opposite is observed for studies from high-income countries.

Third, the type of ALMP (referred in the report as the “what”) plays an important role in determining its success, with entrepreneurship promotion and skills training interventions driving positive outcomes.

Overall, entrepreneurship promotion and skills training interventions report larger impacts than employment services and subsidized employment programmes. This result contrasts with earlier findings by Kluge et al. suggesting that the most recent evidence captures more clearly the differential effect of certain intervention types. Some important nuances emerge across country income groups:

- **In the low- and middle-income countries, studies of entrepreneurship interventions report the largest impacts, followed by employment services and skills training interventions.** This finding may be connected to a lack of private sector jobs, gaps in education systems, and the prevalence of labour market information asymmetries in low- and middle-income countries. In such contexts, young workers are largely employed in the informal economy, young jobseekers have limited knowledge of job opportunities and there are insufficient information channels on job vacancies and career paths.
- **In high-income countries, studies of skills training interventions report the largest impacts, while wage subsidies and public works have only marginal impacts on average.** The evidence is most pronounced for impacts on earnings outcomes. This finding is in line with an earlier meta-analysis of Card et al. (2018) based on ALMPs targeting adult workers, showing that human capital centred ALMPs achieve larger impacts, particularly in the long-run.

Fourth, the “how” an intervention is designed and implemented is an important determinant of better labour market outcomes.

- **In low-and middle-income countries, studies of comprehensive, multipronged ALMPs and those that certify participation report larger impacts on youth labour market outcomes.** ALMPs that offer various services are able to better address the many constraints young people face. This effect is also evident and strongly significant among programmes that integrate soft skills and those that provide certification to participants. An important factor in low- and middle-income countries seems to be that ALMPs enable young jobseekers to better signal their competencies and employability to employers.
- **Interventions in low- and middle-income countries are more successful when the programme duration exceeds four months.** One reason may be that youth in developing contexts face a larger number of different (though often interlinked) labour market barriers and require more intensive interventions.
- **In high-income countries, larger impacts are reported for interventions that are implemented only by public stakeholders as compared to those under public-private engagement.** By contrast, in low- and middle-income countries, public-private partnerships are linked to larger impacts. While less consistent across specifications, this result may speak to the value of partnerships when reaching out specific populations or addressing multiple labour market constraints in developing country contexts.

Fifth, for “whom” impacts are reported matters: impacts are larger for youth from disadvantaged backgrounds, young women, and participants younger than 25 years, particularly in low- and middle-income countries. This result highlights the important role that youth-targeted ALMPs may play in assisting those youth groups severely and disproportionately affected by the COVID-19 crisis, who are more prone to be in informal employment or in inactivity.

Sixth, interventions in high-income countries and across the Africa region implemented at the local or regional levels produce better outcomes for youth than those with national scope. The finding suggests that interventions with a local or regional scope may be better adapted to specific youth labour market challenges. This offers promise for municipal or regional efforts to curbe the youth employment challenge, likely through better targeting and outreach strategies.

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

Conclusion

The systematic review and meta-analysis come at a time when the COVID-19 pandemic has exacerbated the already arduous transition of young people to the labour market worldwide. The crisis pushed youth disproportionately into unemployment, inactivity or lower-quality jobs. Considerable learning losses, including through disruptions in education and training, and fewer opportunities to engage in employment or work-based learning call for substantially upscaling interventions and investments in youth.

Against this backdrop, this report provides the most comprehensive analysis of youth employment interventions to date. The analysis synthesizes the evidence of more than 30 years of impact evaluations. The findings aim to support decision-making by policymakers and practitioners on the allocation of scarce resources for promoting youth employment with a view to unpacking “what” works across different country contexts and developing more cost-effective programmes (“how”).

The report concludes that ALMPs are effective policy measures to improve labour market outcomes of youth, particularly those hardest hit by the COVID-19 recession, and urges for further investments in youth. The analysis shows that ALMPs are well-suited to enhance employment and earnings of young people across the globe. They offer particularly positive prospects for youth in low- and middle-income countries as well as for youth groups at the greatest risk of labour market exclusion. The findings support the case for urgent and targeted investments in youth-targeted ALMPs to address the heightened youth employment challenge.

► 1 Introduction

The COVID-19 pandemic severely affected decent job opportunities and labour market outcomes for young people, exacerbating pre-existing inequalities and stalling the progress of the past years on the transition of young people into decent work. Some youth employment targets of the 2030 Agenda for Sustainable Development are now off track, and their outlook is rather dim.

The recession and job crisis induced by the pandemic have impacted young people through job and income loss, greater barriers in the transition to work, and disruptions in education and training. Between 2019 and 2020, 34 million young workers lost their job, with young women shouldering the worst of the youth employment crisis. The global youth employment loss of 8.2 per cent sharply compares to the employment loss of 3.6 per cent among adults, thus emphasizing the disproportionate effect of the recession on youth (ILO 2022). As economic activity and businesses came to a halt at the start of the pandemic, firms were more likely to dismiss young workers due to their limited experience and seniority or less-stringent contractual arrangements. Government relief measures had an insufficient focus on young workers, and recent studies have shown the narrow ability of those responses to protect them (Barford et al. 2022; Eichhorst et al. 2022). Moreover, in such uncertain times, firms were less likely to hire new workers, posing additional entry barriers for young people.

The COVID-19 pandemic translated into both unemployment and labour force exit (ILO 2021a). In 2020, global youth unemployment increased by nearly 4 million persons, while the number of young people not in employment, education or training (NEET) reached a record high of 282 million. The gender gap has continued and exacerbated in some regions, with young women twice as likely as young men to have NEET status and 1.5 times less likely than young men to be employed (ILO 2022).

Although economies around the globe have made up some of the losses, young people have lagged behind in the labour market recovery and have experienced only slow progress (OECD 2022a; Schady et al. 2022). By the end of 2022, global youth unemployment is projected to improve slightly, remaining at nearly 6 million unemployed persons above the 2019 level (ILO 2022). The pandemic also triggered a learning crisis, disrupting education and training systems all over the world. In 2020, around 98 per cent of technical and vocational education and training institutions experienced at least partial closure due to the pandemic (ILO 2021b). A global online youth survey conducted between April and May 2020 also showed that despite the best efforts of schools and training institutions to provide continuity of education through online learning, one in eight students aged 15–29 was left without any access to courses. And 65 per cent of them reported having learned less since the onset of the COVID-19 pandemic (ILO 2020a).

The jobs and learning losses signify an urgent call for youth interventions because the pandemic and further ongoing crises may have long-lasting, and for now still largely hidden, effects. When it comes to employment, the lesson from past economic shocks is that full recovery to pre-crisis trends is often a myth and leads to scarring, particularly among youth (ILO 2022; Gregg and Tominey 2005). Beyond the scarring of workers who lose their job or switch to worse employment conditions, crises and slowdowns can have long-term negative effects on labour market trajectories and the experience of new entrants to the labour force (Kahn 2010). Such adverse impacts on employment and skills development are confounded by the deterioration of the mental well-being of young people (ILO 2020a), foreshadowing a protracted youth employment crisis (Schady et al. 2022). Targeted and urgent investments in youth employment and skills development are therefore crucial to foster successful transitions into decent work.

Active labour market programmes (ALMPs) have been promoted as an essential component of the COVID-19 recovery plans (Asenjo et al. 2022; ILO 2022; ILO 2020b). Their effectiveness on youth labour market outcomes was documented by Kluge et al. (2017 and 2019¹), demonstrating positive impacts on both employment and earning

¹ Kluge 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 with the World Development Journal, built on the previous report and added additional analyses emerging from multivariate meta-regressions.

outcomes. While positive, the effect of ALMPs to date generally has been deemed small, varying greatly across intervention types and driven by contextual and design features.²

A recent global review of COVID-19 employment policies found that across 224 countries and territories, more than half introduced at least one ALMP, such as job placement, training, or entrepreneurship support, in response to the pandemic. Of these countries, 36 per cent adapted an existing programme and 79 per cent introduced new policies (Gentilini et al. 2022). Immediate recovery strategies that leveraged the ALMPs to assist young people during the first two years of the COVID-19 crisis include the Kickstart and Restart Programme in the United Kingdom, the Rural Employment Guarantee Scheme in India and the Short-time Work Programme in Germany (Barford et al. 2022). These interventions have not yet been evaluated, although in the case of the Kickstart Programme, there were early signs of undersubscription.³ Further research will be needed to assess impact as a response to the COVID-19 crisis.

Recognizing the great focus that policymakers are placing on policies and programmes to accelerate the COVID-19 recovery and the urgency in alleviating the employment challenges of young people, this updated systematic review incorporated the empirical evidence of the past three decades on what works in youth employment and examined the effectiveness of youth-targeted ALMPs to improve labour market outcomes of beneficiaries. The analysis sought to assess the role that context, youth characteristics, evaluation features, programme design and implementation had in moderating the impact on labour market outcomes.

The review upgraded the sample of studies in Kluge et al. (2017 and 2019) and refined the inclusion criteria to focus only on youth-targeted ALMPs, considering young people as those aged 15–35. Eligible interventions comprised skills training, entrepreneurship promotion, employment services, wage subsidies and public works programmes. Outcomes of interest included employment, earnings and business performance. Eligible studies included counterfactual-based impact evaluations, both published and unpublished that were conducted in low-, middle- or high-income countries between 1990 and 2022.

A comprehensive systematic search for relevant evidence across 40 sources, using search terms in English, French, German, Portuguese and Spanish, led to more than 21,000 records that were screened. The search process was completed in March 2022, leading to 228 reports (or impact evaluation papers) assessing 220 interventions within 171 separate ALMPs across 62 countries. In total, information on 5,051 treatment effect estimates were extracted from these primary reports. Additionally, detailed information regarding intervention design and implementation, beneficiary characteristics and evaluation methods was coded.

As a result, the novel systematic review more than doubled the Kluge et al. 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 for the systematic analysis.

To render treatment effect estimates comparable across reports, standardized mean differences (SMDs) were computed following Hedges and Olkin (1985). Hedges' *g* is a measure of effect size and captures the relative magnitude of the treatment effect, in comparison to a counterfactual, in a dimensionless way and adjusts for potential bias from small sample studies. The required information to compute Hedges' *g* was available for 4,225 of the 5,051 treatment effect estimates.

Random-effects meta-analysis methods were employed to synthesize and compare the SMDs reported in the primary reports. Subsequently, multivariate meta-regression models were estimated using both random-effects and weighted least squares methods, and information about intervention-level, study-level and country-level characteristics were included to assess factors associated with the magnitude of reported effect size estimates. This review improved upon the meta-analytical approaches that Kluge et al. (2017 and 2019) employed in several ways, implying that the results are not strictly comparable.

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

³ See www.ft.com/content/d33d6a8a-9c96-4e83-a4bd-1616d38666c0.

The systematic review found that ALMPs have been effective in improving labour market outcomes for young people. The overall effect size across all outcomes is 0.08 SMD with a 95 per cent confidence interval (CI) of 0.07 SMD – 0.09 SMD. The summary effect size for employment outcomes is 0.07 SMD (95 per cent CI: 0.06 SMD – 0.09 SMD), 0.08 SMD for earnings outcomes (95 per cent CI: 0.06 SMD – 0.10 SMD) and 0.12 SMD for business performance outcomes (95 per cent CI: 0.07 SMD - 0.17 SMD). Benchmarking these findings to recent meta-analyses of social science interventions shows progress in the impact of ALMPs on youth labour market outcomes, as an SMD of 0.05 and above is considered of “medium” size. Which, with improved methods, are larger than those that Kluge et al. (2017 and 2019) found in a sample of studies from 1990 to 2014 (namely, an overall SMD of 0.04, with a CI ranging between 0.02 SMD and 0.06 SMD).

On average, ALMPs are more effective in low- and middle-income countries than in high-income countries. In the former, impacts on earning outcomes are on average higher than on employment and business performance outcomes, while the reverse is true in the latter.

The review findings highlight 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 focus on delivering entrepreneurship promotion interventions – and to some extent also those focusing on employment services – have been significantly more effective than other types of interventions in low- and middle-income countries. Overall, skills training interventions yield better labour market outcomes than other intervention types in high-income countries.

As regards to the “how”, interventions that included a component of soft skills training appear more successful in most settings, as do programmes that provided certification to their participants, pointing to the importance of signalling strategies. The effects from youth employment programmes in low- and middle-income countries appear to differ in other important ways from those in high-income countries: ALMPs in low- and middle-income countries are considerably more successful when multiple components are combined, when the programme duration exceeds four months, and when implementation is carried out through a partnership between public and private actors, such as the collaboration between a public employment service agency and a company or an NGO. Interventions implemented across Africa are on average as effective as interventions in other low- and middle-income countries, yet there is no evidence that success is driven by adding soft skills components or providing certification to participants.

The sub-sample analysis of youth from disadvantaged backgrounds, young women, and participants younger than 25 years reports larger effect sizes, particularly in low- and middle-income countries. This result places confidence on the ability of the ALMPs to assist COVID-19 recovery measures to integrate the hardest hit youth groups into the labour market.

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

Consistent with previous reviews, 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 was not systematically correlated with better effects. As regards to 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 low- and middle-income countries, effects do not increase 12 months after the end of the intervention, however they may materialize within that period.

The costs analysis showed a welcome increase in the availability of cost information and cost-benefit analyses, which is more pronounced since the mid-2010s. The trend is observed mainly among skills training interventions. Nearly three out 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 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 organized as follows: Section 2 explains the inclusion criteria and important adjustments made to the review's population, interventions, comparisons and outcomes, in

comparison with Kluge et al. (2017 and 2019). Section 3 covers the search strategy and data-collection process. The characteristics of the studies and interventions included in the review are discussed in Section 4, which also depicts a description of interventions and studies from Africa. Section 5 presents the meta-analysis methods, while Section 6 shows the results from a univariate unconditional approach as well as through multivariate meta-regressions. It also offers findings for Africa and includes a discussion on programmes' costs and the availability of cost-benefit analyses. Section 7 closes with conclusions from the study 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 to 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 the 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 2022b), 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 matching closely 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. 2017 and 2019	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 Organisation for Economic Co-operation and Development 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 the policy or programme intention and corresponding mechanisms set in place for its implementation. In other words, while a programme might pursue 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 2021c). 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 Kluge et al. (2017 and 2019) is 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 reduce employers' social security contributions, those that reduce the labour or wage cost and those that provide 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 the unemployed, provide supplementary employment to the underemployed and generate income and therefore smooth or boost consumption patterns.

Note: Kluge et al. 2017 and 2019 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 that 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 (Kluge et al. 2017 and 2019; 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 Kluge et al. (2017 and 2019) used.

2.3 Comparison group

This systematic review included studies that measure change in at least one outcome of interest among intervention participants and relative to non-intervention participants based on a counterfactual analysis (comparing treated and control groups). Comparison and control groups must have included those 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
Business performance	Sales
	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 and (iii) methods for causal inference under unconfoundedness (regression methods, statistical matching, propensity score matching) and (iv) selection on 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 search

The systematic review focused on expanding the sample of reports from Kluge et al. with evaluations published after 2014. Subsequently, the review searched for new versions of reports and papers previously included by Kluge 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 by 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 Kluge 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 Kluge et al.) led to the identification of 310 keywords⁴. 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. Keyword and controlled vocabulary terms

Controlled vocabulary set	Keywords for studies from Kluge 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* allows the search to capture instances where the text is “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 Kluge et al. 2017.

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

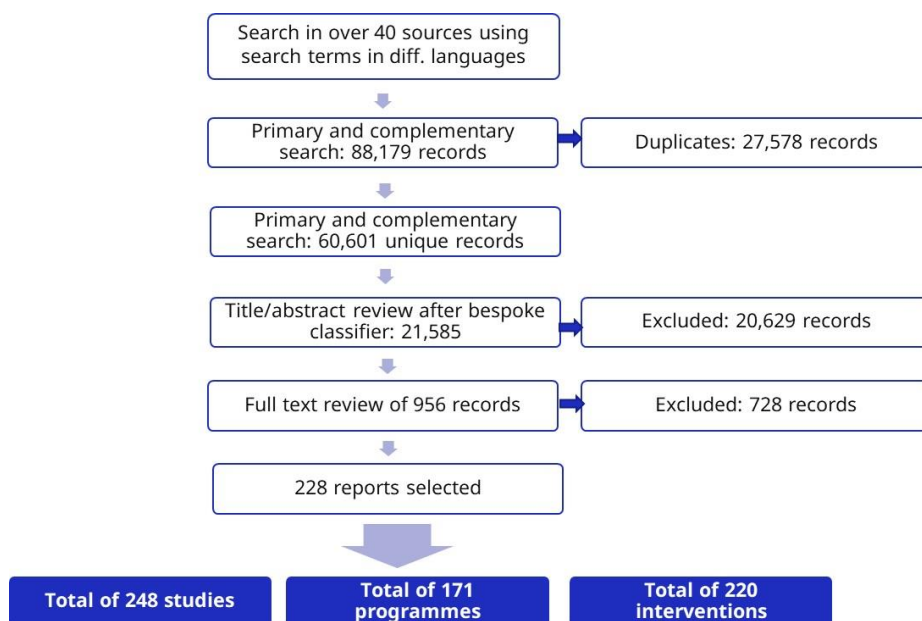
Thirty-four primary search sources were identified, starting with all sources with at least one of the 113 included studies in Kluge et al. An initial focus was placed on records with a publication date of 2013 or onwards, creating a small overlap with the base sample from Kluge et al.⁵ Subsequently, 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 4.0.

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 Kluge et al. (2017 and 2019); (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 meta-analysis by Kemper et al. (2022);⁷ (iv) 99 references provided by 41 of the 94 experts and authors of studies included in Kluge 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.

The primary and complementary searches resulted in **88,179 records** (references) based on **40 sources** from 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**



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, Abdul Latif Jameel Poverty Action Lab (JPAL) 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 forthcoming meta-analysis by Kemper, Stöterau, and Ghisletta (2022) leveraged the database from Kluge 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.

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 a "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.⁸ 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 were taken to the next stage and 21,585 records were double-screened by title and abstract.⁹ 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 more than double the number of those captured by Kluge et al. (at 113 reports).

3.2 Data collection

The data extraction followed the same protocol that Kluge 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.

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

⁸ 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 4.0 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).

⁹ In this edition of the review, the screeners encountered various records with title but not abstract. The search process led to the identification of studies in languages beyond those in the search protocol, such as Japanese, Korean and Hungarian, but with 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.

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 Kluge et al.

