What Works for Active Labor market Policies? A Meta-Analysis of RCT Findings

Autoría ditelliana: Levy Yeyati, Eduardo; Montané, Martín; Sartorio, Lucas
Otras autorías: Gauna, Alfonso (*University of Chicago, US)*

Fecha de publicación: 19/03/2024

Publicado originalmente en: Economía LACEA Journal (e-ISSN: 1533-6239)

Economía LACEA Journal

What Works for Active Labor Market Policies? A Meta-Analysis of RCT Findings

RESEARCH

LSE Press

EDUARDO LEVY YEYATI
MARTÍN MONTANÉ
LUCA SARTORIO
ALFONSO GAUNA
*Author affiliations can be found in the back matter of this article

ABSTRACT

The past 10 years have witnessed a flurry of Randomized Controlled Trials (RCTs) that shed new light on the impact and cost-effectiveness of Active Labor Market Policies (ALMPs) that aim to improve workers´ access to new jobs and better wages. This paper presents the first systematic review of 102 ALMP RCTs comprising a total of 668 estimated impacts. We find that (i) a third of these estimates are positive and statistically significant (PSS) at conventional levels; (ii) programs are more likely to yield positive results when GDP growth is higher and unemployment is lower; (iii) programs aimed at building human capital, such as vocational training, independent worker assistance and wage subsidies, show significant positive impact; and (iv) program length, monetary incentives, individualized follow up and activity targeting are all key features in determining the effectiveness of the interventions.

CORRESPONDING AUTHOR:
Eduardo Levy Yeyati
School of Government at Universidad Torcuato Di Tella, AR
ely@utdt.edu

KEYWORDS:
vocational training; labor policies; wage subsidies; randomized controlled trials

JEL CLASSIFICATION CODES:
J21; J48; E24

TO CITE THIS ARTICLE:
Levy Yeyati, Eduardo, Martín Montané, Luca Sartorio, and Alfonso Gauna. 2025. "What Works for Active Labor Market Policies? A Meta-Analysis of RCT Findings." *Economía LACEA Journal* 24(1): 81–104. DOI: https://doi.org/10.31389/eco.450

DOI: 10.31389/eco.450

1. INTRODUCTION

In 2019, public spending in Active Labor Market Policies (ALMPs) have accounted for more than 0.5 percent of the GDP of OECD countries. ALMPs is a general denomination for several specific policies that could be broadly grouped into four big clusters: vocational training, assistance in the job search process, wage subsidies or public works programs, and support to micro-entrepreneurs and independent workers.

ALMPs are inherently complex interventions and their incidence depends on a broad range of variables that are associated with design, context, and with implementation. This has been implicitly assumed by various recent meta-analyses that centered not so much on the aggregated impact but rather on what specific features make them work (Card et al., 2010; Card et al., 2017; Escudero et al., 2017). However, these meta-analyses usually group different evaluation approaches together, combining quantitative assessments through a simplified metric that merges diverse interventions and contexts to get more statistical powerful results at the expense of precision.

Fortunately, the past 10 years have witnessed a flurry of experimental evaluations through Randomized Controlled Trials (RCT) that have shed new light on the impact and cost-effectiveness of ALMPs (Figure 1). This large production of experimental evaluations of ALMPs in recent years has allowed researchers to generate relatively large samples of RCTs that enable a meta-analysis with greater precision and representative potential, and also a greater depth in the insights that can be derived from this information.

In this paper, we contribute to the blooming literature about ALMPs in two ways. First, we focus exclusively on programs evaluated through Randomized Controlled Trials (RCTs). This choice was not without tradeoffs as it reduced the number of relevant evaluations but allowed us to focus on estimates with high internal validity and to refine the metrics used to analyze results, making the findings from individual evaluations more naturally comparable. The high number of experimental evaluations carried out since 2014 allowed us to assemble a sample large enough to consider exclusively this "gold-standard" methodology that usually represents a minority part of the most extensive reviews of the literature, such as Card et al. (2010) or Card et al. (2017); indeed, two thirds of our sample come from papers published in 2014 or after. In the process, we collected data from old and recent evaluations of ALMPs' effectiveness to build a workable dataset of 668 Intent-to-Treat (ITT) impact estimates on employment and income variables from 102 interventions around the globe, evaluated through 73 rigorous impact evaluations with an experimental design, and which covered the four broad groups of ALMPs mentioned above. To our knowledge, this constitutes the most extensive review of the available empirical evidence on this group of policies that is entirely composed of impact evaluations based on an experimental design.[1]

Second, we created a metric of new variables to capture the key (implementation, context, and target) determinants of ALMP success. More specifically, following Pritchett et al. (2013), we described the design space of ALMPs, i.e. a parsimonious version of the space of all of the possible instances of a class of policy, arrived at by specifying all of the choices necessary for a project to be implemented. To characterize all of the dimensions of this design space we granularly identified a set of standardized variables that tracked (i) the specific elements in which the programs can be decomposed; (ii) the implementation features and the type of public-private participation involved; and (iii) the economic context and the target population of the programs. This allowed us to refine the analysis and identify why policies that are similar on paper can differ in their impact and cost-effectiveness due to differences in their design, implementation, or operating context.

Comparing the overall impact of the four policy clusters analyzed, we found that wage subsidies and supports to independent workers showed the greatest median impact in earnings relative to the control group, with improvements of 16.7 percent and 16.5 percent, respectively. On the other hand, vocational training programs had a median impact of 7.7 percent, while employment services showed an almost negligible impact.

1 Kluve et al.'s (2019) meta-analysis also benefits from this recent batch of ALMPs' RCTs, but they restrict attention to youth-targeted programs and complement their sample with other evaluation approaches.

The reported impacts of ALMPs on employment and earnings outputs, although moderately positive on average, were subject to a large variability, possibly due to the multidimensional design space of these policies. To address this, we developed meta-analytic regressions that exploited the descriptive granularity of the proposed design space, seeking to identify policy components associated with a greater probability of success.

Wage subsidies showed the greatest impact on labor earnings and employment relative to the control group, followed by supports to independent workers and vocational training programs, while the incidence of employment services was almost negligible, but all types of ALMPs showed great variability in reported impacts. Regarding the specific design and implementation features of ALMPs, we documented that individualized coaching and/or follow-up of the participants, the targeting of the training to a specific industry or occupation, and giving monetary incentives to trainees all correlated with better outcomes in vocational trainings programs (the most frequent ALMPs in our dataset). Regarding the contexts in which ALMP tends to be more effective, we showed that the performance of this kind of program positively correlates with per capita GDP growth and negatively with the unemployment rate in the year of implementation.

Although there is little evidence on the delivery costs of the programs, we did find greater volatility in independent training programs, relative to employment services or vocational training programs. On average, employment services were inexpensive, and they stressed the need of further research to focus on cost effectiveness analysis.

The rest of the paper is organized as follows. In Section 2, we present in detail the organization and characteristics of the ALMPs' design space. In Section 3, we describe our database of 102 experimental evaluations. In Section 4, we analyze the effectiveness of the different classes of ALMPs and present the results of the quantitative meta-analysis. In Section 5 we conclude.

2. ACTIVE LABOR MARKET POLICIES AND THEIR DESIGN SPACE

THE MULTIDIMENSIONAL NATURE OF ACTIVE LABOR MARKET POLICIES

The effectiveness of multidimensional and complex policies, such as ALMPs, depends on how they were specifically designed, on the quality of their implementation, on the context in which they were developed, and on their target population. For example, a vocational training program may differ in its cost and duration; in its curricular content; in whether or not, and how, the private sector participates; and it may address a very diverse audience, from experienced software programmers in New York or Berlin to disadvantaged youth in the state of Madhya Pradesh. The number of potentially relevant factors and related aspects may quickly render the dimensionality of the analysis practically unmanageable.

Following Pritchett et al. (2013), our four groups (vocational training; wage subsidies or public works programs; support to micro-entrepreneurs and independent workers; and assistance in the job search process) can all be thought of as "classes" of policies that could be designed and implemented in very different ways and could target diverse demographic groups, with widely varying effectiveness. A review that does not consider this variability could conclude that "wage subsidies work" or "vocational training does not work", statements as imprecise as "the ingestion of chemical components works".[2] To account for the dimensionality of the problem in an operational way, we need to review the evidence from the perspective of a simplified *design space*, namely, a parsimonious version of the space of all of the possible instances of a class of policy, arrived at by specifying all of the choices necessary for a project to be implemented.

2 Pritchett et al. point out that the question "Does the ingestion of chemical compounds improve human health?" is under-specified, as some chemical compounds are poison and some are aspirin or penicillin and their effects will vary widely depending on the frequency of the applied dose or the particular conditions of the individual in which it is applied. According to these authors, the currently conventional approach to "the evidence" is of limited value as a result of the inability to extrapolate the lessons of an impact evaluation of a particular policy to the analysis of another policy that has small changes in some elements of its design (lack of "construct validity") or that has different target populations or is implemented in a different context (lack of "external validity").

