The effectiveness of formalisation policies in Latin America

Matias Golman* Verónica Escudero† Elva López Mourelo‡

This version: December 29, 2025

Click here for the most recent version

Abstract

This paper presents the first comprehensive meta-analysis of formalisation policies in Latin America, analysing 79 impact evaluations with 527 estimates from interventions implemented between 1990 and 2020. We calculate standardised mean differences for 444 estimates and conduct thorough publication bias analysis using FAT-PET-PEESE methods. Unlike previous meta-analyses, we examine interventions targeting both worker and firm formalisation, include Spanish-language studies and unpublished government reports, and provide the first systematic assessment of effect magnitudes for the region. Our findings show that formalisation policies predominantly produce positive effects, with adverse impacts being rare (5.1% of estimates). However, systematic publication bias, when corrected, substantially reduces effect sizes. For formality outcomes, active labour market policies show consistent positive effects (0.238 standard deviations, $p < 0.05$), as do information campaigns (0.226 standard deviations, $p < 0.10$) and labour inspections (0.320 standard deviations, $p < 0.10$). For earnings outcomes, social protection programmes with activation components demonstrate the largest effects (1.258 standard deviations, $p < 0.01$), though negligible formality impact (0.033 standard deviations, not significant). This divergence suggests these programmes primarily affect income rather than formal employment status. Programme benefits strengthen over time, with long-term effects exceeding short-term impacts by approximately 0.10 standard deviations for formality and 0.32 standard deviations for earnings. Multi-component interventions outperform single-component policies by 0.07 standard deviations. These findings underscore the need for realistic expectations while supporting evidence-based, multifaceted approaches to formalisation.

Keywords: formalisation policies, Latin America, meta-analysis, publication bias, active labour market programmes, social protection

JEL Codes: J08, J21, J46, O17, C83

*University of Nottingham. Email: matias.golman@gmail.com.

†International Labour Organization. Email: escudero@ilo.org.

‡International Labour Organization. Email: lopezmourelo@ilo.org.

Contents

1 Introduction

Informality remains a defining characteristic of labour markets across Latin America, with approximately 140 of the region's 263 million workers operating in informal employment as of 2016 (Salazar-Xirinachs and Chacaltana, 2018). Following the standards of the International Labour Organisation, informal employment refers to work that is not covered by formal systems such as labour law, social security, and tax regulations (ILO, 2024). This represents a significant economic and social challenge. Informal workers lack access to social protection, face precarious working conditions, and typically earn lower and more irregular wages than their formal counterparts. Informal firms face substantial barriers to growth, including limited access to credit, restricted participation in government procurement, and exclusion from formal supply chains. At the aggregate level, informality undermines equity, efficiency, tax collection, social security coverage, productivity, and economic growth (Bertranou, 2025).

The persistence of high informality rates has prompted governments in the region to implement diverse policy interventions aimed at promoting transitions to formality. These range from active labour market policies (training courses, internships, and employment subsidies or intermediation) to institutional reforms, including simplified business registration, reduced payroll taxes, enhanced labour inspection, and information campaigns on formalisation benefits (Escudero et al., 2019; Jessen and Kluve, 2021). This diversity reflects both the multifaceted nature of informality and the competing theoretical frameworks that guide policy design. Despite substantial investment in these interventions, systematic evidence on their effectiveness remains limited. Individual impact evaluations exist, but no comprehensive effort has been made to synthesise these findings to identify which approaches work, for whom, and under what conditions.

Previous meta-analyses have examined related questions but with limitations. Escudero et al. (2019) analysed active labour market policies in Latin America, finding positive effects on formal employment, particularly for women and youth, but focused primarily on employment outcomes rather than formalisation itself. Jessen and Kluve (2021) examined formalisation interventions globally, identifying 170 impact estimates from 38 studies, but their analysis was not Latin America-specific and omitted Spanish-language studies and unpublished reports. Floridi et al. (2020) conducted a meta-analysis of formalisation interventions, but restricted their scope to firm outcomes and did not focus specifically on Latin America. In addition, none of these studies calculated standardised mean differences for the impacts of formalisation policy in the region, limiting the comparability of effect sizes in different outcomes and studies.

This paper addresses these gaps by conducting the first comprehensive meta-analysis of formalisation policies specifically for Latin America. We built an extensive database of 79 impact evaluations covering interventions implemented between 1990 and 2020, with

527 impact estimates. Our analysis makes several advances over previous work. First, we calculate standardised mean differences (Hedges' g) for 444 estimates, providing the first systematic assessment of effect magnitudes for formalisation policies in the region. Second, we examine a comprehensive range of policy interventions targeting both worker and firm formalisation, going beyond the focus on either active labour market policies or firm-level interventions in isolation. Third, we include evaluations published in Spanish alongside English-language studies, and incorporate unpublished reports often excluded from global meta-analyses. We address potential quality concerns by applying strict inclusion criteria requiring experimental or quasi-experimental methods. Fourth, we conduct a thorough publication bias analysis using FAT-PET-PEESE correction methods. We examine effectiveness across three key outcomes: probability of being formal (for employees or firms), monthly earnings, and hours worked. We employ both random-effects and fixed-effects meta-analytical models to account for heterogeneity, and conduct extensive subgroup analyses to identify conditions under which different intervention types prove most effective.

Our findings indicate that formalisation policies in the region predominantly produce positive effects, with adverse outcomes being rare (5.1% of estimates). However, uncorrected pooled estimates overstate true effects substantially. For formality outcomes, bias correction reduces estimates by approximately one-third; for earnings outcomes, correction reduces estimates by half or more, depending on the method applied. Among intervention types, active labour market policies show consistent positive impacts on formality outcomes (0.238 standard deviations, $p < 0.05$). Social protection programmes that combine income support with activation components produce the largest effects on earnings (1.258 standard deviations, $p < 0.01$) but show negligible impacts on formality itself. Information campaigns and labour inspections also show positive effects on formality, although evidence bases remain more limited.