► 4 Study and intervention characteristics

4.1 Global overview

Table 5 provides an overview of the 248 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 Kluge et al. (2017 and 2019), 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**

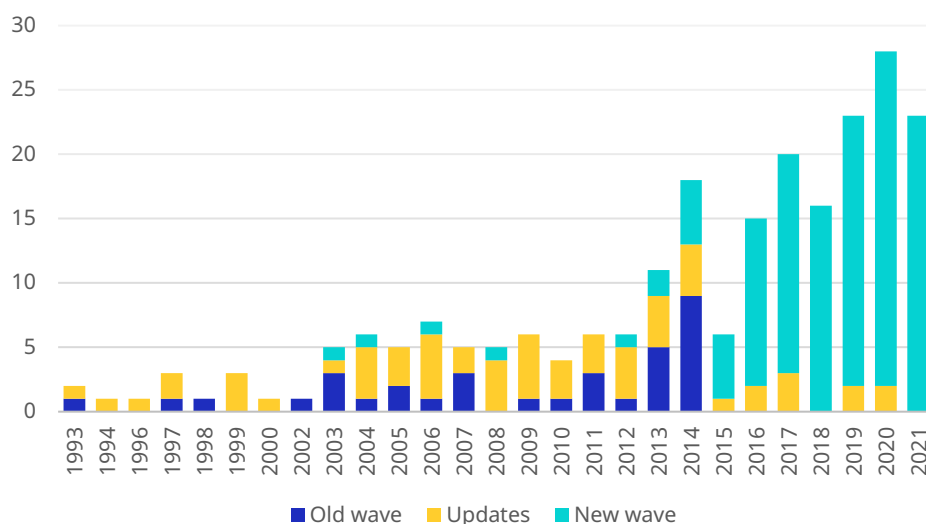
	Studies	%		Interventions	%
Publication period			Country income level		
1990-2000	10	4	High income	92	41.8
2001-2005	15	6	Middle Income	92	41.8
2006-2010	42	16.9	Low Income	36	16.4
2011-2015	47	19	Continent		
2016-2022	134	54	Africa	72	32.7
Publication status			Arab States	4	1.8
Peer-reviewed publication	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 (RCT)	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
Female and male together	184	74.2			

	Studies	%		Interventions	%
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 not be exclusive across the different typologies. For example, one study may estimate multiple outcomes or look 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, female and male together. Similarly, interventions may not be 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.

RCT = randomized controlled trial; 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**



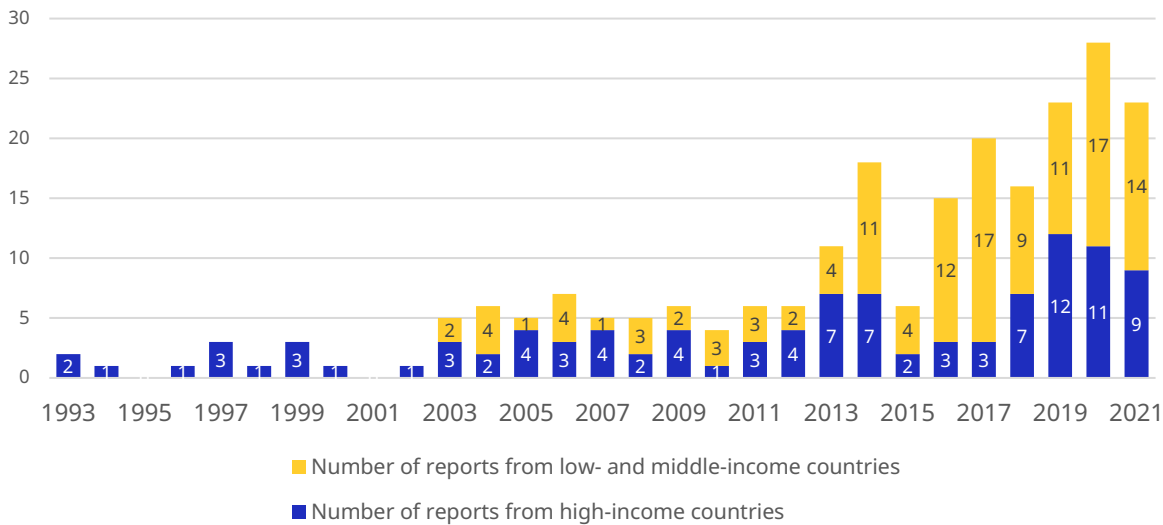
Note: “Old wave” refers to the studies included in Kluge et al. (2017 and 2019). “New wave” includes all new evidence that was not included in Kluge et al. “Updates” refers to more recent versions of studies that Kluge et al. included. While the period of study is 1990–2022, the figure shows the years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

When looking at the sample’s **geographical composition**, 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, as shown in figure 3), represents an important development in the impact evaluation field, compared with the 42 per cent share reported in Kluge et al. (2017 and 2019).

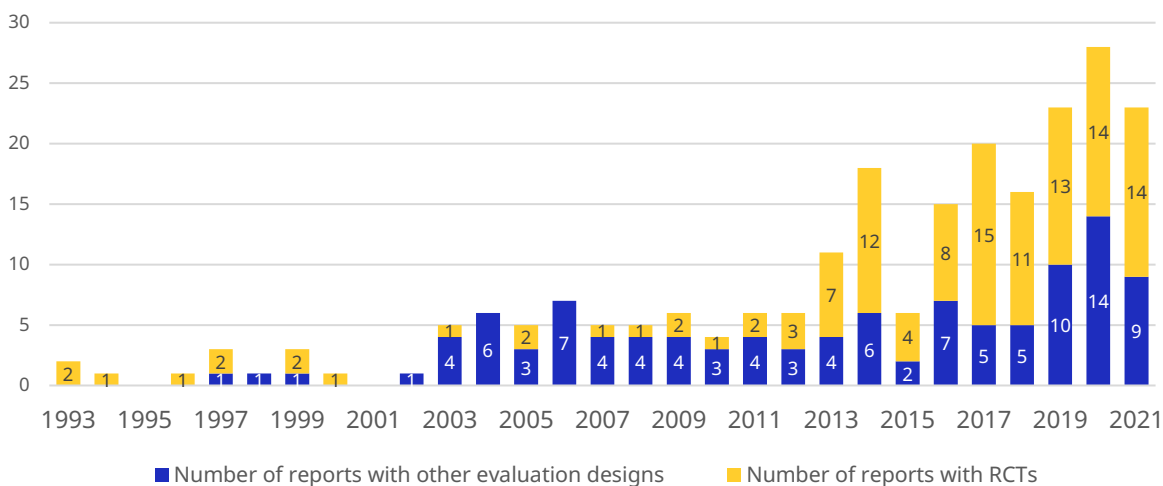
Regarding **publication status and study design**, the extensive search effort was also successful in identifying unpublished studies – the so-called “grey” literature. Around 38 per cent of the studies were from peer-reviewed journals, with the remainder covering technical reports from implementing organizations, working papers and other publication types (thesis and dissertations). The number of experimental studies or randomized control trials (RCTs) has 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**



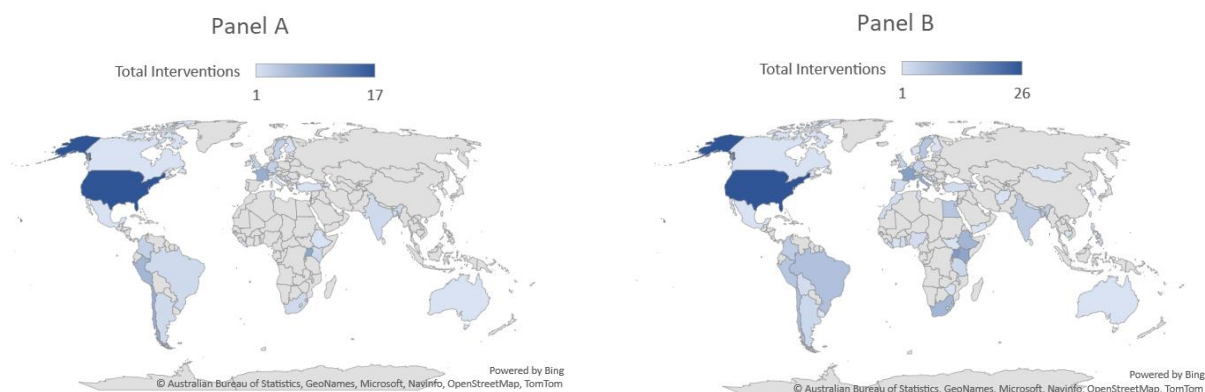
► **Figure 4. Increasing number of randomized control trial reports, 1993–2021**



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 skills-training interventions. Some studies provided a 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 Kluge et al. 2017 and 2019, and Card, Kluge and Weber 2018) that skills development interventions (as main category) have been 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 the classroom and 55 per cent at the 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 skill 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.

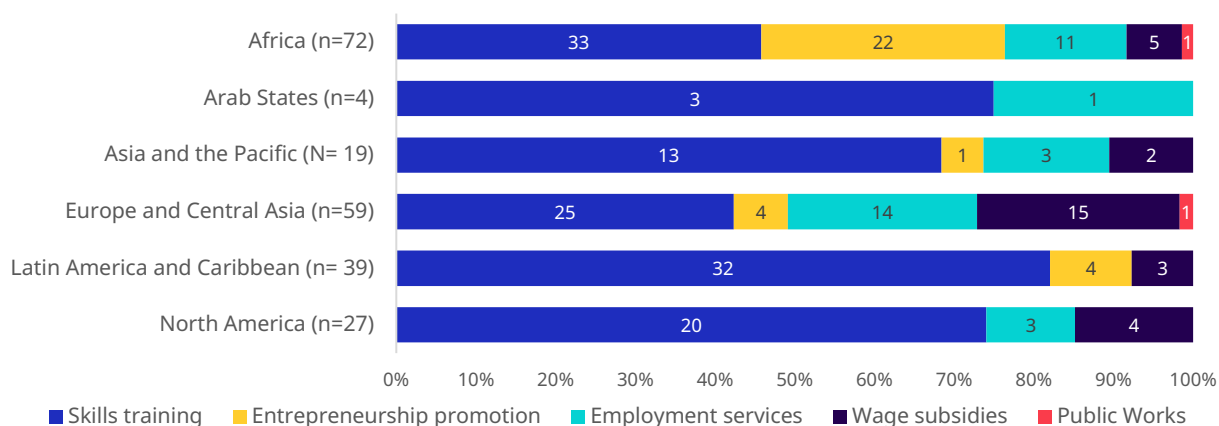
4.2 Regional differences

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 are 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 is relatively small.

Table 6 shows the frequency of intervention types as components within broader comprehensive designs. Skills interventions are 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 has some level of diversity across skills, entrepreneurship promotion and employment services, while in Europe and Central Asia the presence of entrepreneurship related interventions is rather limited.

► **Figure 6. Main categories of interventions, by region**



► **Table 6. Different components of interventions, by region**

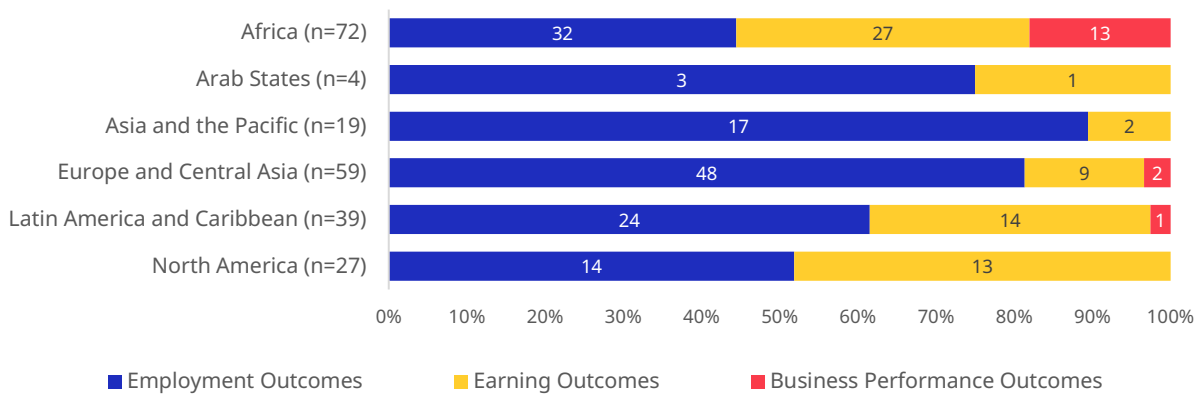
	Africa	Arab States	Asia & the Pacific	Europe & Central Asia	Latin America & Caribbean	North America	Total
Total interventions	72	4	19	59	39	27	220
Has skills training component	36	3	13	36	32	25	145
%	50.0%	75.0%	68.4%	61.0%	82.1%	92.6%	65.9%
Has entrepreneurship promotion component	32	0	2	4	4	1	43
%	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 out of 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 the largest 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, immediately followed by North America and then Latin America and the Caribbean (22 and 20 per cent, respectively).

The review of outcomes measured across regions shows 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.

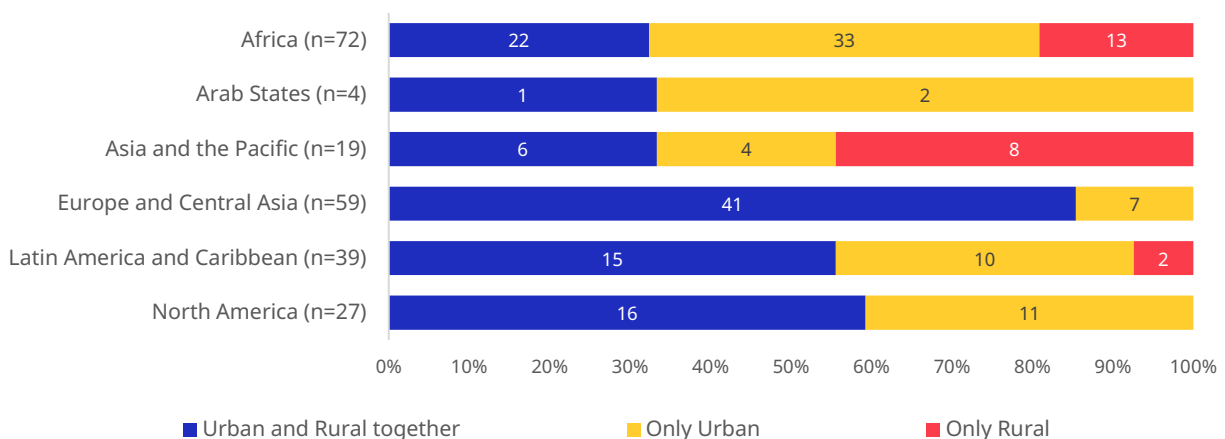
► **Figure 7. Main outcome construct evaluated for each intervention, regional differences**



Regional variation can be observed also when looking at programme financing. Interventions in Europe and Central Asia, as well as in Asia and the Pacific, are mainly financed entirely by public authorities (42 and 63 per cent, respectively). Yet, interventions in Africa (41 per cent) and Latin America and the Caribbean (33 per cent) are 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 (27 per cent), immediately followed by Latin America and the Caribbean (26 per cent). These interventions 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).

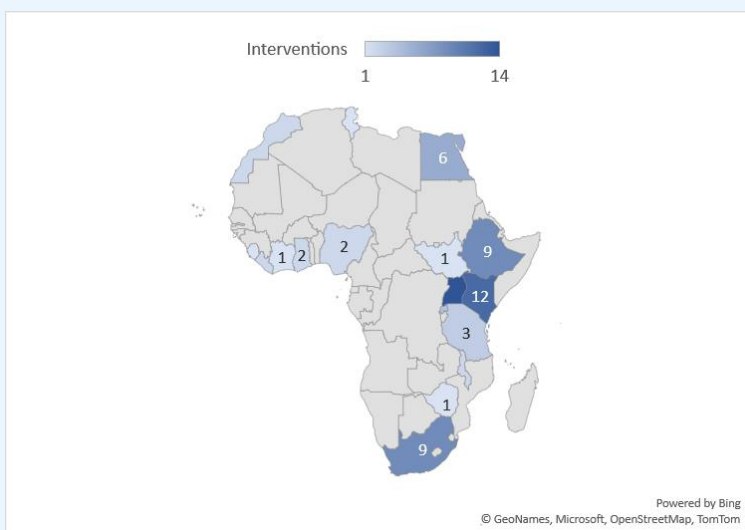
► **Figure 8. Urban versus rural focus, regional differences**



► **Box 1. Africa at a glance**

As shown in tables 5 and 6, 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 Western 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 (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			

► 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 RCT-based designs).¹⁰ 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):

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

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

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 $\Gamma()$ is the gamma function.¹¹

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):

¹⁰ 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).

¹¹ 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.¹² 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 (box 2). 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 2. 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 $i = (1, \dots, m)$ 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.

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

¹² 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).

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.¹³ 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 analysis was then conducted using the newly available suite of meta-analysis command in Stata 17. The summary effect sizes based on these models are presented alongside the 95 per cent confidence intervals in the forest plots in Section 6. In addition to the aggregate effect size, these forest plots display the weight each intervention carries towards the summary effect size. 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:

$$g_{ij} = \beta_0 + u_j + \varepsilon_{ij}$$

where g_{ij} is the i^{th} effect size estimate (SMD) from study j (for different cohorts or different outcome measures, for example). And $u_j \sim (0, \tau_j^2)$ 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.

5.4 Assessing correlates of effect size magnitude

Moderator analyses were performed when there was evidence of heterogeneity in summary effect size estimates based on random effects meta-analysis I-squared statistics. The analyses tested hypotheses of whether variation in the (average) effect sizes reported in studies was associated with differences in study, participant and intervention characteristics (moderators). The same moderators were also used in the multivariate meta-regression to assess correlations of effect size magnitude. This moderator analysis therefore formed the basis for the answers to the research questions regarding factors of intervention effectiveness.

In a first step, the review team conducted subsample analysis along dimensions that were considered as core moderators of effect size heterogeneity according to the theory of change outlined in Kluve et al. (2017). Specifically, the review team investigated heterogeneity *across* the main intervention category and country-income level *within* each outcome category. Ideally, moderator analysis should be conducted with a minimum of ten studies for each individual moderator variable (Borenstein et al. 2009). A decision was made to present forest plots for subgroups (such as intervention types) that had at least four individual interventions. The results from these subsample random effects meta-analysis are presented in Section 6.

In a second step, multivariate meta-regression models were estimated. This allowed the review team to assess factors that 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

¹³ 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".

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).

In the main analysis, an unrestricted weighted least squares (WLS) model was estimated, with weights as the inverse variance of each effect size estimate – a procedure that Stanley and Doucouliagos (2012) recommended. 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:

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

where Z_{jk} represents K ($k = 1, \dots, K$) moderator variables representing study-level or intervention-level heterogeneity. $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ 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, WLS models assume 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). The results from the multivariate random effects meta-regression model are presented side by side with the WLS results in section 6.4. These random effects models are more comparable to the results of the moderator analysis based on forest plots (aside from accounting for publication bias; see the next section). The random effects meta-regression model was estimated using a restricted maximum likelihood method:

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

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

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

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:

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

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).