Systematic reviews generally consider their policies evaluated as low-dimensional (uniform and with few relevant decisions to make in their design) and with smooth and non-contextual response surfaces.[3] This could be the case of purely "logistical tasks",[4] such as conducting vaccination campaigns in which once we know the "optimal design" of the medical solution and all its contraindications and requirements; the effects will have a strong homogeneity and effectiveness in very dissimilar contexts and there will be almost no relevant decisions to be made in their design beyond ensuring the application of standardized protocols of proven performance (i.e. to ensure the application of vaccines). But ALMPs are generally complex policies with high-dimensional design spaces and rugged and contextual response surfaces, and are highly dependent on a good implementation where there is no "optimal design" that will be unfailingly effective in any social, political, or cultural context. Any systematic review that does not exhaustively describe the design space of the policies evaluated or considers the existing variability within the same intervention class, or their interactions with the context and the target population, may have limited use from a practical policy perspective.

A DESIGN SPACE FOR CHARACTERIZING ACTIVE LABOR MARKET POLICIES

Developing a complete and exhaustive design space that homogeneously portrays the evaluated policies requires building a set of standardized variables that characterize at least five fundamental dimensions of the ALMP, such as i) its type; ii) its specific components; iii) the way it is implemented (including the nature of public-private involvement); iv) its implementation cost; and v) its implementation context and its target population.

To characterize these dimensions, we analyzed all of the information available in the academic publications and condensed the description of the criteria used for the identification of each of the variables into unified protocols that articulated the review process. Table 1 summarizes all the variables considered across the five dimensions of the design space. Table A2 in the appendix provides more detail on how these variables were precisely identified across the different studies.

Table 1 Design space taxonomy. NGO: non-governmental organization. PPP: purchasing power parity.

ALMP Type
Vocational trainings
Wage subsidies or public works programs
Supports to micro-entrepreneurs and independent workers
Assistances in the job-search process
Content components
Training in technical skills
Training in soft skills
Individualized mentoring and/or follow-up
Assistance services in the job-search process
Monetary transfer or expenses financing
Voucher system
Asset transference
Loan

(Contd.)

3 Pritchett et al. introduce the concept of *response surface*, defined as the average gain on a target indicator of a selected population exposed to a specific program (as an element of the overall design space) compared to those of an ex-ante identical population not exposed.

4 Following Andrews, Pritchett & Woolcock (2017), a policy intensive in logistical tasks is a policy that requires an important number of agents to implement this policy (is "intensive in transactions") but who do not need to make significant decisions (they do not require "local discretion" from their agents) beyond following established protocols based on known and proven technologies.

DOI: 10.31389/eco.450

Implementation features	
	Duration (in months)
	Participation of the non-public sector (private firms, NGOs or multilateral organizations)
	Active involvement of the non-public sector in program design
	Active financing of the non-public sector
	Content provision by the non-public sector
	Program implemented as field experiment and/or pilot
	Orientation to specific economic sectors or occupations
Implementation costs	
	Average cost per participant (2010 PPP dollars)
Context and target population	
	Average, minimum, and maximum age of the participants
	Proportion of participants with a high-school degree
	Proportion of participants with a university degree
	Average years of study of the participants
	Proportion of unemployed participants at the beginning of the program
	Female/Male proportion of participants

Types

As noted, our taxonomy classifies four big types of ALMP for our design space. First, we considered **vocational trainings**, which are programs essentially based in the provision of technical skills to develop a particular professional activity. Although these can be complemented with additional components, such as hiring subsidies or soft-skills training, the fundamental goal of these programs is to provide technical human capital for salaried workers.

Second, we analyzed **wage subsidies or public works programs**. These are initiatives in which the State explicitly encourages the hiring, either through monetary incentives for private employment (wage subsidies programs) or through direct hiring by the state (public works programs), without any explicit instance of formal training.[5]

Third, we reviewed **supports to micro-entrepreneurs and independent workers**. These are projects that seek to impact employment levels and labor earnings by assisting microenterprises or self-employed workers. They typically include job skills training or business management training (or both), which are explicitly targeted to self-employed workers and are sometimes complemented by cash transfers, subsidized loans, or asset transfers.

Finally, we included **assistances in the job-search process**, which are programs whose fundamental objective is to help individuals to be successful in their salaried job search process without any explicit instance of formal training[6] and/or monetary incentives to private or public employers (wage subsidies or public works).[7] It is the most heterogeneous category of ALMPs as it includes very diverse interventions, such as the organization of job fairs with private employers, workshops to build CVs or digital professional profiles or individualized counseling and mentoring programs during the job search process, among others.

Our design space described the policy type of each program using a unique categorical variable that indicates one of the four possible options.

5 Otherwise, the program is classified as a vocational training.

6 Otherwise, the program is classified as a vocational training.

7 Otherwise, the program is classified as a wage subsidy or a public works program.

Content Components

Next, we granularly mapped all of the individual components of the program design of each intervention through eight dummy variables that describe the content of each ALMP. For example, this allowed us to map if a particular intervention offered specific elements, such as training in technical skills, training in soft skills, individualized mentoring or follow-up, loans, asset transferences, among others.

This content description allowed us to describe more precisely the value proposition of each program. For example, even if the main intervention of an ALMP was training (classifying its program type as "Vocational training"), it could be complemented with some content components from other program types, such as some services in the job-search process.

IMPLEMENTATION FEATURES

Even if the specific components of the content of the policies are similar, there may be considerable variations in their delivery, in the duration of the program, in the existence of incentive schemes for a correct implementation, or in the nature of the public-private collaboration in their execution. This dimension tries to capture variables of the design space that characterize the incentives and the nature of the implementation of the policy, independently of its curricular and formative content.

IMPLEMENTATION COSTS

When the information was available, we added a continuous variable that identifies the average cost per person of the intervention, in 2010 purchasing power parity (PPP) dollars. It is important to stress that only 51 interventions reported this critical variable, and only 22 carried out a rigorous cost-benefit analysis by means of net present value (NPV), internal rate of return (IRR) or payback periods, highlighting an important limitation of the usual practice in the impact evaluation literature.

CONTEXT AND TARGET POPULATION

Finally, our dataset also contains information about the context in which the policies were implemented, and identifies the country and the cities in which the programs were developed, including whether the intervention was conducted in a rural or urban area and some key demographic features of the target population, such as age, educational and gender distribution.

3. DATA

To carry out a systematic review of ALMPs, we focused on policies evaluated through an RCT published either in academic journals or reports from think-tanks or multilateral organizations, or even in current working papers. This exclusive focus in experimental evaluations may have some eventual drawbacks, as one could argue that international agencies and governments might be more prone to fund RCTs of programs with larger predicted effectiveness or that programs evaluated by this methodology might not always be representative of policies implemented at large scale.[8] However, we believe that this criterion allowed us to concentrate within a methodology that has high standards of internal validity, and generated a relatively large sample of evaluations that provided us with greater precision in the analysis while at the same time avoiding comparability problems by combining different evaluation criteria. During the last decade, there was a very important number of experimental evaluations of ALMPs on a worldwide scale (Figure 1) and being able to include these recent evaluations allowed us to work with what we believe to be a relatively large sample that allowed us to make informative inferences without introducing comparability problems of the estimated effects.

8 This does not deny the convenience of eventually comparing our results with those drawn from a broader meta dataset.

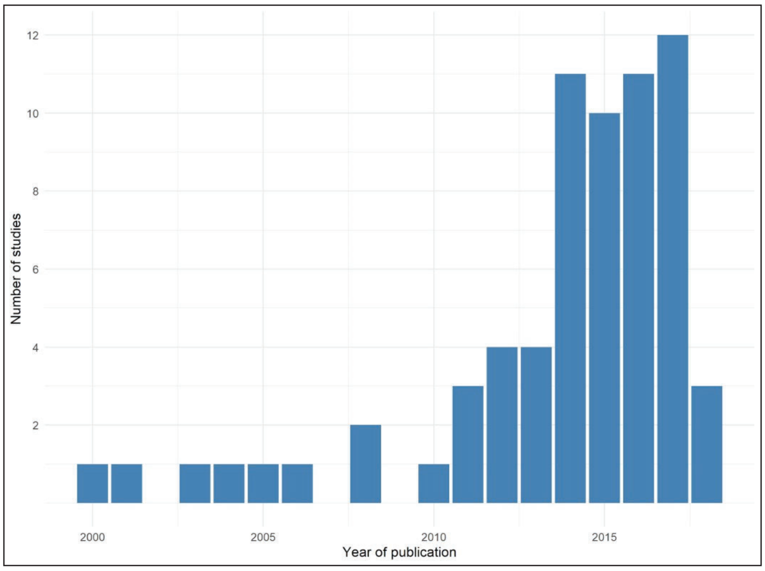

Figure 1 Distribution of studies included in our sample according to the year of publication.

The studies reviewed in this work were collected from two sources. First, we considered ALMPs evaluations from other previous studies such as Card et al. (2017) and McKenzie (2017). These sources were supplemented by a comprehensive search for ALMPs impact evaluations on Google Scholar.

Then, to build our final sample of evaluations, we imposed three selection criteria to filter studies of interest for this work. First, we discarded all studies that we could not clearly classify into any of the four large types of ALMPs previously defined. Then, we only included ALMPs that had been evaluated through RCTs. Finally, we only considered policies whose evaluations had documented two outcome variables that we defined to evaluate the effectiveness of ALMPs: the impact on the total or labor income of individuals and the impact on the employment levels of the participants.