Three additional patterns emerge from our analysis. First, programme benefits strengthen over time. Long-term effects (measured 25+ months after programme completion) exceed short-term impacts by approximately 0.10 standard deviations for formality outcomes and 0.32 standard deviations for earnings outcomes. This pattern has important implications for evaluation design, as studies with follow-up periods shorter than two years may substantially underestimate programme impacts. Second, multi-component programmes outperform single-component interventions. Programmes combining multiple policy tools (such as training with apprenticeships or labour intermediation) show effects 0.07 standard deviations larger than single-component approaches. This suggests potential complementarities between interventions that address multiple barriers to formalisation. Third, methodological choices influence reported effect sizes. Experimental evaluations report effects 0.06 to 0.09 standard deviations smaller than comparable quasi-experimental studies, likely reflecting stricter methodological standards rather than inferior programme performance.

These results have important implications for policy design and evaluation in Latin America's formalisation agenda. Active labour market policies emerge as the most consistently effective tool for achieving formality outcomes, with substantial supporting evidence across diverse contexts. For policymakers seeking to improve earnings among vulnerable populations, social protection programmes with activation components show large effects, though these do not translate into formal employment transitions. Information campaigns offer promise as low-cost complements, particularly where knowledge gaps constitute important barriers. The strengthening of programme benefits over time underscores the importance of sustained policy implementation and longer evaluation periods. The consistent pattern of smaller effects after publication bias correction emphasises the need for realistic expectations about programme impacts.

Several evidence gaps require attention. Rigorous evaluations of payroll tax reductions, labour inspections, and simplification measures remain limited, with the evidence base heavily concentrated in active labour market policies (79% of estimates). Longer-term follow-up studies are essential, as programme benefits appear to strengthen over time. Research on combined policy approaches could identify complementarities between interventions, given that multi-component programmes show larger effects. Understanding mechanisms would enable better adaptation across contexts. Finally, firm-level formalisation requires more attention, as evidence remains concentrated on worker formalisation.

The paper proceeds as follows. Section 2 reviews conceptual foundations and existing empirical evidence on formalisation policies. Section 3 describes our methodology for data collection, coding, and meta-analytical techniques. Section 4 presents our results, including overall effect sizes and heterogeneity analysis by policy type, programme characteristics, and target population. Section 5 discusses policy implications and concludes with recommendations for future research and policy design.

2 Theoretical framework and evidence from previous meta-analyses

We define formalisation policies as interventions aimed at increasing the rate of formality among individuals and firms in Latin America. We include six types of policy interventions: (1) Active labour market policies (ALMPs), which include training, apprenticeships and internships, hiring subsidies, and labour intermediation; (2) Social protection policies with activation components; (3) Information campaigns; (4) Labour inspections; (5)

Payroll tax reductions; (6) Simplification of formalisation processes.[1] These categories reflect distinct theoretical mechanisms through which governments attempt to encourage transitions from informal to formal employment and enterprise registration.

2.1 Theoretical perspectives on informality

Understanding why informality emerges and persists is central to the theoretical and policy debates on formalisation. The conceptualisation of informality has evolved substantially since its introduction in the early 1970s. Initially associated with subsistence self-employment, the concept now encompasses a diverse range of occupations in both formal and informal economies (Bertranou, 2025). Three main theoretical perspectives have emerged to explain informality's origins and persistence, each with distinct implications for policy design.

The dualist perspective views informality as a residual sector that absorbs workers excluded from formal employment (Layard et al., 2005). Rigidities in formal labour markets, such as minimum wages, union agreements, or efficiency wage considerations, create a limited number of formal jobs. Workers who are unable to access formal employment turn to informal activities for survival (Fields, 2009). This perspective suggests that expanding formal job creation through economic growth and reducing barriers to formal employment should reduce informality. Formalisation policies under this view should focus on enhancing workers' productivity and employability to enable transitions into available formal positions.

The structuralist perspective emphasises the links between formal and informal economies (Portes et al., 1989). Informality arises from productive decentralisation processes, as formal firms subcontract production to reduce costs and increase flexibility. Informal units supply inputs and services to formal enterprises, creating functional interdependencies (Moser, 1978). Unlike the dualist view, this perspective suggests that economic growth alone may not reduce informality if formal sector expansion relies on informal production chains. Formalisation policies must therefore address both demand-side factors (formal firms' incentives to hire formally) and supply-side factors (informal workers' and firms' capacity to meet formal requirements).

The exit or legalist perspective posits that workers and firms choose informality after evaluating the costs and benefits of formalisation (Maloney, 2004; Levy, 2008). High compliance costs, complex regulations, or weak links between contributions and benefits may make informality attractive. Some workers may access social protection through non-contributory programmes or family members, reducing formal employment benefits (Perry et al., 2007). Firms may avoid registration costs and regulatory burdens by re-

[1] In our analysis, we consider a separate seventh category for policies that combine information campaigns and payroll tax reductions.

maining informal (De Soto, 1989). This perspective suggests that reducing formalisation costs and strengthening benefits should encourage transitions to formality. However, it recognises that even workers who choose informality can face poverty and precarious conditions, having selected this option as the least-bad alternative given their circumstances (Bertranou, 2025).

More recent integrated approaches recognise that the informal sector itself is heterogeneous (Fields, 1990, 2023; Ranis and Stewart, 1999). Some informal workers engage in subsistence activities with minimal formal sector connexions (consistent with dualism). Others are microentrepreneurs who strategically choose partial compliance with regulations (consistent with exit). A third segment comprises workers and small firms subordinated to larger formal enterprises through subcontracting chains (consistent with structuralism). This heterogeneity implies that different policy tools may be effective for different segments of the informal sector.