► 6 Results

6.1 Summary effect sizes

The search process allowed the identification of 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 7, 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 (53 per cent) stemmed from studies that reported on randomized controlled trials, followed by quasi-experimental studies (41 per cent).

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

► **Table 7. 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	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 (RCT)	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 are 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 primary studies (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 final analysis consisted of 210 reports, reflecting 227 independent studies and the evaluation of 204 interventions across 157 programmes.

To summarize effect size estimates across studies, random effects meta-analysis was employed. The results are shown using forest plots in the following figures. Box 3 includes more information on how to interpret forest plots.

These forest plots show summary effect sizes after application of the preselection and aggregation method described in section 5. The summary effect sizes from forest plots do not account for publication bias. The results therefore differ from those in the multivariate meta-analysis presented in the next section, which deals with effect size dependency and publication bias. The results in this section should therefore be considered as descriptive evidence of the underlying data.

▶ Box 3. Reading a summary forest plot

This review presents effect size estimates and confidence intervals for the respective outcomes of interest of an intervention. This information is displayed in forest plots, which can be read as follows:

Each subgroup (such as intervention category) is represented by one line in the plot.

The average standardized mean difference (SMD) is reported under effect size, along with its corresponding 95 per cent confidence interval. The same information is represented graphically through a line. An SMD greater than zero indicates that, on average, the treatment group had a better outcome than the comparison group, which did not receive the treatment. This is considered a positive effect.

The overall effect estimate is reported at the bottom of the plot. Notes below each forest plot provide additional information on the composition of the sample for the forest plot.

Source: Adapted from Kluge et al. 2017.

Figure 9 shows results from the random-effects meta-analysis for subgroups for which a single effect size could be created for each study. Hence, the number of studies aligns with the overall number of studies for each subgroup. The following results emerged:

- ▶ The overall effect of youth employment programmes is positive, at 0.08 SMD, and statistically significant at the 1 per cent level (CI = 0.07 SMD, 0.09 SMD). This effect size is considered “medium” when compared to the social science meta-analyses of educational outcomes that Kraft (2020) conducted.¹⁴ And building on the new evidence and the refined selection criteria, the systematic review revealed a larger average effect size than the 0.04 SMD reported in Kluge et al. (2017 and 2019).

The first panel of figure 9 shows subsample analysis by **main category of intervention**. For all categories, average effect sizes are statistically significant and range from 0.04 SMD to 0.15 SMD.

- ▶ For programmes with skills training as the main category of intervention, the average effect size is 0.08 SMD (CI = 0.07 SMD, 0.10 SMD). This effect is slightly lower in magnitude than related studies, particularly Kemper, Stöterau, and Ghisletta (2022), who report 0.105 SMD, albeit using a different estimation procedure (robust variance estimation). But the impact is substantially larger than the 0.05 SMD for skills training in the predecessor study by Kluge et al. (2019, table 4).
- ▶ The average effect size of entrepreneurship promotion interventions is significantly larger, at 0.15 SMD, although with substantial heterogeneity (CI = 0.10 SMD, 0.21 SMD). The SMD is slightly larger than the 0.12 SMD reported in Kluge et al. (2019). Other meta-analyses of entrepreneurship-related interventions not restricted to youth have described positive effects of varying magnitude.¹⁵
- ▶ Studies of employment services and subsidized employment interventions reported smaller effects on average, which are 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. This is a marked increase from the findings in Kluge et al. (2019), which documented almost zero impacts of employment services and subsidized employment interventions (SMD = 0 and 0.02, respectively). The improved level and statistical significance of average impacts may relate to recent

¹⁴ Where small effects are defined as those below 0.05 SD, medium as those in the 0.05–0.20 range and large as those above 0.20.

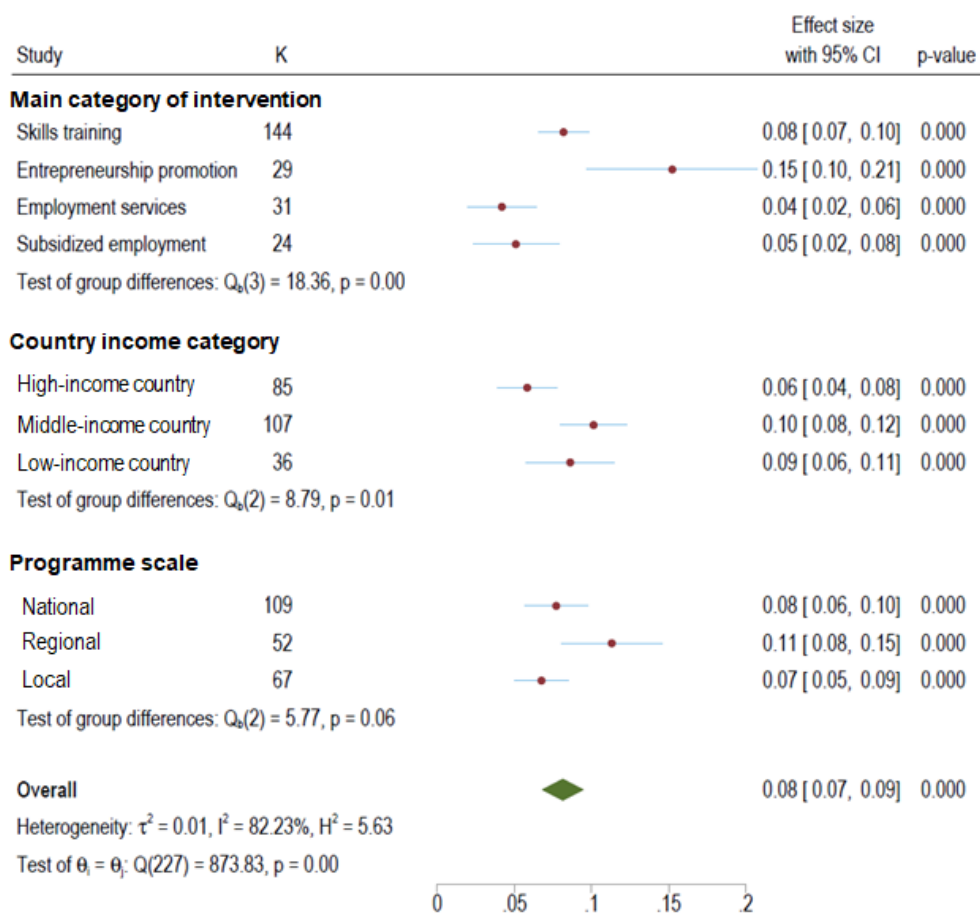
¹⁵ For a sample of evaluations of business training interventions for beneficiaries of all ages, McKenzie (2020) obtained an average impact of 10 per cent on business profits (not directly comparable to SMDs) with a 95 per cent CI of 4–16 per cent, and 5 per cent on sales (95 per cent CI: 0.2–9.2 per 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).

evidence or may suggest some progress across these intervention types in addressing targeting, design or implementation constraints.¹⁶

Figure 9 also shows results from the random effects meta-analysis for subgroups for which a single effect size could be created for each study. Hence, the number of studies aligns with the overall number of studies for each subgroup. The following results emerged:

- The overall effect of youth employment programmes is positive, at 0.08 SMD, and statistically significant at the 1 per cent level (CI = 0.07 SMD, 0.09 SMD). This effect size is considered “medium” when compared to the social science meta-analyses of educational outcomes that Kraft (2020) conducted.¹⁷ And building on the new evidence and the refined selection criteria, the systematic review revealed a larger average effect size than the 0.04 SMD reported in Kluge et al. (2017 and 2019).

► **Figure 9. Summary forest plot for the full sample: Intervention category, country income level and programme scale**



Note: The full sample was 4,060 (of 4,232) non-missing effect sizes that remained after censoring (SMD limit = 1, Inverse standard error limits = 1 to 100), with 2,151 of 4,060 effect sizes selected (1,909 effect sizes excluded as the higher level available). And 1,887 effect sizes were aggregated to 895 groups (intervention category, country income group and programme scale by study ID level).

The second panel of figure 9 divides effect sizes across **country income groups**, leveraging the increased number of studies and relying in particular on a broader evidence base from studies conducted in low- and middle-income countries (see figure 3).

¹⁶ Kluge 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.

¹⁷ Where small effects are defined as those below 0.05 SD, medium as those in the 0.05–0.20 range and large as those above 0.20.

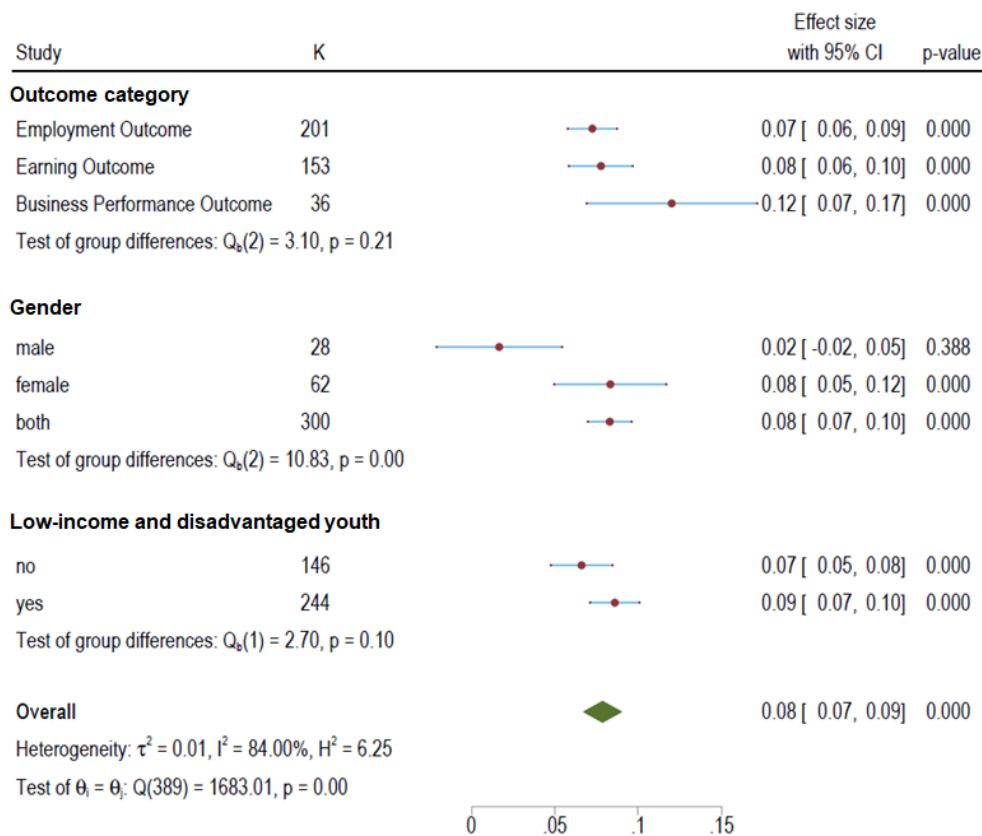
- The analysis shows the highest effects among evaluated ALMPs in middle-income countries, with an SMD of 0.10 (CI = 0.08 SMD, 0.12 SMD), followed by low-income countries, at 0.09 SMD (CI = 0.06 SMD, 0.11 SMD) and then high-income countries, at 0.06 SMD (CI = 0.04 SMD, 0.08 SMD). However, given that the confidence intervals overlap for each subgroup, the difference in the descriptive evidence is not statistically significant, at 5 per cent.

The last panel of figure 9 divides effect sizes by the **scale of implementation** to assess whether studies of local or regional programmes reported different effect sizes on average than those implemented at the national levels. The regional effect sizes are generally larger than the national effect sizes.

- Studies that measured the impacts of regional programmes reported larger effect sizes (SMD = 0.11, CI = 0.08 SMD, 0.15 SMD) than those focusing on a local scale (municipalities) (SMD = 0.07, CI = 0.05 SMD, 0.09 SMD) as well as national programmes (SMD = 0.08, CI = 0.06 SMD, 0.10 SMD).

Figure 10 shows results from the random effects meta-analysis for subgroups that could include several effect size estimates for a single study. For example, the same study may have reported effect size estimates for employment and earning outcomes as well as separately by gender for each of the outcome constructs. Hence, the number of studies does not align with the overall number of studies for each subgroup. Thus, the following results emerged.

► **Figure 10. Summary forest plot for the full sample: Outcome categories, gender and youths' income level**



Note: Grouped by outcome category, effects are measured across gender categories and across youths' income level. The full sample was 4,060 (of 4,232) of non-missing effect sizes that remained after censoring (SMD limit = 1, Inverse standard error limits = 1 to 100), with 2,151 of 4,060 effect sizes selected (1,909 effect sizes excluded as the higher level available). And 1,735 effect sizes were aggregated to 817 groups (outcome category, effects measured across gender categories and effects measured across youths' income level by study ID level).

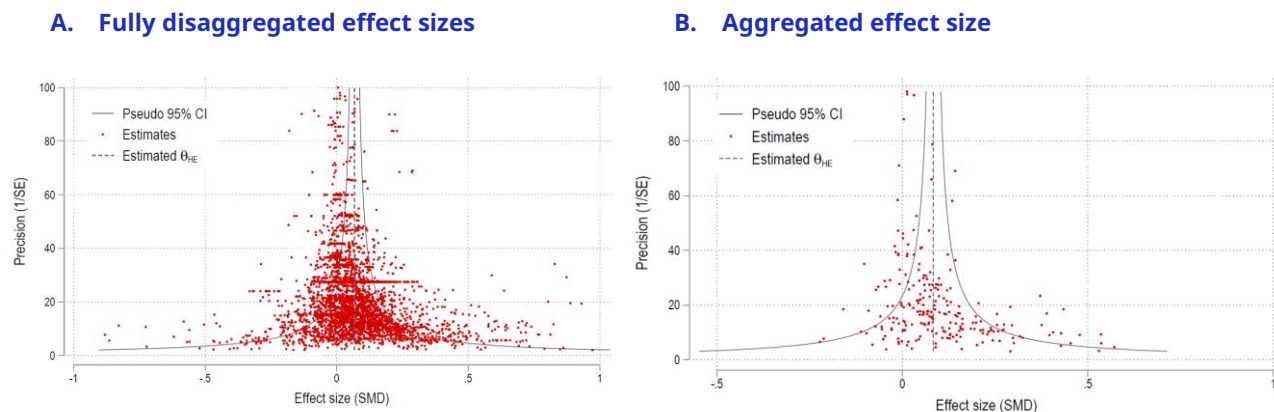
- The first panel of figure 10 divides effect sizes across **outcome categories**. Average effect sizes are positive and statistically significant for all subsamples. The overall effect of youth ALMPs on employment and earnings outcomes are similar, with 0.07 SMD and 0.08 SMD, respectively (compared to 0.05 SMD and 0.04 SMD in Kluge et al. (2019), table 4). Business performance effects of youth ALMPs exhibit greater heterogeneity, yet the SMD of 0.12 is statistically significant, with a 95 per cent confidence interval ranging from 0.07 SMD to 0.17 SMD.

- The second panel shows the univariate moderator analysis by **gender** and indicates that effect sizes reported for female participants (SMD = 0.08, CI = 0.05 SMD, 0.12 SMD) or for both genders combined (SMD = 0.08, CI = 0.07 SMD, 0.10 SMD) are larger on average than those reported for male participants only (SMD = 0.02, CI = -0.02 SMD, 0.05 SMD). The latter is, however, not statistically significant.
- The last panel presents effects sizes reported for young ALMP participants whether or not identified as low income or economically disadvantaged. The SMD for those identified as participants from low-income households are slightly larger, at 0.09 SMD (CI = 0.07 SMD, 0.10 SMD) than participants not identified as low income (SMD = 0.07, CI = 0.05 SMD, 0.08 SMD).

6.2 Publication bias

Figure 11 illustrates funnel plots for the pooled sample of effect size estimates. The dashed line refers to the unweighted mean effect size from the random-effects model. For reference, the 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 on the study level. The results are similar in quality and direction. As expected, more precise estimates generally centre more closely around the estimated true mean effect size. Figure 11 shows that the distribution of reported effect size estimates is 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.

► **Figure 11. Funnel plots**



Note: For the fully disaggregated effect size plot, the full sample was 4,060 (of 4,232) non-missing effect sizes remaining after censoring (SMD limit = 1, Inverse standard error limits = 1 to 100), with fully disaggregated effect sizes. For the aggregated effect size plot, the full sample was 4,060 (of 4,232) non-missing effect sizes remaining after censoring (SMD limit = 1, Inverse standard error limits = 1 to 100), with 2,151 of 4,060 effect sizes selected (1,909 effect sizes excluded due to a higher level available), with 1,909 effect sizes aggregated to 911 groups (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 that was estimated using the WLS procedure is significantly smaller than the random effects estimates shown in the forest plots (0.044 average effect from WLS, table 8, column 1 versus 0.08 SMD in figure 9).¹⁸ The FAT-PET test on the presence of publication bias in the pooled sample (table 8, column 2) suggests the presence of publication bias but little indication of a genuine effect beyond the reporting bias. The PEESE model therefore accounted for reporting bias using the standard error² (table 8,

¹⁸ 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.

column 3). This further reduced the summary effect size, compared to estimates from the model that did not account for publication bias (from 0.44 standardized difference to 0.37 standardized difference in the pooled sample).

► **Table 8. Test for publication bias**

	(1) No publication bias adjustment	(2) FAT-PET test	(3) PEESE model
Average effect	0.044	0.013	0.037
	0.000	0.311	0.000
Standard error		1.016	
		0.000	
Standard error²			4.897
			0.000
Estimates	4 060	4060	4060
Interventions	204	204	204
Reports	210	210	210
Studies	227	227	227

Note: Censoring SMD at +/-1 and inverse standard error at 1 and 100, Rho = 0.8, publication bias: clustered standard errors at study ID level.

6.3 Multivariate meta-regression results

The multivariate meta-regression introduces study- and intervention-level characteristics as covariates to understand how they drive the effectiveness of youth labour market interventions. The meta-analysis regresses the SMDs (see Hedge’s g in Section 5) on sets of explanatory variables that were selected based on data availability and the theory of change laid out in Kluge et al. (2017). To structure the discussion, the covariates were clustered in five groups.