The process of filtering and selecting evaluations for the final sample is detailed in Figure 2. We did not impose any geographic filter or select by type of participants, be it age, gender, or economic level, unlike other meta-analyses in the literature. In addition, the review includes both policies that have been effectively implemented by governments and experiments or pilot projects. Finally, the review considers RCTs published in peer-reviewed academic journals and working papers as well as reports from multilateral institutions and research organizations. This allowed us to generate a final sample with 73 papers or studies that contemplate 102 interventions, programs or policies evaluated through an RCT.[9]

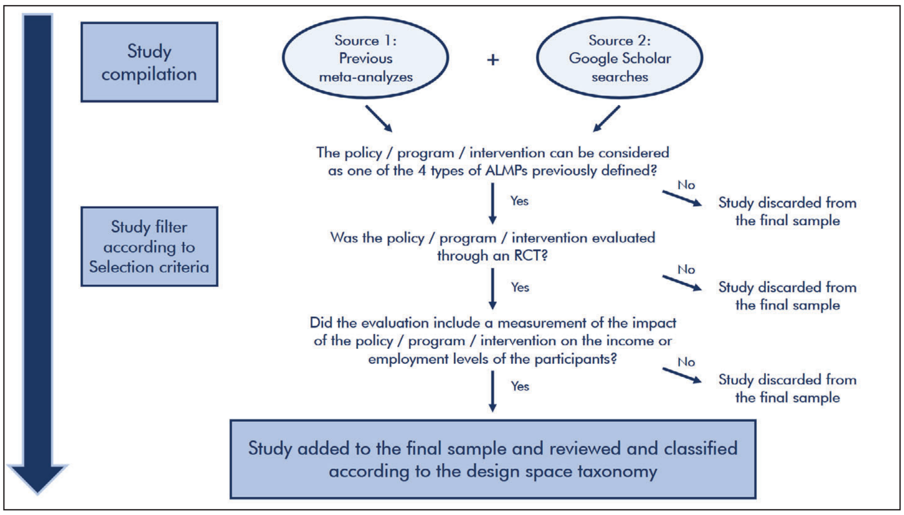

Figure 2 Review criteria and selection and filter process of ALMP impact evaluations.

9 Available on request.

4. META-ANALYSIS

SAMPLE STRUCTURE

Our sample has a high percentage of vocational trainings (45.1 percent), in line with other meta-analyses. Entrepreneurial training and employment services are almost evenly represented with 24 (23.5 percent) and 22 (21.6 percent) interventions respectively. Wage subsidies and public works programs account for the remaining 10 percent of interventions (Table 2).

Regarding the target of the programs, only 10 interventions were specially designed for women, and 23 targeted exclusively participants that were 29 years old or younger. Although the share of programs that targeted the youth may look small, the unweighted average of the mean age reported by all the interventions in the sample is only 27.7 years: either by design or by demand, participants tend to rather young.

The interventions took place in countries with widely different income levels, with approximately one half of them being implemented in lower-income or lower-middle-income economies.

TYPE OF INTERVENTION	n	%
Skill training	46	45.1
Entrepreneurship training/promotion	24	23.5
Employment services	22	21.6
Wage subsidies	7	6.9
Public works	3	2.9
Target groups		
Gender		
Only women	10	9.8
All genders	92	90.2
Age		
Maximum age 29 or less	23	22.5
Maximum age 30 or more	36	35.3
No age restrictions	43	42.2
Country income		
Low-income	27	26.5
Lower-middle-income	22	21.6
Upper-middle-income	20	19.6
High-income	33	32.4
Region		
Africa	31	30.4
Latin America and the Caribbean	20	19.6
North America	25	24.5
Asia	18	17.6
Europe	8	7.8
Information about costs		
Costs per participant	51	50

Table 2 Summary statistics on selected variables of our sample of 102 interventions.

Table 3 presents details of the implementation features of the programs in our sample. We can see that the participation of the non-public sector (private sector, non-government organizations(NGOs) or multilateral organizations) was highly frequent in our full sample (66.3 percent), especially in entrepreneurship promotion or assistance to independent worker programs (70.8 percent) and vocational trainings (88.9 percent) while this involvement of the non-public sector was considerably less frequent (27.3 percent) in our set of assistance in the job search process programs (or "employment services"). When considering the full sample, we can appreciate that this participation of the non-public sector usually implied its involvement in the direct provision of the main component of the policy (59.4 percent) while its participation in the design (39.7 percent) or the financing (17.8 percent) of the policy was less frequent.

Moreover, we observe that almost half of our interventions were field experiments or pilot projects (47.5 percent) and that only 22.8 percent had an explicit orientation to a particular set of economic sectors, a feature that was relatively more usual in vocational trainings (33.3 percent) and in entrepreneurship promotions or assistances to independent workers (29.2 percent).

Table 4 displays some descriptive statistics regarding the content components of the four classes of ALMPs. Interestingly, vocational trainings were not necessarily limited to technical training modules, but they tended to include with relative high frequency a component on soft-skills training (48.9 percent) and a work experience, internship or on-the-job training component (46.7 percent). However, only in 6.7 percent of the cases in our sample did vocational trainings offer a voucher system with which the participant could freely choose their training provider, and these programs usually discretionally assigned a provider selected by the policymaker/designer to the beneficiary. These programs also did not tend to include job-search assistance services, such as mentoring for interviews, help in designing a CV or digital profile, or the organization of job fairs (6.7 percent of total cases in the sample).

Assigning mentors or to include an individualized follow up of the participants tended to be also a relatively usual practice in employment services (50 percent), in assistances to entrepreneurs or self-employed workers (54.2 percent) and in vocational trainings (42.2 percent). On the other hand, giving some monetary transfers to participants was not a frequent feature of employment services (13.6 percent) or of vocational trainings (24.4 percent). Regarding the financial assistance component of entrepreneurship promotion programs, it usually took the form of a direct monetary transfer (33.3 percent) or asset transferences (37.5 percent) but less frequently through loans (8.3 percent).

Table 3 Summary statistics on implementation features of our sample of 102 interventions.

TYPE OF PROGRAM	EMPLOYMENT SERVICES (%)	ENTREPRENEURSHIP TRAINING/PROMOTION (%)	VOCATIONAL TRAINING (%)	WAGE SUBSIDIES (%)	FULL SAMPLE (%)
Participation of the non-public sector	27.3	70.8	88.9	50.0	66.3
Participation of Private Firms	22.7	20.8	46.7	30.0	33.7
Participation of NGOs	4.6	41.7	40.0	20.0	30.7
Participation of multilateral organizations	0.0	8.3	0.0	0.0	2.0
Participation of non-public sector in the design	4.6	54.2	53.3	20.0	39.6
Participation of non-public sector in the financing	0.0	25.0	20.0	30.0	17.8
Participation of non-public sector in the implementation	27.3	62.5	82.2	20.0	59.4
Field experiment or pilot	54.6	58.3	40.0	40.0	47.5
Orientation to a specific sector	4.6	29.2	33.3	0.0	22.8

TYPE OF PROGRAM	EMPLOYMENT SERVICES (%)	ENTREPRENEURSHIP TRAINING/PROMOTION (%)	VOCATIONAL TRAINING (%)	WAGE SUBSIDIES (%)	FULL SAMPLE (%)
Training in soft skills	4.6	29.2	48.9	50.0	34.7
Training in technical skills	9.1	66.7	88.9	0.0	57.4
On-the-job training	0.0	0.0	46.7	100.0	30.7
Mentoring	50.0	54.2	42.2	10.0	43.6
Assistance in the job-search process	95.5	0.0	6.7	0.0	23.8
Monetary transfer	13.6	33.3	24.4	0.0	21.8
Voucher System	0.0	0.0	6.7	0.0	3.0
Asset transference	0.0	37.5	0.0	0.0	8.9
Loan	0.0	8.3	0.0	0.0	2.0
Average duration (in months)	3.73	8.18	5.38	5.11	5.66

Table 4 Summary statistics on content components of our sample of 102 interventions.

DESCRIPTIVE EVIDENCE

Impacts are quantified by the evaluation coefficients. To a question of the type: 'what is the effect of a training program on the labor income of program participants relative to comparable workers that do not participate in the program?', an evaluation study would respond with a coefficient that identified the mean differential effect over the "treated" sample (participants) relative to the control sample (non-participants), as well as some measure of the precision of the coefficient (ideally, its standard error). In our study, we concentrated only on Intent-to-Treat (ITT) effects to ensure an easier interpretation of our aggregate metrics. It should be noted that, due to imperfect compliance, this choice may underestimate the true impacts of the programs when compared with other relevant outcomes, such as Treatment-on-the-Treated (TOT) or Local Average Treatment Effects (LATE).[10]

From our dataset, we collected 668 coefficients obtained from 102 interventions, i.e. approximately six coefficients per intervention. By grouping these coefficients according to the type of program and the outcome category, we find that (i) there are slightly more coefficients related to employment outcomes (55 percent) relative to earnings and, more importantly, (ii) vocational training accounts for 54 percent of all coefficients. As some types of interventions, such as wage subsidies, are underrepresented in our sample, we worked with several meta-regressions with different subsamples to address this bias. Figure 3 illustrates this composition.