2.2 Formalisation policy mechanisms

Individual studies measure formality using varied operational definitions reflecting the ILO conceptual framework but adapting to data availability. For workers, common definitions include social security registration, pension system affiliation, possession of formal employment contracts, or access to employer-provided health insurance. For firms, definitions typically capture business registration with tax authorities, legal entity status, or participation in formal social security systems. We exclude studies using weak proxies such as contract type, earnings thresholds, or bookkeeping practices. Our meta-analysis addresses this heterogeneity by focussing on whether studies measure transitions towards formality, regardless of the specific operational definition employed.

2.2.1 Active labour market policies

Active labour market policies (ALMPs) address skill deficits, search frictions, and employment costs that may keep workers in informal employment (Layard et al., 2005; Kluve, 2010). Training programmes aim to improve human capital and productivity, with effects materialising as skills accumulate (Card et al., 2010). Apprenticeships and internships combine classroom learning with workplace experience, addressing skill gaps and information asymmetries about worker quality (Lerman, 2019). Labour intermediation services improve job matching efficiency and reduce search costs (Schmid et al., 2001). Employment subsidies alter relative hiring costs between formal and informal workers (Katz, 1996), although sustained impacts depend on whether temporary support enables permanent productivity increases (Abel et al., 2022).

Training programmes aimed at disadvantaged youth represent the most frequently evaluated intervention, which typically combines three months of classroom instruction with similar-length private sector internships. Major programmes include Colombia's *Jóvenes en Acción*, Peru's *Projoven*, and the Dominican Republic's *Juventud y Empleo* (Attanasio et al., 2011; Díaz and Jaramillo, 2006; Card et al., 2011). More intensive approaches include Brazil's *PRONATEC* offering 1,200 hours of technical instruction (Camargo et al., 2021) and Chile's *FOTRAB* combining 500 classroom hours with 360 hours of on-the-job training (Doerr and Novella, 2024). Labour intermediation reforms in Colombia and Mexico transformed basic job-matching into integrated systems offering counselling, job search assistance, and placement services (Pignatti, 2016; Morales-González et al., 2019). Hiring subsidies include direct worker bonuses, wage supplements, and combined subsidy-intermediation models (Abel et al., 2022; Berniell and de la Mata, 2017; Novella and Valencia, 2022).

2.2.2 Social protection with activation components

Social protection programmes with activation components combine income support with requirements or incentives for labour market participation. Income support allows for a more effective job search by reducing the pressure to accept immediately available opportunities (Marimon and Zilibotti, 1999), while activation requirements address skills and information barriers that prevent access to formal employment.

Temporary public employment initiatives constitute the primary type of intervention, providing work on community service projects or infrastructure development alongside income support. The duration of the programme typically ranges from three to eight months. Argentina's *Jefes de Hogar* Programme, implemented during the 2002 crisis, required twenty hours of community work, training, or private company employment for household heads who received 150 pesos monthly (Gasparini et al., 2009). Peru's *Construyendo Perú* provided four-month employment on small-scale urban infrastructure projects with job training (Escudero, 2018; Macroconsult, 2012). *Uruguay Trabaja* combined six to eight months of community service with competency workshops, vocational training, health services, and documentation assistance (Blanchard et al., 2025; Nogueira Puentes, 2018).

Impact pathways involve immediate effects (participation creates temporary formal employment and income) and longer-term effects through improved employability. Skills training, work experience, and formal work history signals should increase participants' attractiveness to private formal sector employers.

2.2.3 Information campaigns

Information campaigns address knowledge gaps and misperceptions about formalisation processes, costs, and benefits. Workers and firms may remain informal due to lack of information about registration procedures, overestimation of associated costs, or unawareness of benefits such as social protection access, credit availability, and government procurement opportunities (Bruhn and McKenzie, 2014). These interventions aim to correct knowledge deficits without altering the underlying economic incentives.

Several studies examined campaigns promoting Brazil's *Microempreendedor Individual* programme through in-person consultancy, virtual assistance through mobile messaging, and information sessions with text message reminders (Zucco et al., 2020, 2023; Lenz, 2017). Other interventions targeted domestic worker employers in Argentina through email and letter campaigns (Ohaco, 2023; Feld, 2024). Some approaches combined information with complementary elements: Brazil's *Descomplicar* tested brochures about *SIMPLES* registration procedures alongside fee waivers and free accounting services (Bruhn, 2013), while Colombia's Act No. 1429 workshops combined information provision with payroll tax reductions (Galiani et al., 2017).

These low-cost interventions are effective when information gaps constitute the primary barrier to formalisation. Where economic factors such as high taxes or procedural complexity dominate, information campaigns alone may generate limited responses (De Mel et al., 2013).

2.2.4 Labour inspections

Labour inspections aim to increase formalisation by raising the expected cost of remaining informal through monitoring and enforcement. Firms compare compliance costs (taxes, social security contributions, administrative requirements) against expected costs of noncompliance (Allingham and Sandmo, 1972; Becker, 1968). Increased inspection intensity raises detection probability, shifting calculations towards compliance. Inspections may also generate spillover effects if firms perceive generally higher enforcement (Ronconi, 2010).

Studies examine substantial variation in institutional design throughout the region, including decentralised systems with provincial variation in enforcement intensity in Argentina (Viollaz, 2018b), graduated enforcement regimes where requirements vary by firm size in Peru (Viollaz, 2018a), expansion of inspector numbers in Colombia (Pignatti, 2020), and interactions between trade liberalisation and local inspection intensity in Brazil (Ponczek and Ulyssea, 2022). While intensified enforcement raises detection probability and pushes firms towards compliance, higher labour costs may simultaneously reduce employment or shift production towards capital-intensive methods (Ronconi, 2010).

2.2.5 Payroll tax reductions

Payroll tax reductions lower non-wage labour costs associated with formal employment. In Latin American countries, formal employment typically involves employer contributions to social security, health insurance, and mandatory levies that substantially increase total labour costs beyond gross wages. Reducing these costs narrows the price differences between formal and informal labour, making formal hiring more attractive to employers and potentially increasing workers' take-home pay in competitive markets (Gruber, 1997). Effect magnitudes depend on labour supply and demand elasticities and on whether tax savings are captured by firms, passed to workers, or shared.