- i. **Outcomes of interests:** namely employment, earnings and business performance. The review included an indicator variable for whether the outcome was measured conditional or unconditional on other (upstream or intermediary) outcomes. One example is whether earnings are measured and compared only for those in employment (conditional on employment) or also for individuals not in employment (unconditional).¹⁹ When breaking down models by outcome categories, only results for employment and earnings are presented, given the small number of observations on business performance outcomes.
- ii. **Country income level:** separating results for high-income countries from the low- and middle-income country results.
- iii. **Main category of intervention and design features of the intervention:** the main category of intervention and design features of the intervention. The latter includes availability of additional services, such as: combining components from various intervention types; the provision of soft skills as a programme component; the provision of incentives for programme participation (such as childcare, catering, transport, allowances or stipends or a salary); the existence of compliance mechanisms or monitoring of participants; the availability of certification of programme participation or completion; the duration of the intervention; the intervention’s scale (national, regional or local); and the involvement of public and private actors²⁰ in programme implementation.
- iv. **Characteristics of the programme participants:** capturing the impact estimates from studies that differentiate programme effects on disadvantaged youth (constructed as youth from low-income households

¹⁹ The conditional measured earnings would likely include lower or zero values for those who are not in employment (some individuals might still report some form of income, depending also on the study), whereas conditional measures exclude these individuals from the sample (thus tend to lead to higher average incomes).

²⁰ Private entities include civil society organizations and the private sector.

or with low educational levels); both female and male, only for young men and only young women; and those younger than 25 or aged 25 and older.

- v. **Characteristics of the study:** this includes whether the underlying report was peer reviewed, if the evaluation design reflected ITT or ATT, ATET and LATE estimates, if the evaluation design was experimental (RCTs) and whether the follow-up assessment was done at least one year after the baseline.

The analysis captured results in five specifications, both for weighted least squares regressions for Hedges' g and the random-effects model. The specifications increasingly add explanatory variables:

- **Specification I:** main category of intervention;
- **Specification II:** *specification I*, plus characteristics of the study and country income level;
- **Specification III:** *specification II*, plus outcomes of interest and characteristics of programme participants;
- **Specification IV:** *specification III*, plus intervention characteristics (extra services, soft skills component, incentives for participants, monitoring of participants, certification); and
- **Specification V:** *specification IV*, plus additional intervention characteristics (intervention duration, scale and implementing actor).

Results are presented for the entire sample (the "pooled sample" in table 9) as well as for the subsample of only employment or only earnings outcomes (table 10). The analysis divided the sample into studies conducted in high-income countries and studies from low- and middle-income countries (table 11). The tables show that the regressions rely on a decreasing sample as a larger number of covariates are introduced. This is a result of missing information due to many impact evaluations having provided limited information on programme design and participants' characteristics. For instance, for the pooled sample model presented in table 9, the analysis moved from 4,060 effect size estimates in specification I to 3,213 in specification V (a reduction of 21 per cent of the sample).

6.3.1 Outcomes of interest

The analysis found a similar effect size of ALMPs on employment and earnings outcomes. This is in line with the univariate analysis in section 6.1. However, for low- and middle-income countries, the earnings outcomes show a significantly larger effect size, at around 0.02 SMD, across all specifications (at the 1 per cent level of significance). Conversely, in high-income countries, employment and business outcomes show slightly larger effect sizes than earnings outcomes, although the differences are not stable across all specifications and are significant at the 5 and 10 per cent levels (table 11).

No differential effect is observed for conditional outcomes at the global level (pooled sample) or in low- and middle-income countries. In the high-income countries, outcomes measured conditional on another labour market indicator show considerably smaller effect sizes (with coefficients implying a reduction of between 0.03 and 0.05 SMD in the random-effects model). Given that the multivariate model simultaneously accounts for the outcome construct, this can be interpreted as the differences between overall earnings versus (hourly) wages (conditional on being employed) and employment probability versus job quality (conditional on being employed). Thus, the finding broadly indicates that ALMPs in high-income countries have slightly smaller impact on wages or job quality, in other words, the effects are more driven by the first-stage impact on employment probability. This is in line with the observation that the effect on employment (relative to earnings) is larger in high-income countries than in low- and middle-income countries.

6.3.2 Country income level

The analysis found that ALMPs in low- and middle-income countries have been more effective than in high-income countries. For the pooled sample (table 9), the interventions in the low- and middle-income countries appear more effective across all specifications, with the difference varying between 0.02 SMD and 0.04 SMD and confirming the findings of Kluge et al. (2017 and 2019) and Betcherman et al. (2007). Interestingly, there is a divergence between employment and earnings outcomes. For employment outcomes, the interventions in the low- and middle-income countries appear slightly more effective than in high-income countries (around 0.02 SMD and statistically significant in most specifications, table 10), while for earnings outcomes, the difference appears much larger, with a coefficient

between 0.04 SMD and 0.06 SMD (see table 10, specifications II, III and IV for both WLS and random effects models and specification IV for the random effects model).

6.3.3 Intervention characteristics

Overall, the results from the meta-regression suggest that the characteristics of interventions matter for reported effect sizes. Both the “what” (the main category of intervention) and the “how” (the intervention’s design and implementation) have important roles in determining an intervention’s success.

Looking at the “what”: When comparing against skills training, the interventions focusing on entrepreneurship promotion have resulted in larger effect sizes in the low- and middle-income countries. In the subsample of employment outcomes (table 10), entrepreneurship promotion interventions outperformed consistently with skills training interventions, by a margin of up to 0.07 SMD (table 10, specification V under the random effects model) and likely capturing, above all, an impact on self-employment. For earnings, the estimates are less stable, although a few random effects model specifications show a positive and statistically significant effect (between 0.03 SMD and 0.05 SMD for specifications III and IV). The strong role of entrepreneurship interventions becomes evident in the subsample analysis of low- and middle-income countries, where effect sizes are consistently positive, statistically significant and increasingly large – up to 0.13 SMD in the random effects model specification V (table 11). By contrast, entrepreneurship interventions appear less effective than skills training in the high-income countries, although the results are not always statistically significant. This is likely due to a lack of observations: only three of the 31 interventions with entrepreneurship promotion as main category came from high-income countries.

Employment services have not been any more or less effective than skills training, on average. However, there is some evidence that they have been more effective than skills training interventions in low- and middle-income countries, with up to 0.04 SMD in the random effects model specification V (table 11). In the case of subsidized employment interventions, that is, wage subsidies and public works, the estimates suggest that they are less effective than skills training, an effect most consistently found in high-income countries, by a margin of up to 0.07 SMD (see table 11, specification IV under the random effects model). Overall, the skills-training interventions appear to have delivered better labour market outcomes among youth in the high-income countries than the other intervention types.

Looking at the “how”: There is strong evidence to suggest that comprehensive, multipronged interventions have large positive effects on youth labour market outcomes, particularly in developing country contexts. The analysis found that extra services (apart from the major intervention category), additional components such as soft skills, certification mechanisms for youth participants and longer, more intensive interventions correlate with larger effect sizes (with some caveats, depending on the outcome category or country income subsample).

- **Multiple components:** In low- and middle-income countries, combining multiple components leads to more effective programmes. The coefficients on the indicator variable “extra services” are consistently positive (between 0.03 SMD and 0.04 SMD; see table 11) and statistically significant. In high-income countries, programmes with multiple components do not seem to have been more effective than programmes with a single component.
- **Soft skills:** Adding training components on soft skills associates with larger effect sizes. This finding holds true for the pooled sample and when dividing by outcomes category as well as by country income group (with an additional statistically significant effect between 0.01 SMD and 0.05 SMD). The finding deviates from Kluge et al. (2019), which found programmes with a soft skills component to be less effective than those without. The updated data from recent years suggest that adding soft skills elements in training curricula can be beneficial in boosting the labour market outcomes of youth.²¹
- **Certification:** Programmes that provide certification to participants perform significantly better than programmes without certification mechanisms in place. The result is stable across specifications and more pronounced in low- and middle-income countries, as well as for earning outcomes (particularly under the random effects model, where point estimates for programmes providing certification to participants reaches about 0.05 SMD and is statistically significant at the 1 per cent level). The finding supports the

²¹ The result does not represent an estimate for programmes that exclusively focus on soft skills.

argument made in Kemper et al. (2022) that ALMPs, notably skills training interventions in developing and emerging country contexts, enable jobseekers to signal their competencies and employability to employers.

- **Duration of intervention:** There is only slight evidence that interventions with longer duration yield better outcomes than short ones. Interventions with a duration of four to eight months are associated with better results, but this is largely driven by the subsamples of employment outcomes and low- and middle-income countries. Coefficients are only statistically significant in the random-effects model. Similar country income subsample results appear for programme duration of more than eight months. For high-income countries, programmes with a duration of four months or more do not show better results than programmes with a shorter duration.

Additional design characteristics, such as the geographic scope of a programme (national, regional or local), and implementers (public or private) appear to matter as well, although the findings apply only to specific country groups.

- **Geographical scope:** Interventions implemented at the local or regional levels produce better outcomes for youth than those with national outreach (up to an additional 0.03 SMD for the pooled sample, significant at the 1 per cent level, for the random effects model). The finding suggests that interventions with a local or regional scope may be better adapted to specific youth labour market challenges, yet it is important to stress that the result emerges from – and it is statistically significant only in – interventions in high-income countries.
- **Implementation by public and private sector actors:** Interventions that rely only on public stakeholders for their implementation (compared to public–private engagement) lead to better outcomes in high-income countries (additional effect of 0.03 SMD, statistically significant at the 1 per cent level). By contrast, in low- and middle-income countries, there is some evidence pointing to greater effectiveness of interventions implemented by public and private stakeholders together, although results are not always statistically significant: In the WLS model, private implementers reduce the SMD by 0.03 as compared to public–private arrangements (the coefficient is also negative for public implementers but not statistically significant). In the random effects model specification, programmes implemented by a public entity alone perform 0.03 SMD worse than interventions with public–private arrangements (with a negative, yet non statistically significant coefficient for ALMPs implemented only by the private sector).
- **Incentives:** Providing in-kind or monetary incentives to stimulate the participation of young people in ALMPs are not associated with better outcomes. In fact, the systematic review found a consistently negative and statistically significant effect for the pooled sample, as well as all other subsamples. The negative relationship between providing incentives and labour market outcomes could in part be mediated by the targeting criteria of participants: Programmes that provide additional incentives for participants might group young women and men that face additional labour market constraints. Further research is needed to open the bundle of incentives and understand the underlying transmission mechanisms, including the targeting and selection of youth beneficiaries.

When interpreting the relevance of these different correlates of effect size magnitude, it needs to be kept in mind that they are likely correlated among each other and hence pick up the general quality of an intervention. For example, it could be that programmes with certification require a validation of the intervention model (such as a curriculum) and/or include a more rigorous level of quality control. Certification may also be a sign that the programme relies on a “payment by results” instrument, by which the training provider is paid (for instance, by the government) contingent upon participants’ success in the programme.

6.3.4 Participant characteristics

The characteristics of the sample participants can also impact the effectiveness of ALMPs. Some groups of young people appear better placed to benefit from the interventions or, put differently, some ALMPs seem to be designed in ways that provide certain types of young people more opportunities than others.

The results suggest that ALMPs on average have larger effects on the earnings of youth from disadvantaged backgrounds. Coherent with previous results, this effect can only be observed for interventions in low- and middle-income countries. The multivariate regressions find differential treatment effects by gender and age. For the pooled sample, estimates only capturing results among young men are significantly lower than for combined or female only samples. By contrast, when results for young women are measured, the effect on labour market outcomes of youth is higher than for cases when results are reported for both gender or men only (around 0.01 SMD, statistically significant at the 1 and 5 per cent levels, depending on the model). The effect holds across both outcome subsamples and low- and middle-income countries.

Last, the regressions show that when results are measured for participants younger than 25 years, the effect sizes are consistently better than for participants 25 and older. Earnings outcomes appear more positive for younger youth as well as certain specifications for both high-income countries and low- and middle-income countries.

6.3.5 Evaluation and study characteristics

Experimental evaluations (RCTs) correlate with lower impacts. Point estimates range from a reduced impact associated with RCTs between 0.02 SMD and 0.07 SMD (statistically significant at least at the 5 per cent level). The result is comparable with other meta-analyses and appears consistent across outcome subsamples. However, it is mostly driven by studies conducted in low- and middle-income countries that relied on an experimental approach (the coefficient for high-income countries is negative but smaller and only significant in one specification (II in the random effects model in table 11).

Publications that were peer reviewed do not systematically outperform non-peer-reviewed studies (this analysis took into account study design). The same is true for intention-to-treat estimates, for which point estimates are close to zero across all specifications (only one specification in the low- and middle-income countries subsample suggests that studies with intention-to-treat correlate with lower impact).

For interventions implemented in high-income countries, effect sizes measured at least one year after the end of the intervention are significantly higher (between 0.02 SMD and 0.04 SMD in table 11), which is consistent with previous findings, such as Kluve et al. (2019) and Card, Kluve and Weber (2018).²² Their estimation results pointed to a pronounced temporal pattern of impacts increasing over time after the end of the programme. This review's finding is particularly pronounced for skills training interventions. However, for low- and middle-income countries, the meta-analysis found negative and statistically significant differential effects, suggesting that effects for developing countries do or do not materialize within the first year after exiting the programme but do not expand thereafter.

²² The Card et al. sample mostly included ALMPs not targeting youth.

► **Table 9. Meta-regression results, full sample**

	Weighted least squares Hedges' g regressions					Random effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V
SMD sampling variance	4.424***	3.014***	2.912***	2.744***	3.554***	2.512***	2.299***	2.468***	2.100***	2.838***
	0.002	0.001	0.002	0.002	0	0	0	0	0	0
<i>Mainly skills training</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Mainly entrepreneurship promotion	0.02	0.019	0.040**	0.047**	0.081*	0.01	0.012	0.034***	0.049***	0.083***
	0.382	0.353	0.037	0.017	0.064	0.237	0.216	0.009	0	0
Mainly employment services	-0.02	-0.01	-0.016	-0.008	-0.006	-0.015**	0.002	-0.004	0.014	0.007
	0.186	0.405	0.165	0.519	0.631	0.048	0.819	0.6	0.11	0.587
Mainly subsidized employment	-0.022	-0.027	-0.043***	-0.024	0.004	-0.035***	-0.030***	-0.034***	-0.013	0.003
	0.24	0.147	0.006	0.152	0.848	0	0.001	0.001	0.196	0.849
<i>Publication not peer reviewed</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Peer-reviewed publication		-0.004	-0.005	0.001	0.005		-0.001	0.001	0.002	0.005
		0.633	0.532	0.872	0.657		0.745	0.878	0.601	0.364
<i>Non-intention-to-treat estimate</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Intention-to-treat estimate		0	0.004	0.002	0.004		0.002	-0.001	-0.006	-0.005
		0.981	0.513	0.697	0.291		0.706	0.877	0.259	0.345
<i>Non-experimental design (IV, RDD)</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Experimental (RCT)		-0.024**	-0.025***	-0.031***	-0.037***		-0.033***	-0.032***	-0.038***	-0.040***
		0.019	0.004	0.001	0.001		0	0	0	0
<i>Measured before one year after exit from the programme</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Measured more than one year after exit from programme		0.007	0.002	0.003	0.006		-0.002	-0.003	-0.002	-0.003
		0.614	0.896	0.811	0.697		0.553	0.428	0.717	0.573
<i>Low- and middle-income country</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
High-income country		-0.040***	-0.032***	-0.031**	-0.015		-0.033***	-0.031***	-0.041***	-0.018***
		0	0.001	0.013	0.31		0	0	0	0.006
<i>Earnings outcome</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Employment outcome			-0.002	-0.002	-0.005			0.001	0	-0.004
			0.718	0.737	0.475			0.826	0.97	0.285
Business outcome			0.004	0.009	0.031			-0.016	-0.011	0.018
			0.813	0.58	0.279			0.27	0.465	0.318
<i>Unconditional outcome</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Conditional outcome			0.007	0.007	0.001			-0.001	0	-0.003
			0.358	0.445	0.956			0.854	0.941	0.611
<i>Pooled sample</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Sample of disadvantaged youth only			0.001	0.008	-0.003			-0.006	-0.002	-0.002
			0.927	0.456	0.84			0.328	0.79	0.826
<i>Both male and female participants</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Sample of male participants only			-0.001	0	-0.004			-0.012**	-0.012**	-0.014**
			0.91	0.992	0.717			0.036	0.029	0.016

	Weighted least squares Hedges' g regressions					Random effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V
Sample of female participants only			0.009	0.012**	0.013**			0.009*	0.010**	0.014***
			0.155	0.03	0.021			0.052	0.033	0.01
<i>Participants aged 25 and older</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Sample of participants younger than 25			0.015**	0.011*	0.005			0.012**	0.010*	0.004
			0.024	0.066	0.497			0.016	0.057	0.444
<i>Intervention without extra services</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
With extra services				-0.011	0.001				-0.007	0
				0.344	0.947				0.181	0.967
<i>Intervention without soft skills component</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
With soft skills component				0.008	0.021**				0.021***	0.036***
				0.299	0.025				0	0
<i>Intervention does not provide incentives for participation</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
Participation incentives				-0.02	-0.046***				-0.023**	-0.046***
				0.147	0.004				0.035	0
<i>Intervention does not monitor participation</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
Monitoring of participants				0.002	0.002				0.003	0.004
				0.876	0.826				0.51	0.509
<i>Intervention does not provide certification</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
Certification				0.019**	0.028***				0.033***	0.035***
				0.02	0				0	0
<i>Intervention shorter than 4 months</i>					<i>ref.</i>					<i>ref.</i>
4–8 months					0.012					0.020***
					0.412					0.003
8 or more months					-0.008					-0.01
					0.556					0.15
<i>Intervention with regional or local scale</i>					<i>ref.</i>					<i>ref.</i>
National scale					-0.034***					-0.027***
					0.002					0
<i>Implemented by private and public entities combined</i>					<i>ref.</i>					<i>ref.</i>
Public entity only					0.001					0.007
					0.889					0.312
Private entity only					-0.006					0.009
					0.657					0.179
Constant	0.042***	0.079***	0.065***	0.057***	0.069***	0.058***	0.090***	0.084***	0.075***	0.067***
	0.003	0	0	0	0	0	0	0	0	0
Estimates	4 060	3 909	3 665	3 665	3 213	4 060	3 909	3 665	3 665	3 213
Interventions	204	180	160	160	123	204	180	160	160	123
Reports	210	186	166	166	132	210	186	166	166	132
Adjusted R-squared	0.036	0.109	0.144	0.153	0.176					

Note: RDD = regression discontinuity design; RCT = randomized controlled trial.