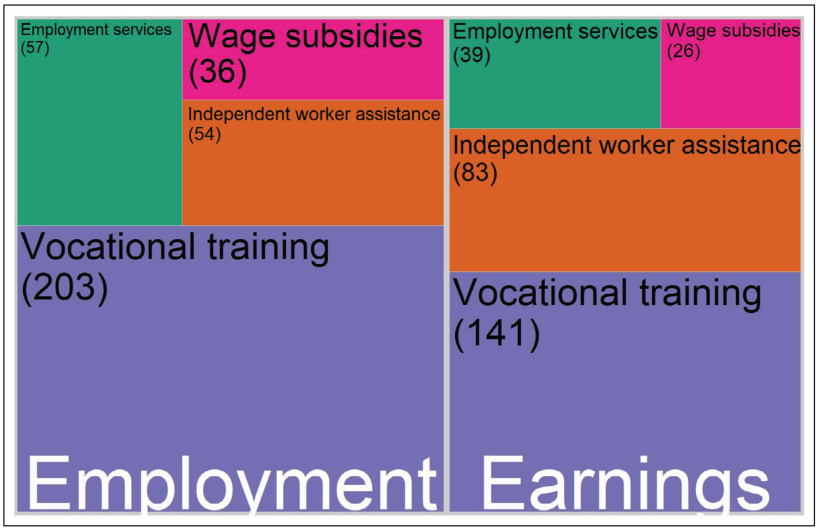

Figure 3 Distribution of the 668 reported coefficients in the 73 selected articles. The left-hand side shows the composition of employment-related outcomes. The right-hand side shows the composition of the earnings-related outcomes.

10 While it would be useful to verify that the compliance rate does not vary significantly across types of training to ensure that the variation in impact is mostly driven by the type of training and not by differing take-up behavior, unfortunately the studies do not include information on take-ups.

If we focus on the median impact on earnings, wage subsidies and independent worker assistance show the greatest impact relative to the control group, with improvements of 16.7 percent and 16.5 percent, respectively. Vocational training programs have a median impact of 7.7 percent, while employment services show an almost negligible impact. The median impact on employment outcomes exhibits a similar pattern, with wage subsidies being the type of program that reports the highest impact on this outcome category, while independent worker assistance and vocational training show a median impact of 11 percent and 6.7 percent, respectively. Interestingly, employment services interventions have a median impact of 2.6 percent, which is consistent with short-lived and inexpensive interventions that do not attempt to help build human capital, but rather aim to improve the propensity of the participants to find employment.

Importantly, there is a substantial variability in reported impacts on earnings and employment outcomes. This is especially true for the types of interventions in which we have more than 10 cases, such as employment services, independent worker support or assistance and vocational training (Figure 4). We think that this is evidence of the multidimensional nature of the design space in which the programs we analyzed are deployed. This variability suggests that the impact should be adjusted according to the different components of each program, the context in which it is implemented, and the target population for which it was designed.

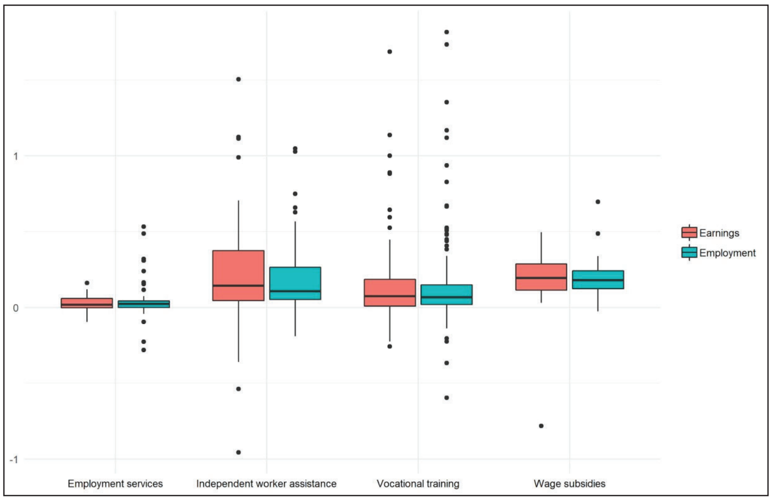

Figure 4 Boxplot of the 668 coefficients according to the estimated effect relative to the control group. Group by type of program and outcome category. Boxes represent the 50 percent central coefficients reported. The horizontal lines show the median value. The vertical lines show the last coefficient that falls into the +/–1.5*interquartile range (IQR) limit. Points are observations that lay above or below the +/–1.5*IQR limit.

One key question in this type of intervention is whether the ALMPs have a lasting effect on the participants. Impact evaluations tend to focus on the short term, especially when they rely on survey data instead of on administrative records. Fortunately, our dataset of studies included reports that estimate the effect of programs even after three or four years. The reported coefficients are most dispersed in the first years, including some negative outcomes, and become gradually less volatile after the second year. More interestingly, the coefficients do not seem to lose their statistical significance over time, as illustrated by the 28 coefficients reported after 4 years of program completion (Figure 5).

Although the sample of ALMPs for which we had cost data was limited, we identified some indicative patterns (Figure 6). Wage subsidies, support to independent workers or micro entrepreneurs and vocational trainings have comparable median cost per participant, ranging from 1744 and 1518 2010 PPP U.S. dollars, with much greater variability in the second group. In turn, employment services are notably less expensive policies, with a median cost per participant of 277 2010 PPP U.S. dollars and limited variability across programs.

These important differences in median costs – and the dispersion in the case of independent workers' programs – highlights the need for precise information on the unit costs of these policies as the policy maker should care not just about what works, but also about what is cost-effective.

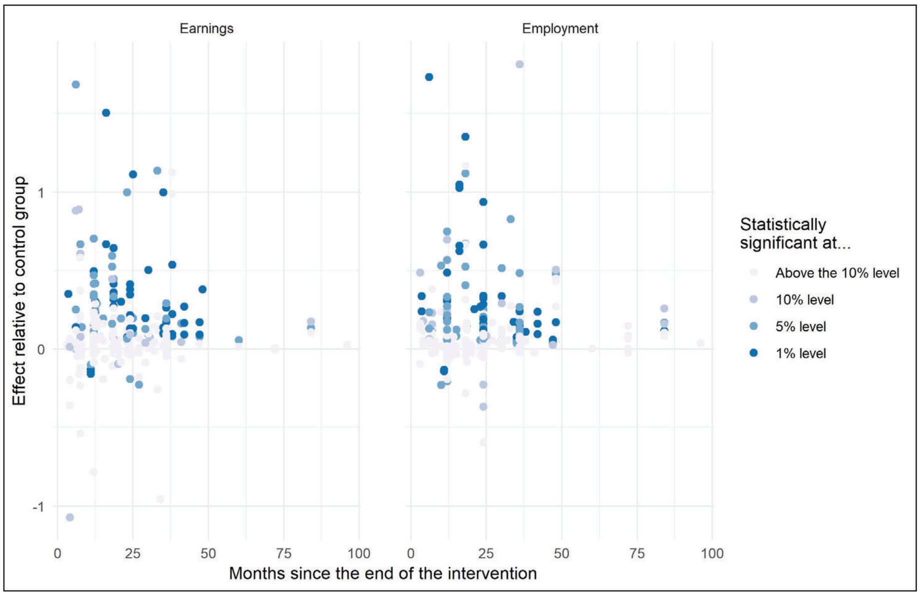

Figure 5 Coefficient reported on earnings and employment status of the treatment group relative to the mean of the control group by months after the end of the program.

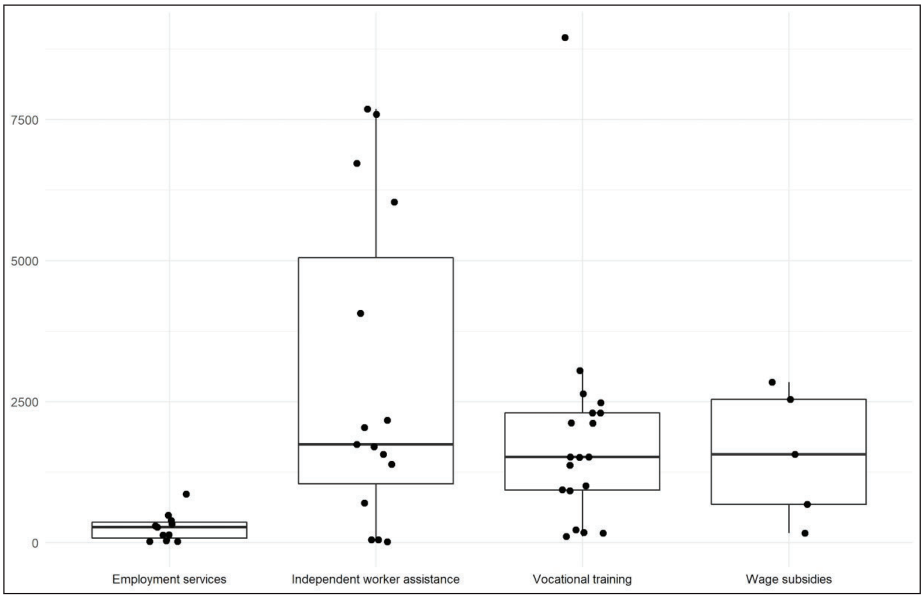

Figure 6 Boxplot of unit costs: cost per treated participant by four-way program classification. 2010 PPP US Dollars. Boxes represent the 50 percent central values reported. The horizontal lines show the median value. The vertical lines show the last coefficient that falls into the +/–1.5*IQR limit. Points are observations that lay above or below the +/–1.5*IQR limit.