Reforms show considerable heterogeneity in design. Colombia's 2012 reform reduced employer payroll taxes and health contributions by 13.5 percentage points for workers earning up to ten times the minimum wage, phased in over 2013–2014 (Bernal et al., 2017; Morales and Medina, 2017; Kugler et al., 2017). Brazil's *SIMPLES* programme consolidated multiple employer taxes into a single payment while substantially reducing overall burdens for eligible micro- and small firms (Fajnzylber et al., 2011; Monteiro and Assunção, 2012; Piza, 2018). The Individual Microentrepreneur Programme extended this to own-account workers, combining web-based registration with reduced social security contributions (Rocha et al., 2018; Pereda et al., 2022). Brazil's 2006 reform for domestic workers allowed employers to deduct social security contributions from personal income taxes, effectively reducing the employer cost of formal registration (Madalozzo and Bortoluzzo, 2011).

Lower tax rates increase formal labour demand by reducing relative costs, with magnitudes shaped by substitution possibilities between formal and informal workers (Ulyssea, 2018). Responses differ substantially between extensive margins (whether firms formalise) and intensive margins (the share of formal workers within firms), particularly when fixed registration costs remain binding for smaller enterprises.

2.2.6 Simplification of formalisation processes

Simplification interventions reduce time and administrative burden associated with formal registration, targeting barriers including number of required procedures, completion time, and compliance complexity (Djankov, 2009). Rather than changing price incentives, simplification reduces fixed formalisation costs. Complex procedures also create corruption opportunities and disadvantage less educated entrepreneurs.

Mexico's Rapid Business Opening System created centralised facilities enabling simultaneous completion of federal, state, and municipal procedures, reducing the average registration time from 30 to 1.4 days (Aparicio, 2014; Bruhn, 2013). Colombia's 2006 reform unified payment systems for health insurance and pensions, simplifying compliance (Calderón-Mejía and Marinescu, 2012). Other approaches combined simplification with tax reductions through web-based platforms, consolidating registration requirements (Rocha et al., 2018; Fajnzylber et al., 2011; Monteiro and Assunção, 2012).

Simplification generates heterogeneous responses across firm types. Registration barriers bind more tightly for retail and service firms with low capital requirements, while manufacturing firms face additional constraints that administrative reforms alone cannot address (Bruhn, 2013; Monteiro and Assunção, 2012). Entrepreneur education levels and geographic proximity to registration offices further shape reform effectiveness.

2.2.7 Combined approaches

Many evaluated policies combine multiple intervention types, recognising that informality arises from intersecting barriers. Combined approaches may produce larger effects than individual interventions if constraints interact. Primary combinations include tax reduction with administrative simplification (Fajnzylber et al., 2011; Monteiro and Assunção, 2012; Piza, 2018; Rocha et al., 2018), regulatory change with enforcement and information campaigns (Feld, 2024), and ALMP packages that incorporate training, internships, job search assistance, and financial support (Alzuá and Brassiolo, 2006; Attanasio et al., 2011; Van Gameren, 2010).

Combined approaches present evaluation challenges: they may generate larger effects by addressing multiple constraints, but hinder identification of which components drive impacts. Most evaluations assess overall packages rather than disentangling individual element contributions, reflecting both practical constraints and the reality that many programmes were designed as integrated packages rather than separable components.

3 Methodology

3.1 Search and Selection Strategy

We selected impact evaluations through a systematic process. First, we established the following inclusion criteria: studies must assess the effects of labour formalisation policies implemented in Latin America, evaluate impacts at the individual or firm level, employ experimental or quasi-experimental methods controlling for selection into treatment, and provide sufficient statistical information (coefficient, standard error, sample size) to calculate effect sizes.

We searched multiple sources to identify relevant studies: international repositories (3ie Impact Evaluation Studies Registry and Inter-American Development Bank), academic databases (Scopus, Web of Science, SSRN, IDEAS, Google Scholar, and social-protection.org), studies from previous meta-analyses by Card et al. (2010), Kluve (2010), Escudero et al. (2019), Jessen and Kluve (2021), and Floridi et al. (2021), and evaluations conducted by government agencies within the region.

Search terms included "formalisation", "registration", "(in)formality", "labour intermediation", "public employment service", "labour inspection", "hiring subsidies", "job training", "social protection", "conditional transfer programmes", "wage subsidies", "payroll tax reduction" and "active policies" in Spanish or English. We conducted searches without restrictions on publication date, capturing studies from the earliest available to the most recent. For each study identified, we conducted forward and backward citation searches using snowballing throughout the entire process. We completed the search at the end of February 2024, identifying 152 studies eligible for inclusion.

Applying the selection criteria reduced the sample from 152 studies to 79 studies with 527 estimates. We excluded 45 studies at the paper level (removing 150 estimates) because they did not meet fundamental inclusion criteria: social protection programmes without activation components (67 estimates), studies not presenting estimations for probability of being formal (38 estimates), studies not evaluating formalisation policy effectiveness as a primary objective (30 estimates), and studies employing non-experimental identification methods (15 estimates).

We then applied estimate-level refinement to the remaining 112 studies, excluding 1,302 estimates for methodological and definitional reasons. We then applied estimate-level refinement to the remaining 112 studies, excluding 1,302 estimates for methodological and definitional reasons. Major exclusions included estimates from individual programme cohorts or intake rounds when aggregate or initial cohort treatment effects were available (247 estimates); for instance, when a study reported separate impacts for the 2005, 2006, and 2007 intakes of a training programme, we retained only the pooled estimate across all cohorts or, when no aggregate was available, the first cohort reported. Other major exclusions included outcome variables outside our scope (427 estimates) and cases where only combined treatments were available when disaggregated effects were required (166 estimates). Methodological exclusions included estimates lacking standard errors (110 estimates) and multiple formality definitions (95 estimates). We created 19 additional pooled gender estimates for studies reporting separate male and female effects but lacking overall estimates. Table A.1 in Appendix A provides a detailed breakdown of all exclusions.