► Table 10. Meta-regression results, subsample outcome category

	Subsample: Employment outcomes										Subsample: Earnings outcomes									
	Weighted least squares Hedges' g regressions					Random-effects SMD regressions					Weighted least squares Hedges' g regressions					Random-effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
SMD sampling variance	5.900***	5.407***	5.848***	5.795***	5.724***	4.243***	4.292***	4.631***	4.364***	4.479***	5.819**	1.478	1.188	0.592	0.86	2.530***	1.763***	1.931***	1.159*	1.482**
	0	0	0	0	0	0	0	0	0	0	0.045	0.315	0.466	0.699	0.635	0	0.005	0.002	0.067	0.036
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	0.035*	0.043*	0.06	0.017	0.029	0.040*	0.053**	0.073*	-0.006	-0.02	0.029	0.037	0.027	-0.011	-0.011	0.031*	0.052***	0.037
Mainly employment services	0.521	0.372	0.094	0.057	0.231	0.32	0.131	0.072	0.018	0.058	0.85	0.548	0.461	0.318	0.475	0.463	0.451	0.076	0.003	0.145
Mainly subsidized employment	-0.019	-0.011	-0.023*	-0.017	-0.007	-0.011	0.001	-0.011	0.002	0.015	-0.023	-0.011	-0.016	0.001	-0.039*	-0.030**	-0.01	-0.01	0.012	-0.027
	0.144	0.402	0.08	0.267	0.662	0.218	0.929	0.299	0.853	0.341	0.312	0.421	0.249	0.952	0.08	0.025	0.433	0.438	0.4	0.154
	-0.012	-0.007	-0.033**	-0.022	0.011	-0.017**	-0.009	-0.023*	-0.01	0.009	0.058***	-0.070***	-0.069**	-0.03	0.011	0.091***	-0.076***	-0.058***	-0.031*	0.006
	0.474	0.685	0.047	0.216	0.603	0.047	0.407	0.067	0.448	0.618	0.001	0.005	0.014	0.337	0.814	0	0	0	0.06	0.771
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	-0.001	0.003	0.009		0.002	0.001	0.001	0.007		-0.011	-0.004	0.003	0.01		-0.009	-0.004	-0.002	0.006
		0.757	0.92	0.747	0.497		0.736	0.906	0.838	0.37		0.451	0.786	0.836	0.543		0.148	0.501	0.78	0.392
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	0.003	0.003	0.009		-0.005	-0.005	-0.008	-0.001		0.012	0.005	0.001	-0.003		0.011	0.004	-0.005	-0.011
		0.656	0.669	0.665	0.16		0.423	0.42	0.225	0.895		0.164	0.435	0.897	0.608		0.115	0.59	0.449	0.118
Non-experimental design (IV, RDD)		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Experimental (RCT)		-0.014	-0.016*	-0.021**	-0.020**		-0.020***	-0.016**	-0.021***	-0.020**		-0.036	-0.039	-0.047*	-0.072***		-0.041***	-0.047***	-0.054***	-0.070***
		0.219	0.085	0.047	0.024		0.002	0.02	0.006	0.027		0.152	0.126	0.063	0.003		0	0	0	0
Measured 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 more than one year after exit from programme		-0.002	-0.011	-0.01	-0.005		-0.006	-0.010*	-0.008	-0.005		0.016	0.018	0.02	0.023		-0.003	0	0.001	0.001
		0.868	0.255	0.29	0.609		0.276	0.085	0.174	0.418		0.483	0.467	0.411	0.387		0.649	0.947	0.83	0.844
Low- and middle-income country		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
High-income country		-0.023**	-0.014	-0.012	0		-0.016***	-0.017***	-0.025***	-0.005		-0.063***	-0.052***	-0.065***	-0.051		-0.054***	-0.046***	-0.062***	-0.037***
		0.022	0.162	0.379	0.979		0.005	0.004	0	0.589		0	0.006	0	0.143		0	0	0	0.001
Unconditional outcome			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.			ref.	ref.	ref.
Conditional outcome			0.012	0.01	0			0.003	0.003	0			-0.001	-0.001	-0.005			-0.012*	-0.006	-0.008

	Subsample: Employment outcomes										Subsample: Earnings outcomes									
	Weighted least squares Hedges' g regressions					Random-effects SMD regressions					Weighted least squares Hedges' g regressions					Random-effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
			0.111	0.219	1			0.712	0.731	0.993			0.901	0.955	0.645			0.05	0.287	0.215
<i>Pooled sample</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Sample of disadvantaged youth only			-0.007	-0.004	-0.02			-0.020**	-0.017**	-0.023**			0.013	0.028**	0.049**			0.021**	0.026***	0.038***
			0.556	0.764	0.19			0.012	0.041	0.033			0.39	0.033	0.042			0.024	0.006	0.001
<i>Both male and female participants</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Sample of male participants only			-0.001	0	-0.005			-0.011	-0.011	-0.014*			-0.004	-0.003	-0.006			-0.008	-0.007	-0.009
Sample of female participants only			0.927	0.983	0.695			0.149	0.146	0.073			0.679	0.792	0.606			0.294	0.33	0.202
			0.008	0.01	0.01			0.011*	0.012*	0.012*			0.015*	0.017**	0.019*			0.018***	0.020***	0.025***
			0.263	0.116	0.116			0.095	0.075	0.094			0.08	0.034	0.057			0.007	0.004	0
<i>Participants aged 25 and older</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Sample of participants younger than 25			0.008	0.009	0			0.003	0.003	-0.007			0.022**	0.005	-0.004			0.013*	0.004	0.001
			0.305	0.243	0.982			0.635	0.655	0.401			0.021	0.678	0.801			0.063	0.541	0.862
<i>Intervention without extra services</i>			<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>	
With extra services				-0.007	0.01				0	0.016*				-0.012	-0.025				-0.015*	-0.024**
				0.578	0.383				0.961	0.052				0.551	0.255				0.08	0.017
<i>Intervention without soft skills component</i>			<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>	
With soft skills component				0.008	0.020*				0.013*	0.026**				0.012	0.01				0.033***	0.040***
				0.463	0.062				0.065	0.002				0.456	0.508				0	0
<i>Intervention does not provide incentives for participation</i>			<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>	
Participation incentives				-0.022	-0.047***				-0.030**	0.053**				-0.006	-0.038				-0.008	-0.038**
				0.106	0.004				0.041	0.002				0.803	0.148				0.602	0.032
<i>Intervention does not monitor participation</i>			<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>	
Monitoring of participants				-0.004	-0.002				0.005	0.005				0.015	0.012				0.011*	0.011
				0.758	0.896				0.481	0.537				0.288	0.396				0.087	0.13
<i>Intervention does not provide certification</i>			<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>	
Certification				0.01	0.022*				0.019***	0.024**				0.033***	0.049***				0.040***	0.049***
				0.316	0.05				0.008	0.003				0.006	0.002				0	0

	Subsample: Employment outcomes										Subsample: Earnings outcomes									
	Weighted least squares Hedges' g regressions					Random-effects SMD regressions					Weighted least squares Hedges' g regressions					Random-effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
<i>Intervention shorter than 4 months</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>
4-8 months					0.017					0.022** *					-0.03					-0.012
					0.318					0.009					0.362					0.257
8 or more months					-0.016					-0.025**					0.013					0.005
					0.268					0.01					0.644					0.617
<i>Intervention with regional or local scale</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>
National scale					-0.037***					0.025** *					-0.043*					-0.042***
					0.006					0.003					0.07					0
<i>Implemented by private and public entities combined</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>
Public entities only					-0.007					-0.008					0.004					0.008
					0.709					0.423					0.616					0.379
Private entities only					-0.018					-0.01					0.03					0.027***
					0.19					0.288					0.116					0.006
Constant	0.039***	0.061***	0.060***	0.053***	0.067***	0.051***	0.072***	0.087***	0.079***	0.076** *	0.041**	0.101***	0.066***	0.061***	0.086***	0.060***	0.108***	0.078***	0.071***	0.080***
	0	0	0	0.001	0	0	0	0	0	0	0.035	0	0.008	0.003	0.009	0	0	0	0	0
Estimates	2 391	2 296	2 146	2 146	1 925	2 391	2 296	2 146	2 146	1 925	1 382	1 354	1 290	1 290	1 185	1 382	1 354	1 290	1 290	1 185
Interventions	182	162	142	142	114	182	162	142	142	114	133	123	108	108	85	133	123	108	108	85
Reports	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	0.094	0.097	0.117						0.055	0.275	0.307	0.331	0.366					

Note: RDD = regression discontinuity design; RCT = randomized controlled trial.

► **Table 11. 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					Weighted least squares Hedges' g regressions					Random-effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
SMD sampling variance	6.107** *	5.887** *	5.413** *	4.605**	3.854*	5.052** *	4.358** *	4.716** *	4.174** *	4.533** *	1.881**	2.452**	2.772**	2.454**	3.629***	1.400***	1.781***	2.188***	1.786** *	2.795***
	0.001	0.001	0.005	0.012	0.054	0	0	0	0	0	0.038	0.01	0.01	0.021	0.001	0	0	0	0	0
Mainly skills training	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Mainly entrepreneurship promotion	-0.026	0.057	-0.016	-0.014	0	-0.040*	0.029	-0.071	-0.062		0.017	0.021	0.048**	0.083***	0.113**	0.015*	0.016*	0.045***	0.087** *	0.132***
	0.228	0.192	0.477	0.451	.	0.074	0.526	0.25	0.309		0.385	0.317	0.024	0	0.011	0.089	0.086	0	0	0
Mainly employment services	-0.005	-0.001	-0.031*	-0.027	-0.03	0	0.006	-0.024*	-0.016	-0.049**	-0.017	-0.001	0.005	0.017	0.038	-0.012	0.005	0.008	0.029**	0.040**
	0.618	0.92	0.052	0.134	0.307	0.992	0.615	0.066	0.259	0.034	0.221	0.938	0.814	0.447	0.114	0.269	0.678	0.547	0.04	0.016
Mainly subsidized employment	0.005	-0.015	0.063** *	-0.035*	0.001	-0.007	-0.026**	0.067** *	-0.038**	-0.021	0.057** *	-0.052**	-0.008	0.01	0.028	0.049***	-0.029**	0.001	0.026	0.011
	0.79	0.513	0	0.061	0.953	0.457	0.032	0	0.017	0.406	0.002	0.016	0.745	0.723	0.269	0	0.042	0.962	0.129	0.627
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	-0.011	0.007	0.006		0.019** *	-0.017**	0.001	0.002		0.01	0.013	0.004	0.002		0.016***	0.020***	0.010*	0.01
		0.123	0.143	0.493	0.661		0.008	0.023	0.914	0.879		0.44	0.307	0.808	0.897		0.003	0	0.094	0.161
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	0.008	0.007	0.006		-0.006	0	0	0.005		0.005	0.004	-0.003	0.011		-0.001	-0.004	-0.013*	-0.004
		0.756	0.206	0.152	0.109		0.433	0.983	0.955	0.598		0.649	0.744	0.792	0.296		0.926	0.516	0.067	0.533
Non-experimental design (IV, RDD)		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.
Experimental (RCT)		-0.018	-0.01	-0.01	-0.026		0.025** *	-0.01	-0.017	-0.025		0.036** *	0.040***	0.037***	-0.031*		0.035***	0.035***	0.033** *	0.025***
		0.163	0.369	0.398	0.121		0.004	0.314	0.122	0.144		0.008	0.005	0.009	0.068		0	0	0	0.001
Measured before one year after exit from the		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.		ref.	ref.	ref.	ref.

► The impact of active labour market programmes on youth

	Subsample: High-income countries										Subsample: Low- and middle-income countries									
	Weighted least squares Hedges' g regressions					Random-effects SMD regressions					Weighted least squares Hedges' g regressions					Random-effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
<i>programme</i>																				
Measured more than one year after exit from programme		0.031	0.022	0.028	0.039*		0.018**	0.015*	0.025** *	0.029** *		-0.013	-0.014	-0.015*	-0.023**		-	-	-0.012**	-
	0.139	0.321	0.171	0.087		0.026	0.057	0.002	0.003		0.107	0.11	0.093	0.04		0.008	0.009	0.014	0	
<i>Earnings Outcome</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Employment Outcome		0.010	0.012	0.01			0.016**	0.016**	0.01			-	-	-			-	-	-	
		0.289	0.126	0.113			0.019	0.02	0.187			0.001	0	0			0.001	0	0	
Business Outcome		0.115** *	0.109** *	0			0.170*	0.157*				-0.001	0.001	0.012			-0.025*	-0.022	-0.006	
		0	0	.			0.054	0.072				0.962	0.954	0.704			0.063	0.109	0.711	
<i>Unconditional outcome</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Conditional outcome		-0.015	-0.01	-0.024			-	-	-			0.01	0.012	0.011			0.006	0.009	0.007	
		0.444	0.623	0.37			0.039** *	0.030** *	0.046** *			0.207	0.133	0.12			0.272	0.113	0.188	
<i>Pooled sample</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Sample of disadvantaged youth only		-0.022**	0.006	0.015			-	-	-			0.02	0.037*	0.032			0.006	0.032** *	0.044**	
		0.036	0.707	0.414			0.040** *	-0.025*	-0.033			0.292	0.073	0.112			0.493	0	0	
<i>Both male and female participants</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Sample of male participants only		0.018	0.021	0.021			0.014	0.019*	0.021*			-0.021**	-0.020**	-			-	-	-	
		0.289	0.133	0.176			0.16	0.069	0.07			0.024	0.032	0.005			0.029** *	0.029** *	0.033**	
Sample of female participants only		-0.003	0.004	0.002			-0.008	0.005	0.008			0.012	0.012*	0.017**			0	0	0	
		0.814	0.672	0.881			0.385	0.593	0.525			0.129	0.092	0.009			0.014**	0.011**	0.016**	
<i>Participants aged 25 and older</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	
Sample of participants younger than 25		0.017** *	0.013**	0.005			0.01	0.003	-0.002			0.004	0.012	0.012			0.014*	0.016**	0.013	

► The impact of active labour market programmes on youth

	Subsample: High-income countries										Subsample: Low- and middle-income countries									
	Weighted least squares Hedges' g regressions					Random-effects SMD regressions					Weighted least squares Hedges' g regressions					Random-effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
			0.001	0.017	0.277			0.221	0.74	0.842			0.822	0.424	0.392			0.055	0.037	0.106
<i>Intervention without extra services</i>			<i>ref.</i>	<i>ref.</i>					<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
With extra services			-0.031*	-0.03				0.035**	-0.023				0.030*	0.036**				0.036**	*	0.039***
			0.051	0.163				0	0.214				0.089	0.01				0	0	
<i>Intervention without soft skills component</i>			<i>ref.</i>	<i>ref.</i>					<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
With soft skills component			-0.004	0.014				0.026**	0.048**	*			0.014	0.026				0.015**		0.024***
			0.66	0.502				0.016	0.003				0.407	0.146				0.028	0.002	
<i>Intervention does not provide incentives for participation</i>			<i>ref.</i>	<i>ref.</i>					<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
Participation incentives			-0.018	0.052**	*			-0.030**	0.056**	*			-0.027	-0.084*				-0.011		-0.048*
			0.304	0.006				0.039	0.003				0.555	0.074				0.607	0.083	
<i>Intervention does not monitor participation</i>			<i>ref.</i>	<i>ref.</i>					<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
Monitoring of participants			0.013	0.006				0.014*	-0.009				-0.013	-0.004				-0.014**		-0.013*
			0.29	0.744				0.08	0.462				0.314	0.811				0.036	0.088	
<i>Intervention does not provide certification</i>			<i>ref.</i>	<i>ref.</i>					<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>				<i>ref.</i>	<i>ref.</i>
Certification			0.014	0.017				0.016*	0.019*				0.033**	0.039***				0.054**	*	0.054***
			0.174	0.121				0.074	0.092				0.027	0.005				0	0	
<i>Intervention shorter than 4 months</i>				<i>ref.</i>					<i>ref.</i>					<i>ref.</i>						<i>ref.</i>
4-8 months				0.01					0.004					0.018						0.025***
				0.715					0.817					0.252						0.001

► The impact of active labour market programmes on youth

	Subsample: High-income countries										Subsample: Low- and middle-income countries									
	Weighted least squares Hedges' g regressions					Random-effects SMD regressions					Weighted least squares Hedges' g regressions					Random-effects SMD regressions				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
8 or more months					-0.004					-0.016					0.009					0.032***
					0.897					0.448					0.656					0.001
<i>Intervention with regional or local scale</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>
National scale					-					-					-0.01					-0.006
					0.035**					0.045**					0.503					0.384
					*					*										
					0.004					0.005										
<i>Implemented by private and public entities combined</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>					<i>ref.</i>
Public entities only					0.028**					0.033**					-0.025					-
					*					*										0.027***
					0.004					0.01					0.358					0.007
Private entities only					0.016					0.011					-0.031**					-0.012
					0.35					0.577					0.021					0.113
Constant	0.018**	0.019	0.019	0.001	0.019	0.024**	0.044**	0.052**	0.027*	0.064**	0.074**	0.087**	0.078***	0.051**	0.041*	0.078***	0.093***	0.083***	0.049**	0.022*
	*					*	*	*		*	*	*	*	*	*	*	*	*	*	*
	0.006	0.188	0.269	0.943	0.453	0	0	0	0.061	0.019	0	0	0	0.021	0.085	0	0	0	0	0.086
Estimates	1380	1334	1306	1306	1129	1380	1334	1306	1306	1129	2680	2575	2359	2359	2084	2680	2575	2359	2359	2084
Interventions	80	70	65	65	50	80	70	65	65	50	124	110	95	95	73	124	110	95	95	73
Reports	91	81	76	76	59	91	81	76	76	59	119	105	90	90	73	119	105	90	90	73
Adjusted R-squared	0.016	0.039	0.077	0.097	0.134						0.062	0.047	0.092	0.111	0.165					

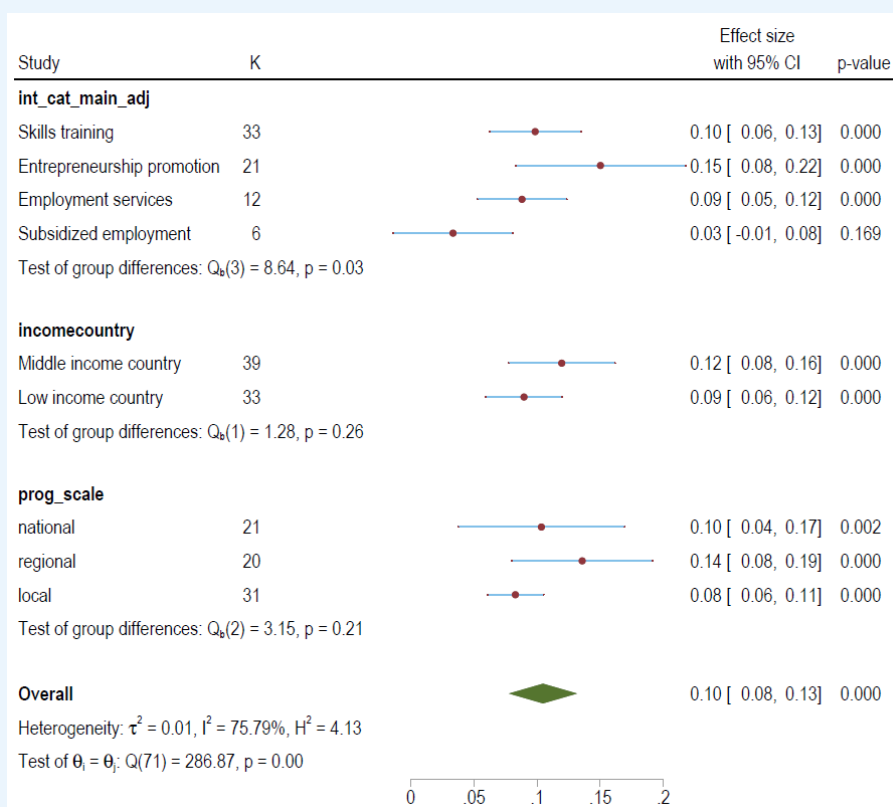
Note: RDD = regression discontinuity design; RCT = randomized controlled trial.