META-REGRESSION ANALYSIS: METHODOLOGY

We performed a meta-regression analysis by estimating two different model specifications. First, we created a Positive and Statistically Significant (PSS) variable that was in line with past meta-analyses (Card et al., 2017; Escudero et al., 2017; Kluve et al., 2019). Then, we ran a meta-regression using a mixed-effects model, using the relative effects and its standard error. We used both measures of effectiveness because some authors do not report the standard errors or an exact continuous p-value of their interventions and document only the significance levels at which the null hypothesis was rejected. This makes it harder to estimate models that require both a reported ITT and its standard error.

PSS Regression

The justification for the PSS regression can be shown as follows. Assume that the estimated impact of an active labor market program on the outcomes of the participants (b) has an approximately normal distribution with mean (β) and a precision (P) that depends both in the sample size and the design features of the study.[11]

11 We use the notation used in Card et al. (2017).

Under this usual assumption, an estimate of one of our studies can have the following representation:

$$b = \beta + P^{-1/2}z \tag{1}$$

where z is a realization of an approximately normal distribution that, if the sample size is big enough, will be close to N (0, 1). $P^{-1/2}z$ can be interpreted as the sampling error captured by *b*.

In turn, we can assume that the limiting program effect, which is the parameter that the impact evaluation studies are trying to estimate, can be decomposed into two terms:

$$\beta = X\alpha + \varepsilon \tag{2}$$

where the multiplication of α, a vector of coefficients, and *X*, that captures the observed sources of heterogeneity (our design space), measures the "explained" part of the effect on the treated, and ε, an unexplained error term that captures other particularities of the programs that could not be observed or are not included in our taxonomy.

Equations (1) and (2) can be combined, yielding the following model of the observed effects:

$$b = X\alpha + \mu \tag{3}$$

where *X* contains the observed characteristics of the programs reflected in our taxonomy, α is the set of parameters that measures the average contribution of the characteristics and μ is the observed error term containing both the unobserved effects and the sampling error.

Unfortunately, because some of the studies do not report the data required to estimate this equation, namely, the estimated average treatment effect *and* its standard error,[12] we estimated an equation based on a Boolean outcome-dependent variable known in the literature as a Positive and Statistically Significant (PSS) variable. The PSS takes the value 1 if the estimated effect is positive and significant below a 10 (or 5) percent threshold and is 0 otherwise. As Table 5 shows, almost no estimate has the wrong sign, so there is no need to estimate the ordered probit model used in Card et al. (2017).

NEGATIVE AND SIGNIFICANT AT THE 5% LEVEL	NOT SIGNIFICANT AT THE 5% LEVEL	POSITIVE AND SIGNIFICANT AT THE 5% LEVEL
23	431	214

Table 5 Distribution of the coefficients of the impact on the employment status of participants or their earnings according to its sign and statistical significance.

Mixed Effects Model

We complemented these results by estimating a mixed-effects model whenever the article reported not only the ITT, average treatment effects (ATE) or LATE of the intervention, but also its standard error and the mean average value of the control group.[13] A mixed effects model can be expressed in the following form:

$$y_j = \beta_0 + \beta X + \epsilon_k + \gamma_k \tag{4}$$

where y_j is the observed ITT, ATE or LATE, β_0 is the intercept, β is a vector of coefficients that measure the impact of the variables that capture the design space (X), ϵ_k is a sampling error and γ_k is a vector of random effects.

Importantly, while this model suffers from a similar loss of observations as would be the case if we had estimated equation (3) directly, we chose to report the results because there is an inherent heterogeneity in the distribution from which we drew the studies and, as a result, a great residual variation after controlling for the factors that can be coded, as is indeed found in the data. This test can be regarded as following the same motivation that led previous ALMP meta-analyses to avoid

12 While this drawback is less severe in our work than in past studies, in 33% of the point estimates the report the *p* value (or a traditional threshold) comes with no information on their standard errors.

13 As noted, this effectively reduces the number of point estimates by about one third, from 668 to 404.

using the standard deviation as an indicator of accuracy because it also captures variations that cannot be explained simply by sampling error (Konstantopoulos, 2006; Card et al., 2017; Escudero et al., 2017).

META-REGRESSION ANALYSIS: RESULTS

Table 6 shows the results for the full sample, ignoring program components and design variables, except for the inclusion of a control for private sector participation. We started from this specification because, even though the interventions are homogeneous in the sense that all of them are active labor market policies or interventions, the components of the programs are usually program-specific, and they can mask the differential impact of a program type.

First, we can tell that splitting the classification into different types of programs captures part of the effectiveness variation. Job search assistance services tend to be less effective than the other programs, while vocational trainings and support to micro-entrepreneurs and independent workers are less effective than wage subsidies. This difference is like the variation shown in the median costs of the programs when we had data about them (see Figure 1).

Second, we analyzed whether the documented effects for any particular population subgroups were significantly higher or lower than the overall average. Our results document that on average the programs have less effectiveness in the specific subgroup of people aged 24 or older than in the overall population, while there seems to be no significant difference in effectiveness between any particular gender or educational level of the participant. Context variables hint that, even though these policies often have the objective of counteracting the impact of the business cycle on the labor market, the GDP growth in the year of the implementation of the program correlates with a greater effectiveness of the program (which signals that ALMP may have a pro-cyclical component), while the income level of the country and the unemployment rate are statistically insignificant at any of the conventional levels.

Table 6 Coefficients, p-value and confidence interval (at the 95 percent level) of a linear probability model with a Positive and Significant Sign (PSS) as a dependent variable.

	COEFFICIENT (p-VALUE)	CONFIDENCE INTERVAL (95%)	
		LOWER BOUND	UPPER BOUND
Type of program			
Vocational trainings	0.2 (0.07)	0.05	0.34
Supports to micro-entrepreneurs and independent workers	0.2 (0.09)	0.03	0.38
Wage subsidy	0.58 (0.13)	0.32	0.84
Population			
Gender (omitted = pooled)			
Women	0.01 (0.06)	–0.1	0.13
Men	–0.06 (0.06)	–0.17	0.05
Age (omitted = pooled)			
Aged 24 or lower	–0.041 (0.09)	–0.21	0.13
Aged above 24	–0.19 (0.07)	–0.33	–0.05
Education (omitted = pooled or no information)			
Incomplete high school or lower	–0.26 (0.07)	–0.4	–0.13
Complete high school or higher	–0.13 (0.08)	–0.30	0.03
Context			
Lower or lower-middle income country	0.12 (0.08)	–0.04	0.28
GDP growth in the year of implementation	0.02 (0.01)	0	0.04
Unemployment in the year of implementation	–0.01 (0.01)	–0.02	0

(Contd.)

	COEFFICIENT (p-VALUE)	CONFIDENCE INTERVAL (95%)	
		LOWER BOUND	UPPER BOUND
Other controls			
Impact measured more than a year after program completion	−0.01 (0.07)	−0.14	0.12
Field experiment	−0.02 (0.06)	−0.15	0.11
Non-public sector participation	0 (0.07)	−0.13	0.13
Number of observations	668		
R squared	0.16		

In our sample of evaluated interventions, it was critical to show different models as some of the variables are dependent on a specific type of program. Table 7 shows coefficients and their significance at conventional level across eight different models. They are the combination of two cut-offs for the PSS binary variable (5 percent and 10 percent) and four subsamples. Several insights arise from this result.

First, it was observed that variables that attempt to capture the context in which the programs were implemented tended to be significant across every subsample. This was especially true among supports to micro-entrepreneurs and independent workers and vocational training programs. GDP per capita growth in the year of the experiment was significant at the 90 percent level or more in three out of four subsamples, and it was always statistically significant whenever employment services programs were excluded. The unemployment rate at the year of the experiment coefficient was always negative, although it was only significantly different from zero in vocational trainings and the specifications with all programs except employment services.

Second, we found little evidence of a positive impact of the nonpublic sector in any of the type of participations we considered. Nonpublic sector financing was positive and statistically different from zero at the 90 percent level whenever we included wage subsidies in the sample. Once we excluded wage subsidies, nonpublic financing seemed to have no effect.

Third, we found some evidence of this kind of policy being more effective at improving formal employment levels of the participants rather than boosting earnings or other types of employment, in line with findings reported in other meta-analyses. However, we think that this could be at least partly due to the null hypothesis testing when working with administrative data, which is always larger than survey data.

Finally, we found that the individualized coaching and follow up of the participants, monetary transfers for the participants and an explicit focus of the program in a particular sector or economic activity are all associated with a higher chance of finding an effective response either in earning or employment outcomes. The explicit activity targeting, and the monetary transfers are also statistically significant for the vocational training subsample.

Table 8 shows the same output, but this time estimates a mixed-effect model detailed in equation (4). We show the results of estimating the variance of the distribution of effect sizes both through Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML). We found some differences in the contribution of the variables of our design space to the effect size of the interventions we study.[14]

First, we saw that on average the greatest impacts were in formal employment and, to a lesser extent, in earnings relative to only employment. The differential impact on formal employment held in every subsample of programs. As in the PSS specification, we did not see any evidence that pointed to impacts that were measured after one year of completion being of less magnitude.

14 While there are differences in the results reported in the tables, which reflect in part the fact that the ME model loses some observations, there are no reversions in cases where results in both tables are statistically significant.

Second, we found an almost negligible association between the type of non-public sector participation, gender or age of the participants, and the size of the effects. We did find greater impacts on the relatively more educated participants in almost all subsamples.