3.2 Data Extraction and Sample Description

We extracted information systematically from studies meeting our inclusion criteria: (i) analysed intervention; (ii) outcome measuring intervention effectiveness; (iii) study characteristics (authors, title, type, and year of publication); (iv) intervention characteristics (country, target population, implementation period, duration, and number of components); (v) empirical analysis (methodological approach, data source, type of estimate and time horizon); and (vi) estimated impact (coefficient, standard errors, treatment, and control group sizes).

When studies reported separate estimates for different interventions, outcome variables, subgroups, or time horizons, we recorded each estimate separately. However, we applied consistency rules to avoid over-representation of individual studies. For studies presenting multiple treatment variations or cohorts, we extracted the aggregate treatment effect rather than the disaggregated results. When multiple identification strategies were reported, we selected the estimate identified as preferred or constituting the main focus of the findings. When multiple definitions of formality were used, we selected the estimate based on the most comprehensive definition.

Two authors selected studies and constructed the database independently, working on the same sample separately and then comparing the results to minimise bias. Additional technical details on data construction procedures, including the handling of interaction terms, the creation of pooled estimates, and sample size adjustments for standardised mean difference calculations, are provided in Appendix A.

Our final sample includes 79 impact evaluation studies with 527 impact estimates from formalisation policies implemented in Latin America between 1990 and 2020. Table 1 presents key characteristics of the sample. On average, each study contributes 6.7 estimates distributed among three outcome categories (formality, hours worked, earnings) and disaggregated by gender, age, and time horizon. This represents the most comprehensive collection of formalisation policy evaluations in the LAC region to date.

The temporal distribution shows growing policy activity, with 44% of estimates evaluating interventions from the 2010s and 34% from the 2000s. Unlike previous global meta-analyses that excluded Spanish-language studies or focused solely on published papers, our systematic search captured a comprehensive regional coverage: 19% of the studies are in Spanish and 44% are published in peer-reviewed journals. This inclusive approach addresses the limitations noted by Jessen and Kluve (2021) regarding Spanish-language studies and unpublished reports.

At the estimate level, 31% employ experimental designs while 69% use quasi-experimental methods. Among quasi-experimental approaches, difference-in-differences is more prevalent (35%), followed by instrumental variables (15%), propensity score matching (11%), and regression discontinuity (8%). Geographic coverage spans major LAC sub-regions: Southern Cone (46%), Andean countries (44%), and Central America and Caribbean (10%). Notably, 63% of estimates evaluate nationally implemented policies, while 37% focus on sub-national interventions.

Active labour market policies represent the largest category (79% of estimates). Our broader scope includes social protection programmes with activation components (10%), payroll tax policies (3%), information campaigns (3%), labour inspections (2%), and simplification measures (1%). Within ALMPs, 88% involve training programmes alone or in combination with other policies. The most common configuration combines training with apprenticeships or internships (49% of ALMP estimates). At the estimate level, 63% evaluate multi-component programmes while 37% assess single-component interventions.

The target population analysis reveals clear policy priorities: 53% focus on youth interventions, 34% target unemployed and job seekers, 8% focus on firms and self-employed workers, and 5% examine the general population. The outcome coverage shows that 59% examine formality, 10% examine hours worked, and 30% examine monthly earnings. Gender analysis is well-represented: 23% provide male-specific results, 24% provide female-specific estimates, and 53% report pooled results. Regarding age, 51% focus on youth populations (aged 15 to 24 or 15 to 30) while 49% examine pooled adult populations (18+). Temporal scope varies: 52% capture short-term effects (0 to 12 months), 22% examine medium-term impacts (13 to 24 months), and 26% assess long-term effects (25+ months).

3.3 Analytical Approaches

We employ two complementary analytical methods. First, we conduct a preliminary sign and significance analysis that examines patterns in the direction and statistical significance of the effects of treatment. This approach requires only the sign and significance reported for each estimate, making it applicable to the full sample of 527 estimates regardless of whether the studies report complete sample size information. Inspired by Escudero et al. (2019) and Kluve et al. (2019), we create a binary outcome variable coded as 1 if the estimate shows a statistically significant positive effect and 0 otherwise, then estimate linear probability models examining how the likelihood of obtaining positive significant results varies with study characteristics.

Second, we calculate standardised mean differences for all estimates with available treatment and control group sample sizes. This forms our primary analytical approach, preserving effect size information and enabling explicit publication bias correction.

3.4 Standardised Mean Differences

Studies in our sample assess the impacts of formalisation policy using three main outcomes: probability of being formal, monthly earnings, and hours worked. Each outcome is measured differently across studies. Table 2 presents the distribution of the outcome variables across the entire sample and by policy type.

The probability of being formal presents the greatest definitional variation. Our sample includes measures from both the employee and the firm perspectives. The definitions of formality for employees include the registration of social security (*RAIS* in Brazil, *SIPA* in Argentina, *PILA* in Colombia), the contribution to the pension and health insurance systems, the possession of formal contracts, and the access to social protection benefits. Firm formality measures include business registration with tax authorities, possession of municipal or state licences, registration in programmes such as *MEI* in Brazil, and compliance with regulatory obligations.

Similarly, earnings measures vary between levels and logarithmic transformations, capturing individual monthly wages, firm-level average wages, or regional wage indicators.