► **Box 4. Impact evaluations of ALMPs in Africa**

Almost one in three (33 per cent, 72 in total) of the interventions covered in this review were implemented across Africa (see Box 1), in either low- or middle-income countries. Main findings from an analysis of the sub-sample of ALMPs in Africa show similarities to findings emerging from the analysis of evaluation in low- and middle-income countries as well as some important deviations:

- **Active labour market programmes evaluated in Africa successfully improved labour market outcomes of youth.** Across all interventions and outcome measures the average estimated effect size is 0.10 SMD, with a 95 per cent CI of 0.08 SMD - 0.13 SMD, as shown in the figure below. This finding is in line with the average effects found for low- and middle-income countries globally (see Figure 9).
- **Entrepreneurship interventions are more effective relative to skills trainings, employment services and subsidized employment programmes.** This finding emerges both from the univariate analysis (see forest plot below) as well as from the multi-variate meta-regressions (presented in the regression table below, which only depicts the results from weighted least squares models). Subsidized employment interventions appear less effective in improving labour market outcomes. Interventions with entrepreneurship promotion as the main category of intervention also accounted for roughly one third (31 per cent) of the sample in Africa (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 full sample was 877 (of 889) non-missing effect sizes that remained after censoring (SMD limit = 1, Inverse standard error limits = 1 to 100), with 275 of 877 effect sizes selected (602 effect sizes excluded as the higher level available). And 594 effect sizes were aggregated to 321 groups (intervention category, country income group and programme scale by study ID level).

- **Multipronged interventions appear more effective, yet the evidence on more comprehensive programmes is considerably more nuanced than for low- and middle-income countries globally.** Studies evaluating interventions with multiple components report considerably higher effect sizes (with an additional average effect between 0.08 SMD and 0.14 SMD, see table below). However, other factors associated with higher effect sizes in low- and middle-income countries globally are not correlated with better labour market outcomes in the Africa sub-sample. For instance, there is no evidence that adding a soft skills training component to ALMPs enhances the impact,

nor that providing certification of participation or attendance is associated with increased effectiveness. Regarding the duration of programmes, a proxy of programme intensity, the evidence is inconclusive: programmes lasting between four to eight months report lower effects (compared to shorter than four months), while ALMPs lasting longer than eight months show larger impacts, on average. In contrast to the findings from low- and middle-income countries at the global level, no (positive or negative) differential effect for interventions providing incentives to programme participants is observed.

- **Young people from economically disadvantaged backgrounds who participated in ALMPs in Africa reported higher labour market outcomes.** The differential impact of up to 0.04 SMD is, however, not significant in all specifications (see table below, specifications III to V). While this finding needs to be interpreted with caution, it is an encouraging trend for a region in which 43 per cent of evaluated interventions target low-income or economically disadvantaged young people (see Box 1).
- **Interventions implemented at the regional or local level report considerably higher effect sizes compared to ALMPs rolled-out at the national level.** The magnitude of the impact is quite large and contrasts with the findings from low- and middle-income countries at global level where no such difference was observed.
- **When measuring labour market outcomes more than a year after the end of the interventions, average effect sizes diminish 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 materialise and are maintained in the medium- and long-run.

► **Meta-regression results, weighted least squares Hedges' g regressions, Africa sample**

	I	II	III	IV	V
SMD sampling variance	-0.885*	-1.165**	-1.608***	-1.313**	0.28
	0.068	0.016	0.004	0.019	0.71
<i>Mainly skills training</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Mainly entrepreneurship promotion	0.030**	0.014	0.057***	0.087***	0.184***
	0.015	0.281	0.005	0	0
Mainly employment services	0.009	0.004	0.02	0.054**	0.014
	0.551	0.781	0.329	0.022	0.555
Mainly subsidized employment	-0.045***	-0.056**	-0.046	-0.065*	-0.1
	0.005	0.038	0.105	0.057	0.106
<i>Publication not peer reviewed</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Peer-reviewed publication		0.041***	0.042**	0.027	0.089***
		0.002	0.015	0.134	0
<i>Non-intention-to-treat estimate</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Intention-to-treat estimate		-0.009	-0.034*	-0.015	0.006
		0.561	0.053	0.432	0.74
<i>Non-experimental design (IV, RDD)</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Experimental (RCT)		-0.061*	-0.055	-0.117***	-0.169**
		0.067	0.141	0.004	0.02
<i>Measured before one year after exit from the programme</i>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Measured more than one year after exit from programme		-0.032**	-0.039**	-0.028*	-0.102***
		0.01	0.011	0.076	0
<i>Earnings outcome</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Employment outcome			0.033**	0.026*	0.015
			0.018	0.055	0.262
Business outcome			0.006	0.009	0.003
			0.74	0.637	0.872
<i>Unconditional outcome</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Conditional outcome			0.014	0.024	0.058**
			0.471	0.232	0.013
<i>Pooled sample</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Sample of disadvantaged youth only			0.028**	0.005	0.036*
			0.048	0.741	0.064
<i>Both male and female participants</i>			<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Sample of male participants only			0.014	-0.014	0.001
			0.505	0.499	0.974

Sample of female participants only	-0.001	-0.025*	-0.006		
	0.946	0.093	0.747		
<i>Intervention without extra services</i>		<i>ref.</i>	<i>ref.</i>		
With extra services		0.078***	0.142***		
		0	0		
<i>Intervention without soft skills component</i>		<i>ref.</i>	<i>ref.</i>		
With soft skills component		0.014	-0.049**		
		0.521	0.04		
<i>Intervention does not provide incentives for participation</i>		<i>ref.</i>	<i>ref.</i>		
Participation incentives		0.03	-0.032		
		0.339	0.5		
<i>Intervention does not monitor participation</i>		<i>ref.</i>	<i>ref.</i>		
Monitoring of participants		-0.01	-0.067***		
		0.463	0		
<i>Intervention does not provide certification</i>		<i>ref.</i>	<i>ref.</i>		
Certification		0.015	-0.011		
		0.445	0.568		
<i>Intervention shorter than 4 months</i>			<i>ref.</i>		
4–8 months			-0.062**		
			0.033		
8 or more months			0.077***		
			0.001		
<i>Intervention with regional or local scale</i>			<i>ref.</i>		
National scale			-0.122***		
			0		
<i>Implemented by private and public entities combined</i>			<i>ref.</i>		
Public entity only			0.139***		
			0		
Private entity only			0.121***		
			0		
Constant	0.083***	0.156***	0.129***	0.162***	0.082
	0	0	0.004	0.001	0.279
Estimates	877	792	641	641	424
Reports	70	62	52	52	40
Interventions	55	47	38	38	29
	I	II	III	IV	V

6.4 Costs

Costs considerations are an essential complement to the evidence on the effectiveness of interventions. Kluge 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.

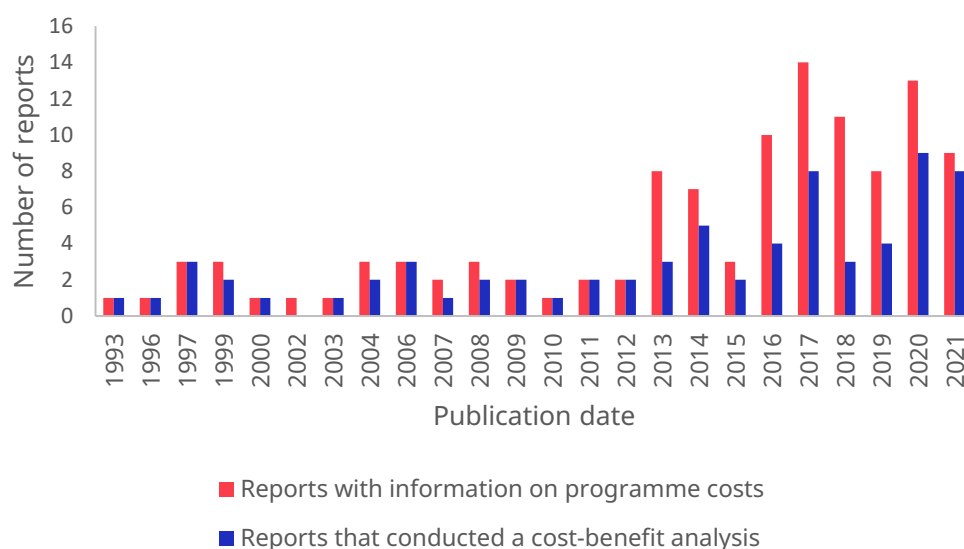
This systematic review collected data on the availability of information on intervention costs and the presence of cost-benefit analyses (which, for simplicity, also includes cost-effectiveness analyses) in reports. Following the definitions in Fiala and Hempel (2011), a cost-benefit analysis identified 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 costs 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 Kluge et al. (2017), the systematic review expanded the coding exercise to integrate information on costs per head and the results of cost-benefit (and cost-effectiveness) analyses when available. The latter distinguished between a net benefit, a net loss or neutral results (costs equal to benefits). In addition, the systematic review included a question on the time (in months) needed for an intervention to yield net benefits.

Cost variables proved to be sensitive in the coding, and their interpretation demands further study to make information standardized and comparable. This section focuses on frequencies of reported information and results reported through the cost-benefit analyses.

Figure 12 shows an increasing number of reports citing costs and presenting cost-benefit analyses. Up to 2012, there were typically three or fewer reports with costing information published each year. But that number increased to approximately ten reports per year from 2013 onwards up to 2021. Although much remains to be done to improve the reporting of costing and generate better evidence on the cost-effectiveness and feasibility of youth employment interventions, the data suggest an encouraging trend.

► **Figure 12. Increasing number of reports citing cost information, 1993–2021**



As reported in table 12, across the 228 reports included in this review, 112 reports (49 per cent) cited cost information, while 70 (31 per cent) included cost-benefit analysis for at least one intervention. When looking at the sample’s income-level composition, most reports cited costs stemming from programmes implemented in high-income (47 per cent) and middle-income (39 per cent) countries. However, when considering the distribution of reports across income levels, the picture becomes more nuanced: Reports studying interventions in low-income countries, while fewer in number, were more likely to cite costing information, at 52 per cent, compared with 51 per cent for high-income countries and 46 per cent for middle-income countries. When looking at the inclusion of cost-benefit analysis, the breakdown of reports across income levels mirrors that of cost information, although the likelihood of reporting cost-benefit analysis is now more comparable across income levels.

Consistent with the distribution of studies in the sample by main intervention type, most reports that included cost or cost-benefit information also evaluated skills development interventions (70.5 per cent for cost and 75.7 per cent for cost-benefit, respectively). Looking at the likelihood of reporting costing information across main intervention category, 54.9 per cent of reports evaluated skills development interventions, while the likelihood fell to 42.1 per cent for those evaluating entrepreneurship promotion and to 43.8 per cent and 33.3 per cent on subsidized employment and employment services, respectively. An analogous pattern held for the inclusion of cost-benefit analysis.

What do cost-benefit analyses say? Seventy-three per cent of cases that reported a cost-benefit analysis yielded a net benefit, compared with 20 per cent for net loss and 7 per cent for a neutral effect. This was computed after missing values and based on the minimum unit of analysis, or SMDs. Among the reported cases with net benefit, 95 per cent related to skills training.

► **Table 12. Information on programme cost and cost-benefit analyses across country income level and main category of interventions**

Panel A	Country income level				Total
	High income	Middle income	Low income		
Total reports	104	95	29		228
Costs					
No. of reports with information on costs	53	44	15		112
% of reports with cost information by country income category	47.3	39.3	13.4		100.0
% of reports with cost information as a share of total number of reports	23.2	19.3	6.6		49.1
% of reports with cost information within country income category	51.0	46.3	51.7		49.1
Cost-benefit analyses					
No. of reports with information on costs and benefits	34	27	9		70
% of reports with costs and benefits information by country income category	48.6	38.6	12.9		100.0
% of reports with costs and benefits information as a share of total number of reports	14.9	11.8	3.9		30.7
% of reports with costs and benefits information within country income category	32.7	28.4	31.0		30.7
Panel B					
	Main category of intervention				Total
	Skills training	Entrepreneurship promotion	Employment services	Subsidized employment*	
Total reports	144	19	33	32	228
Costs					
No. of reports with information on costs	79	8	11	14	112
% of reports with cost information by country income category	70.5	7.1	9.8	12.5	100.0
% of reports with cost information as a share of total number of reports	34.6	3.5	4.8	6.1	49.1
% of reports with cost information within country income category	54.9	42.1	33.3	43.8	49.1
Cost-benefit analyses					
No. of reports with information on costs and benefits	53	4	9	4	70
% of reports with costs and benefits information by main intervention category	75.7	5.7	12.9	5.7	100.0
% of reports with costs and benefits information as a share of total number of reports	23.2	1.8	3.9	1.8	30.7
% of reports with costs and benefits information within main intervention category	36.8	21.1	27.3	12.5	30.7

Note: A report may contain more than one study and more than one intervention and therefore cite more than one cost-benefit analysis per main category of intervention. To align with the total number of reports, the authors selected a category of intervention among four studies that reported cost-benefit analysis for various intervention types. *= Including public works.

► 7 Conclusions

Worldwide, the COVID-19 pandemic has exacerbated the already arduous transition of young people to the labour market, pushing them disproportionately into unemployment, inactivity or lower-quality jobs. It also has led to considerable learning losses, including through disruptions in education and skills development programmes and fewer opportunities to engage in work-based learning, which ultimately affects young people's employability. Together, these developments are likely to hamper labour market outcomes of young people in the long-run, leading to scarring effects. The time to act and reverse the jobs and skills losses is now. This calls for substantially upscaling interventions and investments in youth who are experiencing a much slower labour market recovery and face potentially lasting hits to their career trajectory.

To guide the investments of policymakers and practitioners in youth employment, this report presents the results from the most comprehensive systematic review of youth employment interventions to date. The review examined the effectiveness of youth-targeted ALMPs to improve labour market outcomes of beneficiaries by assessing the role that context, youth characteristics, evaluation features, programme design and implementation had in moderating the impact of ALMPs on employment, earnings and business performance outcomes of youth.

The systematic review synthesized the evidence of more than three decades of impact evaluations. It followed up and leveraged the information and methodologies from Kluge et al. (2017 and 2019) and upgraded the information with new reports or new versions of already-included reports. It also refined earlier inclusion criteria and the search process, including through machine learning, and captured additional pieces of information deemed relevant to the current times.

Empirical evidence from youth employment programmes has increased substantially over the past decade. This systematic review doubled the information gathered and computed by Kluge et al. (2017 and 2019) in all 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 reports depicted the effects of ALMPs in low- and middle-income economies.

As a result, today's evidence is much more balanced worldwide. Across continents, Africa has the largest share (33 per cent) of included studies, followed by Europe (27 per cent) and Latin America and the Caribbean (18 per cent). This enhances the ability of this review to capture intervention models and results from a wide range of contexts. At the same time, the share of experimental evaluations has continued growing, amounting to 47 per cent of the 248 included studies. About three quarters of the studies were either peer reviewed (38 per cent) or published as working papers (35 per cent).

The report offers several substantial findings on the effectiveness of youth labour market programmes:

First, ALMPs have been effective in improving the labour market outcomes of young people. A consistent positive effect of ALMPs is documented across all income groups, outcome categories and when analysing impact evaluations from Africa separately. The average overall effect size is 0.08 SMD, which varies depending on the outcome of interest and method applied. It suggests some progress vis-à-vis the results from Kluge et al. (2017 and 2019), which built on reports published before 2015 and found an average impact of 0.04 SMD. Recent meta-analyses in social sciences define "small" effects as those below a 0.05 SMD and "medium" as those in the 0.05–0.20 SMD range, supporting the case for increased effectiveness of ALMPs.

Second, youth-focused ALMPs are more effective in low- and middle-income countries than in high-income countries on average. The systematic review finds an effect size of 0.09 SMD for low-income countries, 0.10 SMD for middle-income countries and 0.06 SMD for high-income countries. Looking at the sub-sample of ALMPs from Africa, the study documents an overall effect size of 0.10 SMD. To explain the larger impacts of ALMPs in developing countries, Betcherman et al. (2007) and Kluge et al. (2017 and 2019) argued that since youth in low- and middle-income countries face comparatively stronger labour market challenges, a marginal investment to improve their situation pays off more there than in high-income countries.

In low- and middle-income countries, ALMPs appear to be particularly successful at improving the earnings of youth when compared to employment outcomes. Such increase in the earnings capacity of youth may occur via entrepreneurship promotion interventions. In the case of high-income countries, employment outcomes appear

more responsive to youth employment interventions than earnings, which may be explained by the nuanced focus of the ALMPs on the unemployed youth and the presence of unemployment insurance benefits, which provide a financial cushion to targeted youth not in employment.

Third, the “what”, that is the type of intervention, plays an important role in determining the success of ALMPs. Overall, entrepreneurship promotion and skills training interventions report larger impacts than employment services and subsidized employment programmes. This result contrasts with earlier findings by Kluge et al., who found no evidence of certain intervention types systematically outperformed other. It suggests that the recent evidence captures more clearly the differential effect across main categories of intervention. Some important nuances emerge across country income groups:

- In the low- and middle-income countries, studies of entrepreneurship interventions report the largest impacts, followed by employment services and skills training interventions. This finding may be connected to a lack of private sector jobs, gaps in education systems, and the prevalence of labour market information asymmetries in low- and middle-income countries. In such contexts, young workers are largely employed in the informal economy, young jobseekers have limited knowledge of job opportunities and there are insufficient information channels on job vacancies and career paths.
- In high-income countries, studies of skills training interventions report the largest impacts, while wage subsidies and public works have only marginal impacts on average. The evidence is most pronounced for impacts on earnings outcomes. This finding is in line with an earlier meta-analysis of Card et al. (2018) based on ALMPs targeting adult workers, showing that human capital centred ALMPs achieve larger impacts, particularly in the long-run.²³

Fourth, “how” an intervention is designed and implemented is an important determinant of labour market outcomes:

- In low-and middle-income countries, studies of comprehensive, multipronged ALMPs and those that certify participation report larger impacts on youth labour market outcomes. ALMPs that offer various services are able to better address the many constraints young people face. This effect is also evident and strongly significant among programmes that integrate soft skills training and those that provide certification to participants. An important factor in low- and middle-income countries seems to be that ALMPs enable young jobseekers to better signal their competencies and employability to employers.
- Interventions in low- and middle-income countries are more successful when the programme duration exceeds four months. One reason may be that youth in developing contexts face a larger number of different (though often interlinked) labour market barriers and require more intensive interventions.
- In high-income countries, larger impacts are reported for interventions that are implemented only by public stakeholders as compared to those under public-private engagement. By contrast, in low- and middle-income countries, public-private partnerships are linked to larger impacts. While less consistent across specifications, this result may speak to the value of partnerships when reaching out specific populations or addressing multiple labour market constraints in developing country contexts.