Third, we found that low or lower-middle income country interventions are linked to a greater impact on average. Moreover, and in line with the PSS specification, GDP growth at the year of the intervention was positively associated with the size of the impacts, while unemployment was negatively associated.

Finally, regarding the other controls, we found a partial reversal of what was observed in the PSS specification regarding specific-activity training, with some negative associations in the magnitude of the effect in two specific subsamples (this may be as a result of changing the sample between both models). Monetary incentives to the participants were mostly positively associated with the size of the effect in most subsamples, although it was only statistically significant in the ML estimation of independent and vocational training subsample. Individualized mentoring or follow-up of the participants was also mostly strongly positively associated with respect to the magnitude of the impacts. Longer programs were associated with stronger outcomes, especially in the vocational training subsample.

Table 7 Coefficients and p values of the null hypothesis testing of a linear probability model with a Positive and Significant Sign (PSS) as a dependent variable. Each column shows the p value used as cut off for the PSS binary variable. * Significant at the 10% level, ** Significant at the 5% level and *** Significant at the 1% level.

	POOLED PROGRAMS		ALL PROGRAMS EXCEPT EMPLOYMENT SERVICES		INDEPENDENT AND VOCATIONAL TRAINING		VOCATIONAL TRAINING	
	PSS (5%)	PSS (10%)	PSS (5%)	PSS (10%)	PSS (5%)	PSS (10%)	PSS (5%)	PSS (10%)
Intercept	0.09	0.2	0.11	0.21	0.09	0.12	0.13	0.1
Category of outcome (omitted = employment)								
Formal employment	0.13	0.18**	0.08	0.14	–0.06	0.08	0.03	0.18*
Earnings	0.03	0.02	0.03	0.04	0.01	0.03	0.07	0.09*
Impact measured more than a year after program completion	0.13	0.18**	0.08	0.14	–0.06	0.08	0.03	0.18*
Nonpublic sector participation								
Designs	–0.12	–0.11	–0.16	–0.16	0.02	–0.02	–0.02	–0.15
Implements	0.08	0.07	0.07	0.07	0.02	0.02	–0.1	–0.04
Finances	0.27**	0.3***	0.3**	0.33***	0.05	0.16	–0.07	0.15
Population properties								
Gender (omitted = pooled gender)								
Women	0	0.05	0	0.04	–0.04	0.02	–0.05	0.04
Men	–0.09	–0.08	–0.09	–0.07	–0.11	–0.1	–0.1	–0.11
Age (omitted = pooled)								
Aged 24 or lower	–0.08	–0.11	–0.16**	–0.18***	–0.07	–0.08	0.05	–0.05
Aged above 24	–0.19**	–0.27***	–0.2***	–0.27***	–0.13*	–0.21***	–0.04	–0.13
Education (omitted = pooled or not available)								
High school dropout or lower	–0.12	–0.05	–0.08	0.01	–0.07	0.02	–0.06	0.03
Complete high school or higher	–0.26***	–0.33***	–0.1	–0.14	0.15	0.07	0.24	0.16
Context								
Lower or lower-middle income country	0.07	0.04	0.01	–0.04	0.01	–0.05	0.15	0.02
GDP growth in the year of implementation	0.02	0.02	0.02*	0.02*	0.04***	0.04***	0.03**	0.04***
Unemployment in the year of implementation	–0.01	0	–0.02**	–0.02**	–0.02	–0.01	–0.03**	–0.02

(Contd.)

	POOLED PROGRAMS		ALL PROGRAMS EXCEPT EMPLOYMENT SERVICES		INDEPENDENT AND VOCATIONAL TRAINING		VOCATIONAL TRAINING	
	PSS (5%)	PSS (10%)	PSS (5%)	PSS (10%)	PSS (5%)	PSS (10%)	PSS (5%)	PSS (10%)
Other controls								
Field experiment	0.02	0.01	0.07	0.06	0.04	0.06	0.1	0.12
Training aims to a specific industry	0.04	0.1	−0.09	−0.02	−0.12	−0.03	0.18	0.29**
Soft skills module	0.07	0.03	0.02	−0.02	−0.09	−0.12	−0.06	−0.08
Individualized mentoring or follow up	0.05	0.09	0.16*	0.19**	0.19**	0.25***	0.07	0.25**
Monetary incentive for the participants	0.04	0.07	0.07	0.09	0.11	0.15	0.18*	0.26***
Program length (omitted = not available)								
A year or less than a year	0.13	0.06	0.2**	0.12	0.17*	0.06	0.09	−0.12
More than a year	0.21**	0.18*	0.29***	0.25**	0.24**	0.2*	0.33***	0.21**

	POOLED PROGRAMS		ALL PROGRAMS EXCEPT EMPLOYMENT SERVICES		INDEPENDENT AND VOCATIONAL TRAINING		VOCATIONAL TRAINING	
	REML	ML	REML	ML	REML	ML	REML	ML
Category of outcome (omitted = employment)								
Formal employment	0.08*	0.05	0.1**	0.05	0.08**	0.07	0.06**	0.06**
Earnings	0.03	0.01	0.05*	0	0.05**	−0.01	0.02**	0.02
Impact measured more than a year after program completion	−0.01	−0.01	−0.05	−0.03	0.01	−0.02	−0.01	−0.01
Nonpublic sector participation								
Designs	0.03	0.02	−0.11	0.01	−0.07	0.04	0.08**	0.08
Implements	−0.01	0	0.05	0.02	0.01	−0.08	−0.12	−0.12
Finances	0.1	0.06	0.07	0.06	0.12	0.08	−0.29**	−0.29**
Population properties								
Gender (omitted = pooled gender)								
Women	0.03	0.01	0.04	0.02	0.01	0.01	0.01	0.01
Men	−0.04	0	−0.02	0	−0.02	0	0	0
Age (omitted = pooled)								
Aged 24 or lower	−0.05	0.01	−0.04	0.01	−0.03	0.01	0.06*	0.05
Aged above 24	−0.02	−0.01	0.03	−0.01	0.03	−0.01	0.08**	0.08
Education (omitted = pooled or not available)								
High school dropout or lower	0.08	−0.16**	0.14*	−0.16	−0.01	−0.15	−0.01	−0.01
Complete high school or higher	0.01	−0.08**	0.47**	0.12	0.24*	0.13	0.19*	0.19
Context								
Lower or lower-middle income country	0.12**	0.11**	0.05	0.09	0.09**	0.07*	0.09**	0.09
GDP growth in the year of implementation	0.01*	0	0.02**	0	0.02**	0.01	0	0
Unemployment in the year of implementation	0.01*	0.01	−0.02	0	−0.02*	−0.01	−0.03**	−0.03*

(Contd.)

	POOLED PROGRAMS		ALL PROGRAMS EXCEPT EMPLOYMENT SERVICES		INDEPENDENT AND VOCATIONAL TRAINING		VOCATIONAL TRAINING	
	REML	ML	REML	ML	REML	ML	REML	ML
Other controls								
Field experiment	−0.11**	−0.03	−0.09	−0.04	−0.03	−0.01	0.02*	0.02
Training aims to a specific industry	0.11	−0.11	−0.05	−0.12	−0.22**	−0.16	−0.09*	−0.09
Soft skills module	−0.03	0.08*	−0.09*	0.06	0.05	0.07	−0.03	−0.03
Individualized mentoring or follow up	−0.02	0.1**	0.17**	0.16**	0.19**	0.18**	−0.05	−0.05
Monetary incentive for the participants	−0.13**	0.02	−0.07	0.04	0.04	0.18**	0.06	0.06
Program length (omitted = not available)								
A year or less than a year	−0.02	−0.01	−0.13	−0.03	−0.01	0.04	0.2**	0.2**
More than a year	0.01	−0.06	0.09	−0.06	0.09	0.05	0.34**	0.34**

Table 8 Coefficients and p values of the null hypothesis testing of a mixed-effect model with the estimated effect (relative to the control group) as a dependent variable. Each column shows the method used to estimate the variance of the distribution of effects. ML stands for Maximum Likelihood and REML for Restricted Maximum Likelihood. *Significant at the 10% level, **Significant at the 5% level and ***Significant at the 1% level.

5. FINAL REMARKS

In this work, we tried to extract lessons from a systematic review of more than 102 ALMPs evaluated from experimental methodologies (RCTs). To be able to make a granular and standardized description of the ALMPs, we generated a taxonomy and a set of standardized variables that tracked (i) specific content of the policies, (ii) the characteristics of their implementation and (iii) the context and the target population. This harmonized design space allowed us to capture part of the multidimensional nature of these policies and then to extract more granular and specific conclusions of their overall effectiveness in a quantitative meta-analysis.

We found that impacts on employment and earnings outcomes are moderately positive on average. The median impact on the participants' employment outcomes of the program relative to the control group ranges from approximately 11 percent for wage subsidies and independent workers' support, to 2 percent for the employment services. Vocational training lies in the center of this range, reporting a median impact of 6.7 percent The median impact on the participants' earnings outcomes were higher for wage subsidies and independent workers' support, reaching almost a boost of 17 percent in the earnings outcomes of the reported coefficients, and for vocational trainings, which have a median relative impact of 7.7 percent. On the other hand, employment services reported null effects on earnings.