Table 1: Characteristics of included estimates

Characteristic	Freq.	%	Characteristic	Freq.	%
Policy type			Outcome category		
ALMPs	417	79.1%	Formality	313	59.4%
Labour inspections	11	2.1%	Hours worked	55	10.4%
Payroll tax reduction	13	2.5%	Monthly earnings	159	30.2%
PTR + Simplification	10	1.9%			
Sensitization	16	3.0%	Target population		
Simplification	5	0.9%	Youth	279	52.9%
Social protection	55	10.4%	Unemployed and job seekers	181	34.3%
			Firms and self-employed	43	8.2%
Sub-region			General workers	24	4.6%
South Cone	240	45.5%			
Andean countries	233	44.2%	Programme components		
Central America	36	6.8%	Single component	194	36.8%
Caribbean	18	3.4%	Multiple components	333	63.2%
Geographical coverage			Gender		
National	333	63.2%	All	279	52.9%
Sub-national	194	36.8%	Men	119	22.6%
			Women	129	24.5%
Evaluation period					
1990s	96	18.2%	Age		
2000s	180	34.2%	Pooled (18+)	256	48.6%
2010s	230	43.6%	Youth (15–24/30)	271	51.4%
2020s	21	4.0%			
			Timing of follow-up		
Method used			Short-term (0–12 months)	275	52.2%
DID	186	35.3%	Medium-term (13–24 months)	117	22.2%
IV	80	15.2%	Long-term (25+ months)	135	25.6%
PSM	59	11.2%			
RD	41	7.8%			
Randomised	161	30.6%	Total estimates	527	

Notes: ALMPs refers to Active Labour Market Policies, including hiring subsidies, apprenticeships, training programmes, and employment services. PTR refers to Payroll Tax Reduction. Social protection refers to policies with activation components (e.g., cash transfers conditional on training or job search). Formality includes both worker and firm formalisation outcomes. Youth (15–24/30) aggregates estimates targeting individuals aged 15–24 or 15–30, while Pooled (18+) includes estimates for the general adult population. Timing of follow-up refers to the period between programme participation and outcome measurement. Studies can contribute multiple estimates through disaggregation by outcome, time horizon, or population subgroup.

Table 2: Estimates availability by policy type and outcome

Policy Type	Formality	Hours	Income	Total
ALMPs	240	40	137	417
Labour inspections	6	2	3	11
Payroll tax	10	0	3	13
Payroll+Simplification	10	0	0	10
Sensitization	16	0	0	16
Simplification	5	0	0	5
Social protection	26	13	16	55
Total	313	55	159	527

Notes: Rows represent policy categories: ALMPs = Active Labour Market Programmes (training, apprenticeships, labour intermediation, hiring subsidies); Labour inspections = enforcement measures to increase formalization compliance; Payroll tax = reductions in employer social security contributions; Payroll + Simplification = combined interventions reducing taxes and streamlining registration procedures; Sensitization = information campaigns; Simplification = business registration and administrative procedure reforms; Social protection = programmes with activation components linking benefits to formalization. Columns represent outcome measures: Formality = probability of formal employment or firm registration; Hours = weekly hours worked (total or formal sector); Income = monthly earnings or wages (total or formal).

Hours worked estimates span weekly hours in levels or logarithms, with some studies focussing on formal sector hours specifically. Table B.1 in Appendix B provides a complete list of the definitions of formality, earnings, and hours worked used in all studies.

To enable systematic comparison across these different measurement approaches, we calculate standardised mean differences (SMD) for all estimates with available treatment and control group sample sizes. The SMD represents the difference between treatment and control group means expressed in standard deviation units, providing a scale-free metric that facilitates comparison across studies with different outcome measures (Higgings et al., 2023).

We use Hedges' g, which includes a small-sample bias correction, making it preferable to Cohen's d for studies with limited sample sizes. Hedges' g is calculated as:

$$g = \frac{Y_t - Y_c}{S_p} \times \left[1 - \frac{3}{4(df) - 1}\right] \tag{1}$$

where $Y_t - Y_c$ represents the estimated treatment effect, S_p denotes the pooled standard deviation, and the bracketed term provides the small-sample correction factor. Since pooled standard deviations are rarely reported in primary studies, we calculate S_p using:

$$S_p = SE \times \sqrt{\frac{n_c + n_t}{n_c \times n_t}} \times \sqrt{n_c + n_t - 2} \tag{2}$$

where n_t and n_c represent treatment and control group sample sizes, and SE is the standard error of the estimated treatment effect. A standardised effect of 0.2 is generally considered small, 0.5 medium, and 0.8 large, regardless of the original metric or baseline rate.

3.5 Meta-Analytical Models

Our analysis follows the comprehensive approach recommended by Irsova et al. (2024) for modern meta-analysis. We begin by presenting summary measures of overall effects across formalisation policy categories and outcomes. This is followed by a heterogeneity analysis and a publication bias assessment. We then estimate conditional effects through meta-regression models that correct for publication bias while controlling for study characteristics.

Meta-analysis employs two primary statistical models: fixed-effects and random-effects. In meta-analysis terminology, these terms differ from panel data econometrics. The distinction concerns whether studies represent a fixed set (the entire population of interest) or a random sample from a larger population. Fixed-effects models assume that all studies share the same true effect size, with observed variation reflecting only sampling error. These models use weights $w_j = 1/\sigma_j^2$, where σ_j^2 is the within-study variance. Random-effects models allow true effects to vary across studies due to genuine differences in implementation, context, or populations. These models incorporate between-study heterogeneity through weights $w_j = 1/(\sigma_j^2 + \tau^2)$, where τ^2 captures variation across studies beyond sampling error.

We present results from both specifications. Random-effects models are appropriate when true effects genuinely vary across studies due to differences in design, implementation, and context, as is common in policy evaluation research. The model explicitly incorporates between-study variance (τ^2), producing correct standard errors and confidence intervals when heterogeneity is present (Borenstein et al., 2021). However, random-effects models can produce biased estimates when publication selection is present (Stanley and Doucouliagos, 2011; Picchio, 2020). Our publication bias correction methods address this limitation.

Following Irsova et al. (2024), we also report unrestricted weighted least squares (UWLS) estimates. UWLS provides a simple summary statistic that uses inverse variance weighting without imposing additional structure on between-study heterogeneity. UWLS avoids the efficiency-bias tradeoff inherent in FE and RE models, providing a weighted average that does not require assumptions about the structure of heterogeneity between studies (Stanley, 2008).