Fifth, the review looked at “who” benefits from youth-focused ALMPs. It found that impacts are larger for youth from economically disadvantaged backgrounds, young women and participants younger than 25 years, particularly in low- and middle-income countries. This result is crucial because these groups have been severely affected by the job crisis induced by the COVID-19 pandemic, are more prone to be in informal employment or in inactivity and have been comparatively out of reach of the emergency employment policies, which were often designed to support formal workers.

Sixth, the “where” an intervention takes place seems to explain some of the effects of ALMPs in high-income countries as well as across the Africa region. In particular, interventions with a local or regional scope lead to better outcomes than national level interventions. This offers promise for municipal or regional efforts to curbe the youth employment challenge, likely through better targeting and outreach strategies.

²³ There is yet little evidence from youth entrepreneurship programmes in high-income countries, which suggests room for further testing of their impact in such context.

Seventh, and consistent with previous reviews, experimental evaluations (or RCTs) link to lower impacts as compared to quasi- and non-experimental designs. Publications that were not peer reviewed do not systematically outperform those that were. Similarly, the availability of intention-to-treat estimates do not systematically lead to differential effects. Other evaluation features offered more nuanced results across country income subgroups, for instance: in high-income countries, effect sizes measured at least one year after the end of the intervention were linked to higher effect size magnitude, suggesting that in those countries, impacts tend to increase over time after the end of the programme. In the case of low- and middle-income countries, effects do not increase 12 months after the end of the intervention, however they may materialize within that period.

Last, the evidence also shows an encouraging positive trend in the inclusion of cost information and cost-benefit analyses in impact evaluations since the mid-2010s. Nearly three fourths of available cost-benefit analyses reported that programme benefits surpassed the programme costs. These were overrepresented among skills training interventions (at 95 per cent) and call for greater attention to assessing cost-effectiveness in other intervention types. The review of costs highlights the gap in harmonized reporting and stresses the importance of 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.

Similar to Kluge et al. (2017 and 2019), this updated systematic review found limited evidence on the relative effectiveness across intervention components and on the impact of combining different intervention types. The use of network meta-analysis may be a good avenue for future research as it is well suited to unpack relative effectiveness across intervention types by leveraging all available information comparing impacts across interventions and vis-à-vis a control group (Wilson, Tanner-Smith and Mavridis, 2016). Moreover, further research is needed to better understand some contextual and design factors of ALMPs, including the transmission channels that lead to positive labour market outcomes over time and the effects of providing young people incentives to participate in the programmes.

▶ 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.
- 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.
- 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.
- 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 Kluge, 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.
- 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.
- 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.
- Eichhorst, Werner, Paul Marx, Ulf Rinne, and Johannes Brunner. 2022. *Promoting Youth Employment During COVID-19: A Review of Policy Responses*. Geneva: ILO.
- Escudero, Verónica, Jochen Kluge, Elva López Mourelo, Clemente Pignatti. 2019. "Active Labour Market Programmes in Latin America and the Caribbean: Evidence from a Meta-Analysis", *Journal of Development Studies* 55 (12): 2644–2661.
- 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.
- Gregg, Paul, and Emma Tominey. 2005. "The Wage Scar From Male Youth Unemployment", *Labour Economics* 12 (4): 487–509.

Gentilini, Ugo, Mohamed Almenfi, Ian Orton, John Blomquist, Pamela Dale, Luciana De la Flor Giuffra, Vyjayanti Desai, Maria Belen Fontenez, Guillermo Galicia, Veronica Lopez, Georgina Marin, Ingrid Veronica Mujica, Harish Natarajan, David Newhouse, Robert Palacios, Ana Patricia Quiroz, Claudia Rodriguez Alas, Gayatri Sabharwal, and Michael Weber. 2022. *Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures*. Washington, DC: World Bank.

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.

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.

—. 2021a. *World Employment and Social Outlook – Trends 2021*.

—. 2021b. *Skills Development in the Time of COVID-19: Taking Stock of the Initial Responses in Technical and Vocational Education and Training*.

—. 2021c. *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*.

Kahn, Lisa B. 2010. "The Long-Term Labor Market Consequences of Graduating From College in a Bad Economy", *Labour Economics*, 17 (2): 303–316.

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.

Kemper, Johanna, Jonathan Stöterau, and Andrea Ghisletta. 2022 (forthcoming, mimeo). "The Impact of Vocational Training Interventions on Youth Labor Market Outcomes: A Meta-analysis".

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

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

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.
- 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.
- 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. 2022a. *OECD Employment Outlook 2022: Building Back More Inclusive Labour Markets*. Paris: OECD Publishing.
- . 2022b. "[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.
- Schady, Nobert, Alaka Holla, Shwetlena Sabarwal, Joana Silva, and Andres Yi Chang. 2022. *Collapse and Recovery: How the COVID-19 Pandemic Eroded Human Capital and What to Do About It*. Washington, DC: World Bank.
- Stanley, Tom and Hristos Doucouliagos. 2012. *Meta-regression Analysis in Economic and Business*. London: Routledge.
- The Campbell Collaboration. 2021. *Campbell Collaboration Systematic Reviews: Policies and Guidelines Version 1.8*.
- Tripney, 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.
- 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, Volume 12, Issue 1*.

▶ Annexes

A.1. Study and programme information

Variable name	Description	Possible answers
Date of publication	Year of publication of study	Type: Numeric
Publication Status	Publication status of study	Type: Numeric or string. Please select accordingly from the drop-down list: 1: Peer-reviewed Journal 2: Working paper 3: Mimeo 4: Book 5: Policy or position paper 6: Evaluation or technical report 7: Dissertations or theses
Country	Country where programme is implemented	Type: String
Region	Region of country where programme is implemented	Type: String
Programme name	Programme name	Type: String

A.2. Primary search sources

34 primary search sources used in this systematic review

- Directory of Open Access Journals (DOAJ) (type: database; period covered: 1990–2021)
- PASCAL (type: database; period covered: 1990–2021)
- ProQuest Research Library (type: database; period covered: 1990–2021)
- ScienceDirect (type: database; period covered: 1990–2021)
- Scopus (type: database; period covered: 1990–2021)
- Library and Archives Canada (type: database; period covered: 1990–2021)
- Academic Search Complete (type: database; period covered: 1990–2021)
- National Technical Information Service (NTIS) (type: grey literature; period covered: 1990–2021)
- Open Grey (type: grey literature; period covered: 1990–2021)
- UNESCODOC (type: institutional website; period covered: 1990–2021)
- World Bank e-library (type: institutional website; period covered: 1990–2021)
- World Bank OpenKnowledge Repository (type: institutional website; period covered: 1990–2021)
- CORDIS Library (type: institutional website; period covered: 1990–2021)
- ASSIA (Applied Social Sciences Index and Abstracts) (type: database; period covered: 2013–2021)
- RePEc (Research Papers in Economics) / IDEAS Economics and Finance Research: <http://ideasrepec.org/> (type: database; period covered: 2013–2021)
- Web of Science (Social Science Citation Index (SSCI) of) (type: database; period covered: 2013–2021)
- ABI/INFORM Global (type: database; period covered: 2013–2021)
- EconLit (type: database; period covered: 2013–2021)
- ERIC (Education Resources Information Centre) (type: database; period covered: 2013–2021)
- IBSS (International Bibliography of the Social Sciences) (type: database; period covered: 2013–2021)
- Google Scholar (type: database; period covered: 2013–2021)
- PAIS International (type: database; period covered: 2013–2021)
- ProQuest Dissertations & Theses Database (type: grey literature; period covered: 2013–2021)
- 3ie Register of Impact Evaluation Published Studies (RIEPS) (type: institutional website; period covered: 2013–2021)
- Abdul Latif Jameel Poverty Action Lab (JPAL) Evaluation and Publication Database (type: institutional website; period covered: 2013–2021)
- Innovations for Poverty Action (IPA) Database (type: institutional website; period covered: 2013–2021)
- Campbell Collaboration <http://www.campbellcollaboration.org> (type: institutional website; period covered: 2013–2021)
- Poverty and Economic Policy Research Network (PEP): Project List (type: institutional website; period covered: 2013–2021)
- United States Department of Labour, Employment and Training Administration, Research Publication Database (type: institutional website; period covered: 2013–2021)
- Expanded Academic ASAP (GALE) (type: database; period covered: 1990–2021)
- HighWire Press (type: database; period covered: 1990–2021)
- PolicyFile (type: database; period covered: 1990–2021)
- Canadian Research Index (type: database; period covered: 1990–2021)
- Readex – AccessUN (type: institutional website; period covered: 1990–2021)

A.3. Effect size information

Variable	Description	Possible answers
Intervention ID	Intervention ID within programme evaluated by the study accordingly to effect size that is going to be coded.	Type: Alphanumeric
Cohort ID	Cohort ID for intervention for which estimate effects is going to be coded	Type: Alphanumeric
Subgroup	Indication that estimation sample is different from targeted intervention population (authors statement or obvious deviation)	Type: Alphanumeric 0: No 1: Yes
Subsample	Treatment effect is estimated for a subsample of the entire study population (only if estimates for total (broader) sample are also reported) (such as subsample analysis of females)	Type: Alphanumeric 0: No 1: Yes
Outcome category	Outcome category	Type: Alphanumeric 1: Employment outcome 2: Earnings outcome 3: Business performance outcome
Outcome	Outcome for which effect size is measured	Type: Alphanumeric. For employment outcome: 1: Employment probability 2: Unemployment probability 3: Participation rate 4: Worked hours 5: Unemployment duration 6: Quality of employment (have contract, fixed benefits, for example) For earning outcome: 7: Earnings or income 8: Household income 9: Consumption 10: Salary or wage For Business performance outcome 11: Profits 12: Sales 13: Number of employees or jobs created 14: Capital and investment 15: Business creation 16: Business survival
Occupation category	Occupation category for which outcome is measured	Type: Alphanumeric 1: Dependent employment 2: Self-employment 3: Both
Status of occupation	Status of occupation for which outcome is measured	Type: Alphanumeric 1: Formal 2: Informal 3: Both
Conditional outcome	Effect size measures effect conditional on some other primary outcome. This variable indicates whether there is a conditionality for the effects that are measured. Those conditions should only be based on other outcomes or constructs of interest. For example, income could be conditional on employment, where conditionality (employment) is an outcome.	Type: Alphanumeric 0: No 1: Yes
Participation conditions	Description of condition for participants to enter the sample	Type: String. Describe the outcome that must be satisfied for participants to enter the sample population (such as employed)

Variable	Description	Possible answers
Lower-bound age	Group for which effect is estimated: lower-bound age	Type: Numeric
Upper-bound age	Group for which effect is estimated: upper-bound age	Type: Numeric
Gender	Group for which effect is estimated: gender	Type: Alphanumeric 1: Male 2: Female 3: Both
Disadvantaged groups	Group for which effect is estimated: low-income, disadvantaged, at-risk, vulnerable youth, marginalized groups, etc.	Type: Alphanumeric 0: No 1: Yes
Primary education	Group for which effect is estimated: educational attainment, either none or primary school level	Type: Alphanumeric 0: No 1: Yes
Secondary education	Group for which effect is estimated: educational attainment of secondary school level	Type: Alphanumeric 0: No 1: Yes
High-level education	Group for which effect is estimated: educational attainment: tertiary	Type: Alphanumeric 0: No 1: Yes
Geographical area	Group for which effect is estimated: area	Type: Alphanumeric 1: Urban 2: Rural 3: Both
Measurement method	Method of measurement of effect size	Type: Alphanumeric 1: Pre-test and post-test comparison 2: Post-test comparison 3: Follow-up comparison
Type of effect size	Type of effect size measure. This should reflect how the underlying outcome variable is measured, regardless of how reported in study. For example, "employed probability" is binary even if percentage employed is reported.	Type: Alphanumeric 1: Dichotomous and binary 2: Continuous 3: Correlational
Type of dichotomous effect size	This variable reflects how the effect estimates are presented in the report and not what the underlying variable is measuring.	Type: Alphanumeric. 1: Number of events (treatment and control) 2: Event rates (treatment and control) 3: 2x2 contingency table (both events, treatment, control, both non-events) 4: Odds ratio 5: Log odds ratio 6: Risk ratio 7: Risk difference
Type of continuous effect size	Like above, this variable reflects how the effect estimates are presented in the report and not what the underlying variable is measuring.	Type: Alphanumeric 1: Means (treatment and control) 2: Raw difference in means 3: (Covariate) Adjusted difference in means (=unstandardized regression coefficient) 4: Standardized mean difference (= standardized regression coefficient) 5: Log difference in means 6: Log standardized mean difference 7: t-value, f-value, p-value (from a paired t-test) 8: Frequency table (2 groups sample sizes)
Raw effect size	Estimated treatment effect, as reported in the study	Type: Numeric

Variable	Description	Possible answers
Effect size direction	Effect size direction	Type: Alphanumeric 1: Positive (higher values equal more positive outcomes) 2: Negative (higher values equal more negative outcomes)
Confidence interval – lower bound	Lower value of confidence intervals if reported	Type: Numeric
Confidence interval – upper bound	Upper value of confidence intervals if reported	Type: Numeric
Duration of intervention	Duration between individual exits the intervention and data measurement	Type: Numeric Duration is expressed in weeks between the end of the intervention and the follow-up date.
Date	Date at which effect size is measured	Type: Alphanumeric This variable reports the date at which effect size is measured (follow-up date)
Page number	Page number where this effect size was found	Type: Numeric
Significance test	Type of significance test	Type: Alphanumeric 1: t-value 2: p-value 3: F-value (df=1) 4: Chi-Square (df=1) 5: Standard error of coefficient estimate (from regression or matching) 6: Variance 7: Confidence intervals (lower, upper) 8: Sample sizes (treatment and control) 9: Standard deviations (treatment and control) 10: common standard deviation 11: Total sample size 12: Standard deviation of difference
Significance value	Value of significance test	Type: Numeric
Statistical significance at 5% level	Estimated treatment effect is significant on the 5% level	Type: Alphanumeric 0: No 1: Yes
Sample mean	Mean in the total sample	Type: Numeric
Standard deviation	Standard deviation in total sample (post-intervention)	Type: Numeric
Sample size	Number of observations total	Type: Numeric
Outcome of treatment group	Outcome of treatment group at end-line, if reported (mean, for example), uses values after matching and covariate adjustment	Type: Numeric
Standard deviation of treated	Standard deviation in treatment group (post-intervention)	Type: Numeric
Size of treatment group	Number of observations in treatment group for which effect is estimated	Type: Numeric
Outcome of control group	Outcome of control group at endline if reported (mean, for example), uses values after matching and covariate adjustment)	Type: Numeric
Standard deviation of control	Standard deviation in control group (post-intervention)	Type: Numeric
Size of control group	Number of observations in control group for which effect is estimated	Type: Numeric
Pooled standard deviation	Pooled standard deviation (post-intervention)	Type: Numeric

A.4. Description of data used and empirical methods

Variable name	Description	Possible answers
Original data	Data source: collected original data for study	Type: Alphanumeric 0: No 1: Yes
Survey data	Data source: survey data	Type: Alphanumeric 0: No 1: Yes
Administrative	Data source: administrative data	Type: Alphanumeric 0: No 1: Yes
Evaluation design	Impact evaluation research design	Type: Alphanumeric 1: RCT 2: Natural experiment 3: Pipeline 4: Only panel 5: Only cross-section
Statistical method	Statistical methodology	Type: Alphanumeric 1: Instrumental variable, 2: Matching (PSM, etc), 3: Regression-adjusted DID 4: Simple DID 5: DID and matching combined 6: Regression discontinuity design (RDD) 7: Multivariate linear (OLS, ANOVA, others less likely) 8: Multivariate non-linear regression model (Probit, Logit, Tobit) 9: Tabulation (simple differences in mean) 10: Other (specify in comments)
Structure of data	Structure of data underlying the treatment effect estimate	Type: Alphanumeric 1: Cross-sectional data 2: Pooled (repeated) cross-sectional data 3: Panel data
Unit of analysis error	Treatment effect estimate is likely subject to unit of analysis error	Type: Alphanumeric 0: No 1: Yes
Intention-to-treatment	Intention-to-treat estimation explicitly mentioned	Type: Alphanumeric 0: No (only if specified that estimator does not measure intention-to-treat) 1: Yes

A.5. Intervention and programme characteristics

ID	Variable name and description	Possible answers
Skills training	Intervention category: skills training	Type: Alphanumeric 0: No 1: Yes
Entrepreneurship Promotion	Intervention category: entrepreneurship promotion	Type: Alphanumeric 0: No 1: Yes
Employment services	Intervention category: employment services	Type: Alphanumeric 0: No 1: Yes
Subsidized employment	Intervention category: subsidized employment (wage subsidies and employment guarantee schemes)	Type: Alphanumeric 0: No 1: Yes
Public works	Intervention category: public works and public services	Type: Alphanumeric 0: No 1: Yes

ID	Variable name and description	Possible answers
Main category of intervention	Intervention category: main category of intervention	Type: Alphanumeric 1: Skills training 2: Entrepreneurship promotion 3: Employment services 4: Subsidized employment 5 Public works
Subcomponent	Evaluation estimates effect for a subcomponent of a more comprehensive programme	Type: Alphanumeric 0: No 1: Yes

A.6. Programme characteristics: Skills training

Variable name	Description	Possible answers
Technical and/or vocational skills	Type of skills training: technical and/or vocational skills	Type: Binary 0: No 1: Yes
Digital skills	Type of skills training: digital skills	Type: Binary 0: No 1: Yes
Business skills	Type of skills training: business skills	Type: Binary 0: No 1: Yes
Literacy and/or numeracy skills	Type of skills training: literacy and/or numeracy	Type: Binary 0: No 1: Yes
Soft skills	Type of skills training: behavioural, life skills, soft skills	Type: Binary 0: No 1: Yes
Distant learning	Skill training delivered: distant learning (books, online training)	Type: Binary 0: No 1: Yes
In classroom	Skill training delivered: in classroom	Type: Binary 0: No 1: Yes
On-the-job training	Skill training delivered: at the workplace (internships, on-the-job training schemes, non-apprenticeship schemes)	Type: Binary 0: No 1: Yes
Integrated schemes	Skill training delivered: integrated schemes (apprenticeships, in shops with master craftsmen and craftswomen)	Type: Binary 0: No 1: Yes
Duration of skills-training programme	Duration of skill training programme: total number of hours per individual (averages)	Type: Numeric
Duration of in-classroom training	Duration of in-classroom training in hours	Type: Numeric
On-the-job training duration	Duration of work-based learning (on-the-job training, internships, apprenticeship and other types of traineeships) in hours	Type: Numeric
Contract	Is there a contract between firms and individuals for workplace training?	Type: Binary 0: No 1: Yes
Public training institution	Provider of the skill training: public training institution	Type: Binary 0: No 1: Yes
Private training institution	Provider of the skill training: private training institution	Type: Binary 0: No 1: Yes