Although only 51 interventions reported the costs of providing the program per participant, we found that wage subsidies, vocational trainings, and independent workers' support do not differ significantly in the median cost per participant, which is approximately 1500 and 1700 USD (2010 PPP). Employment services are inexpensive policies whose median cost is only 277 USD. However, two caveats should be stressed: (i) there are only five observations of the costs of wage subsidies programs; and (ii) the dispersion in the costs of independent workers' support is high relative to the rest of the programs.

When we focused on the quantitative meta-analysis our main finding was that context matters: GDP per capita growth is positively correlated and the unemployment rate is negatively related to the probability of a program reporting positive and statistically significant (PSS) and to greater impacts. In this regard, we found evidence close to Escudero et al. (2017). Moreover, the individual following of the participants is correlated with better outcomes once employment services are removed from the sample (in line with Kluve et al., 2019).

In the vocational training subsample, in which our database is denser, we found more interesting results. First, context keeps playing an important role, with a tendency of these programs to have a pro-cyclical component, where its effectiveness positively correlates with the business cycle.

Second, as expected and in line with other meta-analyses (Card et al., 2017) longer programs tend to be more effective. Third, monetary incentives for the participants to cover for the opportunity costs of taking the program, or maybe just nudging for the participation becomes a statistically significant coefficient (in line with Kluve et al., 2019). Finally, sector-oriented vocational programs are also associated with a greater probability of success in the PSS regression, but not so in the mixed-effect model.

As we pointed out, without a clearly defined design space to evaluate the characteristics of the design, implementation, context and target population of the policy, learning from the identification of empirical regularities in the determinants of the success is rather limited. The adoption of a taxonomy validated by the evaluating community would help to consolidate the findings from different empirical evaluations in a granular way, and to summarize them in terms of universally comparable variables.

We hope that the proposed design space in this work serves as a preliminary version of a validated protocol for a systematic description of ALMPs. Ideally, each academic publication could specify explicitly and in tabulated form each of the dimensions of the design space for the policies under consideration. This would allow truly informative reviews and aggregate analyses to provide powerful insights for policymakers. Although the accomplishment of empirical meta-analysis has yet to be realized, this systematized description would allow specific elements of the design of each kind of policy to be isolated, identifying empirical regularities that are truly useful for the policymaker.

However, this taxonomy is only a parsimonious version of the many relevant decisions in the design and implementation of an ALMP. The design space suggested in the present work is a first approximation that does not contemplate some crucial aspects of the policies evaluated as these are not systematically described in the compiled studies. Indeed, the information available in academic publications often leaves several aspects of interest undocumented. For example, they can describe with rigor the content of an ALMP (the type of training, its duration, its modality, the disciplines or productive sectors involved, etc.), but say little about its implementation (was there a competitive selection process for training providers, or incentives that linked their remuneration to performance, or a monitoring process and evaluation of their activity?).

While we believe that the boom of experimental evaluations has the potential to provide informative evidence for the design of effective policies, the impact of this empirical research will require a coordinated effort to facilitate a standardized collection of information. This paper, which compared studies based on a homogenous approach that lends itself more easily to the construction of the coefficients used as impact measures, is another step in the continuous effort to extract systematic lessons from policy experience, and as such will be hopefully updated in the future as the dataset is enriched with new interventions and evaluations.

APPENDIX – LESSONS FROM PAST META-ANALYSES

Along with the publication of new impact evaluations of ALMPs came a series of meta-analyses that tried to infer what works in this kind of intervention, controlling for important variations in design, context and implementation.

Heckman et al. (1999), a seminal paper within this group, focused on job training, public works, wage subsidies, and job search assistance services programs in the United States.[15] They found that programs in the U.S. tended to have a modest but overall persistent positive impact on the labor income of the participants, especially in the case of adult women. Importantly, they concluded that, under certain scenarios, these interventions seemed to be 'remarkably costeffective'. They also noted that the programs appeared to be effective in raising the earnings of disadvantaged adult males, but ineffective in delivering the same outcomes when it came to disadvantaged youths. They attributed this heterogeneity, in part, to skill differences across groups, suggesting that ALMPs performed comparatively better with skilled participants.

15 As the authors note, the country bias was not as a result of a preference for ALMPs in the US, but rather thanks to a preference for having these policies evaluated, relative to other developed countries.

Since this seminal contribution several meta-analyses were published (See Table A1 below). We identify three patterns that are present in the existing literature:

1. There has been an important effort not only to estimate the effectiveness of the programs but also to identify features in the design, context and implementation of the interventions associated with positive outcomes. Thus, some meta-analyses focus exclusively or mostly on Latin American and Caribbean countries (Escudero et al., 2017; Kluve 2016) or on low-income countries (Cho and Honorati, 2013; Grimm and Pauffhausen, 2015; McKenzie, 2017), while Kluve et al. (2019) look exclusively into programs that target young workers. Even when the universe of interventions is not restricted, the data are usually subset to analysis the heterogeneity in the programs' effectiveness (Card et al., 2017; Kluve 2016).
2. While all the meta-analyses discuss the quality of the data and the methodology of the impact evaluations they work with, RCT impact evaluations usually represent a minor share of all the evaluations included in the analyses, making up as little as 30 percent in some cases (Card et al., 2017). While the debate of whether to mix experimental impact evaluations with other types of evaluation design is present and discussed at some length in the texts, and even controlled for in the regressions, the main conclusions about what works are largely drawn from non-RCT impact evaluations.
3. There is some basic agreement in the main findings of these meta-analyses. When comparing the effectiveness of the programs between public and nonpublic implementations, authors conclude that ALMPs that are not implemented by the public sector are associated with a better performance (Cho ands Honorati, 2013; McKenzie, 2017; Kluve et al., 2019). Both Kluve (2016) and Card et al. (2017) conclude that public-sector employment programs have negligible impact. Also, there is agreement about the magnitude of the impacts over different time horizons: effects tend to be larger in the medium and long run[16] (Kluve, 2016; Card et al., 2017; Kluve et al., 2019). Another common finding is that programs that target long-term unemployment have larger impacts. There is mixed evidence for a sign of the relationship between growth and performance; in particular, in contrast with our findings, one meta-analysis (Kluve, 2016), shows that programs are *more* effective in periods of slow growth and higher unemployment, although this finding restricts itself to a Latin America and Caribbean sample.

PAPER	INTERVENTIONS UNDER STUDY	FOCUS ON	NUMBER OF STUDIES	NUMBER OF PROGRAMS EVALUATION	NUMBER OF ESTIMATES	IMPACT EVALUATION METHODOLOGY	MAIN FINDINGS
Cho and Honorati (2013)	Interventions that aim at promoting potential or current entrepreneurs. It includes Active Labor Market Policies (ALMPs) designed to enhanced technical, vocational or financial skills for self-employment.	Developing countries	37	Not available	1116	Experimental or quasi-experimental	They don't find statistically significant difference among types of programs, although when interacting training with counseling the magnitude of the effects tends to be higher. Programs impacts estimated for youth and the urban population use to be positive and significant. NGOs are associated with better performance.
Grimm and Pauffhausen (2015)	Access to finance, entrepreneurship training, business development environment.	Low- and middle-income countries	53	Not available	116	Experimental or quasi-experimental	Finance interventions had lower employment effects than training ones. Interventions targeting

(Contd.)

16 We do not find this effect in our model and data. This difference could arise from using only RCT impact evaluations, in which the problem of mixing short-run estimates after the end of the program with some observations when the program was still running is reduced.

PAPER	INTERVENTIONS UNDER STUDY	FOCUS ON	NUMBER OF STUDIES	NUMBER OF PROGRAMS EVALUATION	NUMBER OF ESTIMATES	IMPACT EVALUATION METHODOLOGY	MAIN FINDINGS
	services, wage subsidies, and improvements to the business						small enterprises are more successful than those that target micro-enterprises. Combined interventions did not systematically lead to larger effects, but the combination of finance and training work better compared to when they are isolated.
Kluve and Rani (2016)	Job search assistance, labor market training, private sector employment incentives and public sector employment	No restrictions	207-LAC sample: 44	526	857- LAC sample: 152	Experimental or quasi-experimental	Effect sizes tend to increase from short to medium-run and is slightly negative between medium and long-run. Public sector employment has negligible or negative impacts. Programs targeting only young and older participants have smaller impacts, relative to those of mixed age groups. Long-term unemployed targeted programs have larger impacts. In periods of slow-growth and high unemployment there are larger impacts; but in LAC GDP growth is positively correlated with effectiveness.
Escudero et al (2017)	Active labor market policies (ALMPs) – i.e. training programs, public works, employment subsidies, self-employment and microenterprise creation programs, and labor market intermediation services	Latin America and the Caribbean	51	53	296	Experimental or quasi-experimental	Interventions with short duration are less likely to produce positive impact compared to longer ones. GDP growth is positively correlated with program effectiveness. Females and youth are more likely to benefit from the programs. Training programs are more successful, mostly impacting over employment formality.
McKenzie (2017)	Labor market policies that have provided vocational training, wage subsidies, job search assistance, and assistance moving	Developing countries	24	22		--	Traditional ALMPs have had at most modest impacts on employment and earnings in most cases. Training is more effective when given by private providers. Subsidies may be useful for temporary employment creation.
Card et al. (2017)	Classroom or on-the-job training; job search assistance, monitoring, or sanctions for failing to search;	Latin America and the Caribbean	207	526	857	Experimental or quasi-experimental (30% of RCTs)	Average impacts are "more positive" 2–3 years after competition of the program. Programs that emphasize in human capital accumulation have larger average gains.