3.6 Publication Bias Testing and Correction

Publication bias arises when researchers systematically under-report results that contradict theoretical expectations or policy objectives. In formalisation policy evaluation, researchers and policymakers may be reluctant to report small or statistically insignificant effects, particularly when substantial resources have been invested in programme implementation. When imprecise but large positive estimates are treated more favourably than imprecise but small estimates, a systematic bias emerges that overstates the impact of the programme in the literature (Stanley and Doucouliagos, 2011).

Our sample includes both published peer-reviewed studies and unpublished working papers, reducing but not eliminating publication bias. Academic journals may favour studies based on effect size magnitude and statistical significance rather than methodological quality alone (Stanley and Doucouliagos, 2011; Brodeur et al., 2016).

We assess publication bias through visual inspection, formal statistical tests, and multiple bias correction approaches. Visual inspection relies on funnel plots that display effect sizes against precision measures (inverse standard errors). In the absence of selection bias, studies should be distributed symmetrically around the mean effect, with greater dispersion among less precise estimates, creating an inverted funnel shape. Asymmetry, particularly sparse representation of small effects when precision is low, indicates systematic under-reporting.

For formal testing, we implement the Funnel Asymmetry Test-Precision Effect Test (FAT-PET) methodology (Egger et al., 1997; Stanley, 2008). This regresses the effect sizes on their standard errors:

$$g_i = \alpha + \beta \times SE(g_i) + \varepsilon_i \tag{3}$$

where g_i denotes the standardised effect size for the estimate i. The intercept α provides the precision-effect test (PET): the estimated effect when the precision is perfect ($SE = 0$), representing the average effect adjusted for publication bias. The slope coefficient β provides the funnel asymmetry test (FAT): a test of whether the effect sizes systematically relate to their standard errors. A statistically significant β coefficient indicates funnel plot asymmetry, suggesting publication bias. The model uses weighted least squares with inverse variance weights $(1/SE(g_i)^2)$ to address known heteroskedasticity.

Following Irsova et al. (2024), we employ multiple approaches to estimate Equation 3. We estimate the FAT-PET model using weighted least squares following Stanley (2008), ordinary least squares with standard errors clustered at the study level to account for within-study dependencies, fixed effects specifications, instrumental variables using sample size as an instrument for standard errors to address potential endogeneity, and precision-weighted regression using inverse standard errors as weights. This multi-method approach reveals the sensitivity of conclusions to different assumptions.

When FAT-PET indicates nonzero effects, we use the Precision Effect Estimate with Standard Error (PEESE) model, which employs quadratic approximation to minimise bias (Stanley and Doucouliagos, 2014). PEESE replaces the standard error term with the variance term:

$$g_i = \alpha + \beta \times SE(g_i)^2 + \varepsilon_i \tag{4}$$

The quadratic specification corrects for the non-linear relationship between effect sizes and their variances when publication bias is present. Linear FAT-PET (Equation 3) underestimates the true effect when the genuine effect is non-zero because it assumes a constant relationship between effect size and standard error. PEESE allows this relationship to vary with the level of imprecision, which better approximates how publication selection operates in practice (Stanley and Doucouliagos, 2014). The PEESE constant term provides the bias-corrected-effects estimate. We implement PEESE with clustered standard errors at the study level.

3.7 Meta-Regression with Covariates

Beyond publication bias correction, we address additional sources of effect heterogeneity by including comprehensive control variables. We account for within-study dependencies through clustered standard errors at the study level. Study-specific factors include programme type, duration, and multi-component design indicators, along with participant characteristics such as age and gender distributions. Methodological controls account for differences in evaluation design between studies. We include GDP growth rates in the year of programme implementation to control for macroeconomic conditions that may influence labour market transitions and programme effectiveness.

To incorporate these factors while maintaining publication bias correction, we extend the FAT-PET specification to include covariates:

$$g_i = \alpha + \beta \times SE(g_i) + \gamma X_i + \varepsilon_i \tag{5}$$

where i indexes individual estimates, X_i represents the covariate vector, and γ contains the estimates of the corresponding parameters. For PEESE specifications with covariates:

$$g_i = \alpha + \beta \times SE(g_i)^2 + \gamma X_i + \varepsilon_i \tag{6}$$

Our preferred specification uses the PEESE methodology with random-effects meta-regression, incorporating clustered standard errors at the study level to account for dependencies within the study estimate, following Irsova et al. (2024) recommendations for robust inference in meta-analysis.

4 Analysis and Results

We present our meta-analytical findings in four parts. First, we conduct a preliminary sign and significance analysis examining the distribution of positive, negative, and insignificant results across the entire sample. Second, we examine the overall effect sizes and heterogeneity patterns using standardised mean differences. Third, we assess publication bias using multiple diagnostic approaches. Fourth, we estimate conditional effects through a meta-regression analysis that simultaneously corrects for publication bias and controls for study characteristics.

4.1 Sign and Significance Patterns

We conducted a preliminary sign and significance analysis on the full sample of 527 estimates. This approach examines whether estimates are statistically significant and positive without requiring complete sample size information. Of all estimates, 287 (54.5%) show statistically significant positive effects, 213 (40.4%) are statistically insignificant, and 27 (5.1%) show statistically significant negative effects. Linear probability models regressing a binary indicator for positive significant results on study characteristics reveal three notable patterns. First, experimental evaluation designs consistently show large negative associations with the probability of reporting significant positive results, with coefficients ranging from -0.301 to -0.349 ($p < 0.01$), likely reflecting stricter methodological standards rather than worse programme performance. Second, information campaigns show positive associations of 0.505 to 0.543 ($p < 0.10$). Third, country fixed effects substantially increase explanatory power from $R^2 = 0.130$ to $R^2 = 0.223$, indicating country-level factors explain considerable variation beyond observable programme characteristics.