Skill training provider: NGO	Provider of the skill training: foundation or community-based, non-government or civil society organization	Type: Binary 0: No 1: Yes
Lump-sum budget	Payment system to training providers: lump-sum budget	Type: Binary 0: No 1: Yes
Payment for services delivered	Payment system to training providers: payment for services delivered	Type: Binary 0: No 1: Yes
Payment by outcomes	Payment system to training providers: payment by outcomes	Type: Binary 0: No 1: Yes
Selection of skills – national government	Selection of skills: identified by national government	Type: Binary 0: No 1: Yes
Selection of skills – regional or local government	Selection of skills: identified by regional or local government	Type: Binary 0: No 1: Yes
Selection of skills – civil society	Selection of skills: identified by civil society	Type: Binary 0: No 1: Yes
Selection of skills – private sector	Selection of skills: Identified by private sector	Type: Binary 0: No 1: Yes
Selection of skills – donors or development agencies	Selection of skills: identified by donors or development agencies	Type: Binary 0: No 1: Yes

A.7. Programme characteristics: Entrepreneurship promotion

Variable name	Description	Possible answers
Business advisory and mentoring	Type of intervention: business advisory and mentoring	Type: Binary 0: No 1: Yes
Business skills	Type of intervention: business skills	Type: Binary 0: No 1: Yes
Access to markets and value chains	Type of intervention: access to markets and value chains	Type: Binary 0: No 1: Yes
Credit or access to credit	Type of intervention: credit or access to credit	Type: Binary 0: No 1: Yes
Grants	Type of intervention: grants (monetary or in-kind)	Type: Binary 0: No 1: Yes
Microfranchising	Type of intervention: microfranchising	Type: Binary 0: No 1: Yes
Provider: public institution	Provider of the entrepreneurship services: public institution	Type: Binary 0: No 1: Yes
Provider: private institution	Provider of entrepreneurship services: private institution	Type: Binary 0: No 1: Yes
Provider: NGO	Provider of the entrepreneurship services: foundation or community-based, non-government or civil society organization	Type: Binary 0: No 1: Yes

Lump-sum budget	Payment system to service providers: lump-sum budget	Type: Binary 0: No 1: Yes
Payment for services delivered	Payment system to service providers: payment for services delivered	Type: Binary 0: No 1: Yes
Payment by outcomes	Payment system to service providers: payment by outcomes	Type: Binary 0: No 1: Yes
Selection process: business plan or idea competition	Selection process: business plan or idea competition	Type: Binary 0: No 1: Yes
Selection process: Survey	Selection process: survey, interview or test	Type: Binary 0: No 1: Yes
Selection process: none	Selection process: any youth within target population is eligible	Type: Binary 0: No 1: Yes

A.8. Programme characteristics: Employment services

Variable name	Description	Possible answers
Job counselling	Type of employment service: job counselling, job search assistance and mentoring	Type: Binary 0: No 1: Yes
Job placement	Type of employment service: job placement	Type: Binary 0: No 1: Yes
Financial assistance	Type of employment service: financial assistance for job search	Type: Binary 0: No 1: Yes
Provider: public institution	Provider of the employment services: public institution	Type: Binary 0: No 1: Yes
Provider: private institution	Provider of employment services: private institution	Type: Binary 0: No 1: Yes
Provider: NGO	Provider of the employment services: foundation or community-based, non-government or civil society organization	Type: Binary 0: No 1: Yes
Lump-sum budget	Payment system to service providers: lump-sum budget	Type: Binary 0: No 1: Yes
Payment for services delivered	Payment system to service providers: payment for services delivered	Type: Binary 0: No 1: Yes
Payment by outcomes	Payment system to service providers: payment by outcomes	Type: Binary 0: No 1: Yes

A.9. Programme characteristics: Wage subsidies

Variable name	Description	Possible answers
Social security contributions	Type of subsidy: reduction in employer social security contributions	Type: Binary 0: No 1: Yes
Reduction in employer labour or wage costs	Type of subsidy: reduction in employer labour or wage costs	Type: Binary 0: No 1: Yes
Direct payment	Type of subsidy: direct payment to the individual (such as voucher)	Type: Binary 0: No 1: Yes
Subsidy setting: government	Subsidy setting: the absolute level of the subsidy is fixed by the government	Type: Binary 0: No 1: Yes
Subsidy setting: variable	Subsidy setting: the absolute level of the subsidy is variable	Type: Binary 0: No 1: Yes
Duration of the subsidy	Maximum duration of the subsidy in months per individual	Type: Numeric
Eligibility – open	Eligible employers: any employer is eligible	Type: Binary 0: No 1: Yes
Eligibility – formal contracts	Eligible employers: only employers who offer formal contracts and/or only formal employers	Type: Binary 0: No 1: Yes
Eligibility – sector	Eligible employers: only employers in certain sectors	Type: Binary 0: No 1: Yes
Eligibility – number of employees	Eligible employers: only employers with certain number of employees	Type: Binary 0: No 1: Yes
Conditionality – none	Conditionality for eligibility: none	Type: Binary 0: No 1: Yes
Conditionality – none	Conditionality for eligibility: the labour contract must be at least for half-time employment	Type: Binary 0: No 1: Yes
Conditionality – contract	Conditionality for eligibility: employers need to offer a contract after the subsidy expires	Type: Binary 0: No 1: Yes

A.10. Programme characteristics: Public Works

Variable name	Description	Possible answers
Infrastructure	Public work type: infrastructure development projects (public works in rural and urban areas – construction, and maintenance of public works)	Type: Binary 0: No 1: Yes
Social development	Public work type: social development and community works and services projects (children's care, sick and older person care, security, health)	Type: Binary 0: No 1: Yes
Duration	Duration of the works in months per individual (average)	Type: Binary 0: No 1: Yes
Selection: national or regional government	Works or services selected by regional or national government	Type: Binary 0: No 1: Yes

Selection: local or regional government	Works or services selected by local or regional government	Type: Binary 0: No 1: Yes
Selection: civil society	Works services selected by civil society (NGOs, communities, youth organizations)	Type: Binary 0: No 1: Yes
Selection: private sector	Works services selected by private sector	Type: Binary 0: No 1: Yes
Selection: donors	Works services selected by donors	Type: Binary 0: No 1: Yes
Minimum wage	Wage setting: programme wage in relation to the minimum wage (as stated in paper)	Type: Alphanumeric 1: Lower 2: Equal 3: Greater 4: There is no minimum wage policy
Market wage	Wage setting: programme wage in relation to the market wage for unskilled labour (as stated in paper)	Type: Alphanumeric 1: Lower 2: Equal 3: Greater 4: There is no minimum wage policy

A.11. Targeting strategy

Variable name	Description	Possible answers
Age (lower bound)	Target group of intervention: lower-bound age	Type: Numeric
Age (upper bound)	Target group of intervention: upper-bound age	Type: Numeric
Gender	Target group of intervention: gender	Type: Alphanumeric 1: Male 2: Female 3: Both
Area	Target group of intervention: area	Type: Alphanumeric 1: Urban 2: Rural 3: Both
Low-income or disadvantaged	Target group: low-income, disadvantaged, at-risk, vulnerable youth, marginalized groups, etc.	Type: Binary 0: No 1: Yes
Primary education	Target group: educational attainment: none or primary school at the start of the intervention	Type: Binary 0: No 1: Yes
Secondary education	Target group: educational attainment: secondary school at the start of the intervention	Type: Binary 0: No 1: Yes
High education	Target group: educational attainment: tertiary at the start of the intervention	Type: Binary 0: No 1: Yes
Only unemployed	Target group of intervention: only unemployed at intervention start	Type: Binary 0: No 1: Yes
Already employed	Target group of intervention: already employed or entrepreneur at intervention start	Type: Binary 0: No 1: Yes
First-time jobseekers	Target group of intervention: only first-time jobseekers	Type: Binary 0: No 1: Yes

Mandatory targeting	Target group of intervention: participation mandatory	Type: Alphanumeric 1: No (offered, active decline) 2: Yes 3: Voluntary, self-select or apply
Welfare benefits	Welfare benefits scheme provided during programme participation	Type: Binary 0: No 1: Yes
Incentives	Incentives provided to programme participants	Type: Alphanumeric 1: In-kind (non-monetary benefits, such as childcare, catering, transport) 2: Monetary benefits (stipend, transport allowance) 3: Both non-monetary and monetary 4: Salary 5: None
Certificate	Programme provides certificate to participants (only if stated)	Type: Binary 0: No 1: Yes
Monitoring	Monitoring of participants or compliance of beneficiaries	Type: Binary 0: No 1: Yes
Sanctions	Sanctions for non-participation or non-compliance (linking programme participation to benefit reception) NOTE: does not include loss of programme benefit	Type: Binary 0: No 1: Yes
Start date	Starting date of programme	Type: Date
Programme scale	Scale of programme	Type: Alphanumeric 1: National 2: Regional 3: Local
Pilot	Programme intended as a pilot to test intervention	Type: Binary 0: No 1: Yes
Programme duration	Average duration for a single cohort to stay in the programme, in months	Type: Numeric
Gender	Programme intended to target gender subgroups	Type: Binary 0: No 1: Yes
Design of programme: government	Design of programme: government	Type: Binary 0: No 1: Yes
Design of programme: NGO	Design of programme: NGO or non-profit	Type: Binary 0: No 1: Yes
Design of programme: multilateral	Design of programme: multilateral	Type: Binary 0: No 1: Yes
Design of programme: donor	Design of programme: donor-organized NGO	Type: Binary 0: No 1: Yes
Design of programme: private sector	Design of programme: private sector	Type: Binary 0: No 1: Yes
Design of programme: other	Design of programme: other	Type: Binary 0: No 1: Yes
Awareness raising	Programme includes awareness-raising activities about the programme to eligible participants	Type: Binary 0: No 1: Yes
Implementer: government	Implementer of programme: government	Type: Binary

Implementer: NGO	Implementer of programme: NGO or non-profit	0: No 1: Yes Type: Binary
Implementer: multilateral	Implementer of programme: multilateral	0: No 1: Yes Type: Binary
Implementer: donor	Implementer of programme: donor-organized NGO	0: No 1: Yes Type: Binary
Implementer: private sector	Implementer of programme: private sector	0: No 1: Yes Type: Binary
Implementer: other	Implementer of programme: other	0: No 1: Yes Type: Binary
Funding of programme: government	Funding of programme: government	0: No 1: Yes Type: Binary
Funding of programme: NGO	Funding of programme: NGO or non-profit	0: No 1: Yes Type: Binary
Funding of programme: multilateral	Funding of programme: multilateral	0: No 1: Yes Type: Binary
Funding of programme: donor	Funding of programme: donor country	0: No 1: Yes Type: Binary
Funding of programme: individual donors	Funding of programme: individual donors (foundations, companies, etc.)	0: No 1: Yes Type: Binary
Funding of programme: beneficiaries	Funding of programme: beneficiaries	0: No 1: Yes Type: Binary
Funding of programme: employer of beneficiaries	Funding of programme: employer of beneficiaries	0: No 1: Yes Type: Binary
Funding of programme: other	Funding of programme: other	0: No 1: Yes

A.12. Cost-benefit information variables

Variable name	Description	Possible answers
Costs	Study cites costs of the programme	Type: Binary 0: No 1: Yes
Cost per head	What are the per head costs of the intervention or programme (if not available at intervention level)?	Type: Dollars
Cost-benefit analysis	Study includes cost-benefit analysis	Type: Binary 0: No 1: Yes
Cost-benefit calculation	If there is a cost-benefit calculation for the intervention or programme (including "back-of-envelope"): does intervention or programme involve a net benefit, loss or is the outcome neutral?	Type: Alphanumeric 0: Neutral 1: Net benefit (benefit>costs) 2: Net loss (costs> benefit)
Time cost-benefit	If cost-benefit calculation=1: after what time does the benefit of the intervention or programme turn positive?	Type: Numeric, fill-up in months

A.13. Study-level variables

Variable name	Description	Possible answers
study_id	Impact evaluation study ID	Type: Alphanumeric, as in effect sizes tab
st_name	Study name	Type: Alphanumeric format: "Author_Year_Title_Publication" without blanks between words
st_costs	Study reports costs of the programme	Type: Alphanumeric 0: No 1: Yes
st_cba	Study includes cost-benefit analysis	Type: Alphanumeric 0: No 1: Yes
st_out	Study reports outcome not able to code	Type: Alphanumeric 0: No 1: Yes
st_out_desc	Description of outcome not coded examples: contract, days, tenure, education, etc.)	Type: String
st_subg	Study reports subgroup analysis not able to distinguish through existing variables	Type: Alphanumeric 0: No 1: Yes
st_subg_desc	Description of subgroup analysis not coded (example: single mothers, etc.)	Type: String
st_eprob	Authors mentions empirical identification problems or shortcomings of the method	Type: Alphanumeric 0: No 1: Yes
st_eprob_desc	Description of empirical identification problems as stated by the author	Type: String
st_impl_prob	Study mentions programme implementation problems	Type: Alphanumeric 0: No 1: Yes
st_geneq	Study mentions general equilibrium effects (such as deadweight losses and substitution effects)	Type: Alphanumeric 0: No 1: Yes
st_geneq_desc	Description of general equilibrium effects (such as deadweight loss, substitution effects)	Type: String
st_comm	Other comments	Type: String Note everything that was difficult for coding this study. Note any particularities regarding the intervention design.

A.14. Risk of bias assessment variables

Variable name	Description	Possible answers
rob_desi_dat	Design: Data: study reports pre-post intervention data for both groups.	Type: Alphanumeric 0: No 1: Yes
rob_bias_bala	Selection bias: Balance: are the mean values or the distributions of the covariates at baseline statistically different for beneficiaries and non-beneficiaries (p<0.005)	Type: Alphanumeric 0: No 1: Yes
rob_bias_coun	Selection bias: Confounding: if there are statistically significant differences in covariates relevant for the model between beneficiaries and non-beneficiaries,	Type: Alphanumeric 0: No 1: Yes

	are these differences controlled for in the impact evaluation?	
rob_bias_attr	Selection bias: Attrition: indicate attrition level (such as ratio between those without end-line information and all included at baseline)	Type: Numeric between 0 and 1
rob_perf_hawt	Performance bias: Hawthorne effect: Are differences in outcomes across the groups influenced by participant motivation because of programme implementation and, or monitoring?	Type: Alphanumeric 0: No 1: Yes
rob_perf_john	Performance bias: John Henry effect: Is control group outcome influenced by participants' exposure to the intervention or motivation?	Type: Alphanumeric 0: No 1: Yes
rob_perf_spil	Performance bias: Spillovers: Are comparisons sufficiently isolated from the intervention (participants and non-participants are sufficiently geographically or socially separated) or are spillovers estimated by comparing non-beneficiaries with access to the intervention to nonbeneficiaries without access to the intervention and/or through social network analysis?	Type: Alphanumeric 0: No 1: Yes
rob_sign_unit	Statistical significance: Unit of analysis problem: Does the study account for lack of independence between observations within assignment clusters if the outcome variables are clustered?	Type: Alphanumeric 0: No 1: Yes
rob_sign_hete	Statistical significance: Heteroskedasticity: Do the authors control for heteroskedasticity and/or use robust standard errors?	Type: Alphanumeric 0: No 1: Yes
rob_rand_equi	Methodology implementation: Randomization: Do the authors report the statistical power (such as one minus probability of error type I) to determine sample size or conversely report the statistical power of the sample size?	Type: Alphanumeric 0: No 1: Yes
rob_disc_allo	Methodology implementation: Regression discontinuity designs: Is the allocation of the programme based on a predetermined continuity on a continuous variable and blinded to the beneficiaries or if not blinded, individuals cannot reasonably affect the assignment variable in response to knowledge of the participation rule?	Type: Alphanumeric 0: No 1: Yes
rob_psm_matc	Methodology implementation: PSM: Are beneficiaries and non-beneficiaries matched on all relevant characteristics?	Type: Alphanumeric 0: No 1: Yes
rob_psm_excl	Methodology implementation: PSM: Does the study exclude observations outside the common support?	Type: Alphanumeric 0: No 1: Yes
rob_psm_exc	Methodology implementation: PSM: Does the study exclude observations outside the common support?	Type: Alphanumeric 0: No 1: Yes
rob_inst_sign	Methodology implementation: Instrumental variable estimation: Are the instruments individually significant ($p \leq 0.05$)?	Type: Alphanumeric 0: No 1: Yes
rob_inst_exog	Methodology implementation: Instrumental variable estimation: Does the study qualitatively assess the exogeneity of the instrument or identifier (both externality as well as why the variable should not enter by itself in the outcome equation)?	Type: Alphanumeric 0: No 1: Yes
rob_heck_excl	Methodology implementation: Heckman models: Is there is a variable that is statistically significant in the first stage of the selection equation and excluded from the second stage?	Type: Alphanumeric 0: No 1: Yes
rob_data	Level of risk of bias due to study's data-collection design and how it is reported.	Type: Alphanumeric This is automatically filled-up once you fill questions related to this type of bias.

rob_bias	Level of risk of bias due to selection bias and confounding	1: Low risk of bias 2: Some concerns 3: High risk of bias Type: Alphanumeric This is automatically filled-up once you fill questions related to this type of bias.
rob_perf	Level of risk of bias due to performance: behavioural responses to the intervention and spillovers	1: Low risk of bias 2: Some concerns 3: High risk of bias Type: Alphanumeric This is automatically filled-up once you fill questions related to this type of bias.
rob_sign	Level of risk of bias due to statistical significance and statistical power of the sample	1: Low risk of bias 2: Some concerns 3: High risk of bias Type: Alphanumeric This is automatically filled-up once you fill questions related to this type of bias.
rob_meth	Level of risk of bias due to hard assumptions during the implementation of the statistical methodology	1: Low risk of bias 2: Some concerns 3: High risk of bias Type: Alphanumeric This is automatically filled-up once you fill questions related to this type of bias.
rob_overall	Overall risk of bias. Please report the overall risk of bias in the log shared file in drive "2021_06_24_Full screening and coding follow up tool"	1: Low risk of bias 2: Some concerns 3: High risk of bias