(Contd.)

PAPER	INTERVENTIONS UNDER STUDY	FOCUS ON	NUMBER OF STUDIES	NUMBER OF PROGRAMS EVALUATION	NUMBER OF ESTIMATES	IMPACT EVALUATION METHODOLOGY	MAIN FINDINGS
	subsidized private sector employment; subsidized public sector employment						There are larger impacts for females and participants who enter from long-run unemployment. ALMPs are more likely to show positive impacts in a recession. Public sector employment has negligible impacts.
Kluve et al. (2019)	Youth-targeted active labor market interventions: training and skills development, entrepreneurship promotion, employment services, and subsidized employment interventions	No restrictions	113	87	3105	Experimental or quasi-experimental (66% of RCTs)	There is no evidence that some programs outperform others but those which integrate multiple services are more successful. Programs in middle- and low-income countries and are more successful. The intervention type is less important than design and delivery. Profiling of beneficiaries, individualized follow-up systems and incentives for services providers (only in high-income countries) matter. In lower income settings, implementation by non-public actors reports larger effect sizes than joint ones. Impacts are of larger magnitude in the long-term.

Table A1 Summary of previous meta-analyses of ALMPs. LAC means Latin America and the Caribbean.

DIMENSION	VARIABLE	VARIABLE TYPE	DESCRIPTION
ALMP Type	ALMP Type	Categorical	Indicates the ALMP type corresponding to each program among the following categories: i) vocational training, ii) wage subsidies or public works programs, iii) supports to micro-entrepreneurs and independent workers and iv) assistances in the job process
Content components	Training in technical skills	Dummy	Indicates if the program has any type of academic or “in-class” training that sought to provide the participant with the technical skills required in their work occupation
Content components	Training in soft skills	Dummy	Indicates if the program has any type of instruction of coaching that sought to provide the participant with “soft skills” that are potentially relevant to their professional performance
Content components	Work experience, internship or on-the-job training	Dummy	Indicates if the program has any type of instruction in real workplace settings whether through formal work experiences (e.g., internships) or simulated trainings
Content components	Individualized mentoring or participants tracking	Dummy	Indicates if the program has any type of structured guidance to the participants with assigned coaches/mentors or a system of individualized follow-up with regular meetings with the participant
Content components	Assistance services in the job-search process	Dummy	Indicates if the program offers any help in the job-search process like mentoring for interviews, help in designing a CV or digital profile, organization of job fairs or assistance in the networking with potential employers (among other possibilities)

(Contd.)

DIMENSION	VARIABLE	VARIABLE TYPE	DESCRIPTION
Content components	Monetary transfer or expenses financing	Dummy	Indicates if the program includes any monetary assistance to the participants (either in the form of incentives or financing of daily expenses)
Content components	Voucher system	Dummy	Indicates if the program offers a voucher (or any similar mechanism) with which the participant can freely choose their training provider (public or private) instead of discretionally assigning him a provider selected by the policymaker
Content components	Asset transference	Dummy	Indicates if the program offers any transfer of productive assets for participants to use in their work activities
Content components	Loan	Dummy	Indicates if the program offers any loan for participants to incorporate assets to use in their work activities
Implementation features	Duration	Continuous	Indicates the duration of the program measured in months
Implementation features	Participation of the non-public sector	Dummy	Indicates the participation in the program of any institution of the non-public sector (e.g. private firms, NGOs, multilateral organizations)
Implementation features	Participation of private firms	Dummy	Indicates the participation in the program of private firms
Implementation features	Participation of NGOs	Dummy	Indicates the participation in the program of NGOs
Implementation features	Participation of multilateral organizations	Dummy	Indicates the participation in the program of multilateral organizations
Implementation features	Participation of the non-public sector in the design of the program	Dummy	Indicates the active involvement of the non-public sector in the design of the program
Implementation features	Participation of the non-public sector in the financing of the program	Dummy	Indicates the active involvement of the non-public sector in the financing of the program
Implementation features	Participation of the non-public sector providing the main content of the policy	Dummy	Indicates the active involvement of the non-public sector in the provision and implementation of the program
Implementation features	Field experiment and/or pilot	Dummy	Indicates if the program is an actual public policy implemented by a national or local government or a field experiment (or pilot test) on a smaller scale or controlled setting
Implementation features	Orientation to specific economic sectors or occupations	Dummy	Indicates if the policy has an explicit orientation to a particular economic sector (e.g. manufacturing, tourism, IT, etc.) or specific occupations (e.g. IT programmers)
Implementation costs	Implementation costs	Continuous	Indicates the average cost per participant of the intervention, in 2010 PPP dollars
Context and target population	Average, minimum and maximum age of the participants	Continuous	Indicates the average, minimum and maximum age of the participants in the program
Context and target population	Proportion of participants with a high-school degree	Continuous	Indicates the proportion of participants in the program that earned a high-school degree
Context and target population	Proportion of participants with a university degree	Continuous	Indicates the proportion of participants in the program that earned a university degree
Context and target population	Average years of study of the participants	Continuous	Indicates the average years of study of the participants in the program
Context and target population	Proportion of unemployed participants at the beginning of the program	Continuous	Indicates the proportion of participants in the program that were unemployed at the beginning of the program
Context and target population	Gender distribution of the participants	Continuous	Indicates the gender distribution of the participants in the program

Table A2 Description of variables included in the ALMP design space.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Eduardo Levy Yeyati orcid.org/0000-0001-7518-1422
School of Government at Universidad Torcuato Di Tella, AR

Martín Montané
Research associate with the Center for Evidence-Based Policy (CEPE) at Universidad Torcuato Di Tella, AR

Luca Sartorio
Research associate with the Center for Evidence-Based Policy (CEPE) at Universidad Torcuato Di Tella, AR

Alfonso Gauna
University of Chicago, US

REFERENCES

Andrews, M., L. Pritchett, and **M. Woolcock.** *Building state capability: Evidence, analysis, action*. Oxford University Press, 2017. DOI: https://doi.org/10.1093/acprof:oso/9780198747482.001.0001

Card, D., J. Kluve, and **A. Weber.** "Active labor market policy evaluations: A meta-analysis." *The Economic Journal*, 120, no. 548 (2010): F452–F477. DOI: https://doi.org/10.1111/j.1468-0297.2010.02387.x

Card, D., J. Kluve, and **A. Weber.** "What works? A meta-analysis of recent active labor market program evaluations." *Journal of the European Economic Association* 16 no. 3 (2017): 894–931. DOI: https://doi.org/10.1093/jeea/jvx028

Cho, Y., and **M. Honorati.** *Entrepreneurship programs in developing countries: A meta regression analysis*. The World Bank, 2013. DOI: https://doi.org/10.1596/1813-9450-6402

Escudero, V., J. Kluve, E. López Mourelo, and **C. Pignatti.** "Active labour market programmes in Latin America and the Caribbean: Evidence from a meta-analysis." *The Journal of Development Studies* (2017): 1–18. DOI: https://doi.org/10.2139/ssrn.3045735

Grimm, M., and **A. L. Paffhausen.** "Do interventions targeted at microentrepreneurs and small and medium-sized firms create jobs? A systematic review of the evidence for low- and middle-income countries." *Labour Economics* 32 (2015): 67–85. DOI: https://doi.org/10.1016/j.labeco.2015.01.003

Heckman, J. J., R. J. LaLonde, and **J. A. Smith.** "The economics and econometrics of active labor market programs." In *Handbook of labor economics* 3, (1999): 1865–2097. Elsevier. DOI: https://doi.org/10.1016/S1573-4463(99)03012-6

Kluve, J., and **U. Rani.** *A review of the effectiveness of Active Labour Market Programmes with a focus on Latin America and the Caribbean*. Geneva: International Labour Organization, 2016. https://www.ilo.org/media/434176/download

Kluve, J., S. Puerto, D. Robalino, et al. "Do youth employment programs improve labor market outcomes? A quantitative review." *World Development*, 114 (2019): 237–253. DOI: https://doi.org/10.1016/j.worlddev.2018.10.004

Konstantopoulos, S. "Fixed and mixed effects models in meta-analysis." *IZA Discussion Papers*, No. 2198 (2006). DOI: https://doi.org/10.2139/ssrn.919993

McKenzie, D. "How effective are active labor market policies in developing countries? a critical review of recent evidence." *The World Bank Research Observer*, 32, no. 2 (2017): 127–154. DOI: https://doi.org/10.1093/wbro/lkx001

Pritchett, L., S. Samji, and **J. S. Hammer.** "It's all about MeE: Using Structured Experiential Learning ('e') to crawl the design space." *Center for Global Development Working Paper*, (2013): 322. DOI: https://doi.org/10.2139/ssrn.2248785

TO CITE THIS ARTICLE:
Levy Yeyati, Eduardo, Martín Montané, Luca Sartorio, and Alfonso Gauna. 2025. "What Works for Active Labor Market Policies? A Meta-Analysis of RCT Findings." *Economía LACEA Journal* 24(1): 81–104. DOI: https://doi.org/10.31389/eco.450

Submitted: 04 May 2024
Accepted: 08 November 2024
Published: 19 March 2024

Economía LACEA Journal is a peer-reviewed open access journal published by LSE Press.