Most other factors show no consistent relationship with the likelihood of obtaining positive significant results. Programme type differences between ALMPs, social protection, and other interventions are not statistically significant. Similarly, neither the timing of effect measurement (short, medium, or long-term), programme duration, nor demographic targeting (gender or age focus) systematically predicts whether a study reports positive significant findings. This preliminary analysis provides descriptive context but discards information about effect magnitudes and does not correct for publication bias, motivating our focus on standardised mean differences in the remainder of this section. The complete results appear in Appendix B.

4.2 Sample Coverage and Overall Effects

From the 527 estimates available in our full sample, 444 (84.3%) disclosed sample sizes for treatment and control groups, allowing calculation of standardised mean differences using Hedges' g. Table 3 presents the distribution of estimates with SMD data across programme types and outcome categories. Active labour market policies dominate with 366 estimates (82% of the SMD sample), while social protection programmes with an activation component contribute 55 estimates. Several categories show limited representation: labour inspections (1 estimate), payroll tax interventions (5 estimates), and simplification measures (0 estimates). This concentration limits our ability to draw conclusions about these types of intervention.

Table 3: Estimates with Hedge'g availability by policy type and outcome

Policy Type	Formality	Hours	Income	Total
ALMPs	222	31	113	366
Labour inspections	1	0	0	1
Payroll tax	3	0	2	5
Payroll+Simplification	2	0	0	2
Sensitization	15	0	0	15
Simplification	0	0	0	0
Social protection	26	13	16	55
Total	269	44	131	444

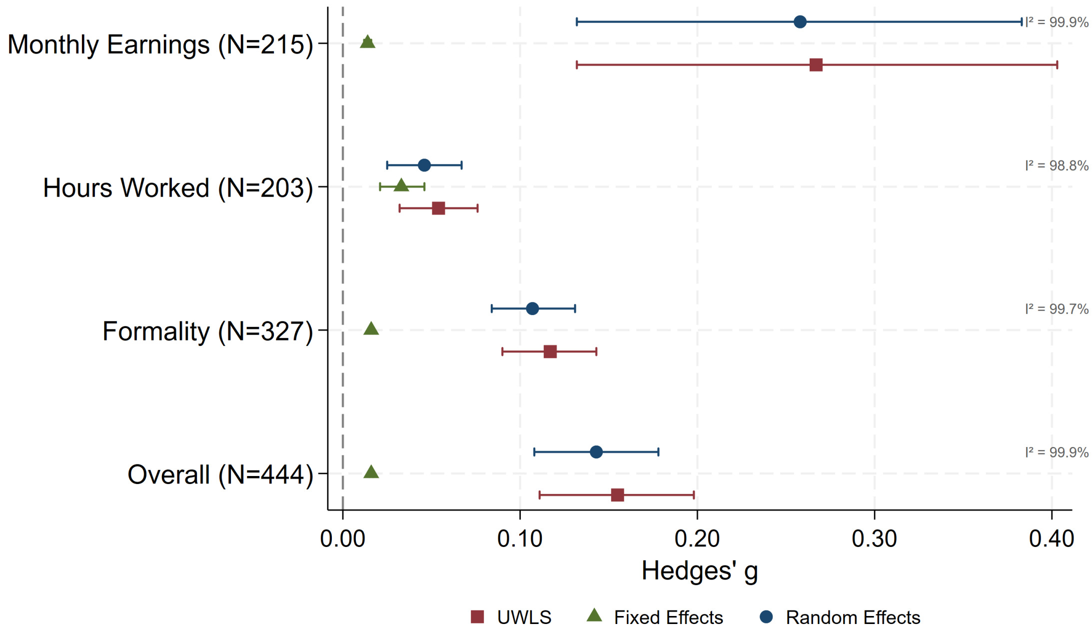
Notes: Rows represent policy categories: ALMPs = Active Labour Market Programmes (training, apprenticeships, labour intermediation, hiring subsidies); Labour inspections = enforcement measures to increase formalization compliance; Payroll tax = reductions in employer social security contributions; Payroll+Simplification = combined interventions reducing taxes and streamlining registration procedures; Sensitization = information campaigns and awareness-raising activities; Simplification = business registration and administrative procedure reforms; Social protection = programmes with activation components linking benefits to formalization. Columns represent outcome measures: Formality = probability of formal employment or firm registration; Hours = weekly hours worked (total or formal sector); Income = monthly earnings or wages.

Figure 1 presents pooled effect sizes across all studies using the three estimation approaches described in Section 3.5. The random effects estimate for the pooled sample shows a positive and statistically significant effect of 0.143 standard deviations, representing approximately one-seventh of a standard deviation improvement in outcomes. The substantial difference between fixed effects (0.016) and random effects estimates indicates considerable between-study heterogeneity. The I^2 statistic of 99.9% confirms that variation across studies reflects genuine differences in treatment effects rather than sampling error alone.

The effects vary by outcome type. The formalisation outcomes show consistent and statistically significant effects in all estimation methods: UWLS of 0.117, FE of 0.017, and RE of 0.107. Hours worked shows smaller effects with RE of 0.046, suggesting that formalisation policies primarily affect employment status and earnings rather than work intensity. Monthly earnings shows the largest point estimates with RE of 0.258, but with wide confidence intervals reflecting high uncertainty.

Programme-specific analyses appear in Appendix B Table B.2. ALMPs produce

Figure 1: Meta-Analytic Effect Sizes by Outcome Category



Notes: Meta-analytic effect sizes with 95% confidence intervals comparing three estimation methods: unrestricted weighted least squares (UWLS), fixed effects (FE), and random effects (RE). Effect sizes represent standardised mean differences (Hedges' g). Sample sizes (N observations) shown in parentheses on y-axis. I^2 statistics indicate percentage of variation across studies due to heterogeneity rather than chance, shown for random effects estimates. The "Overall" row presents effects pooled across all three outcome categories. Sample includes estimates from 79 formalisation policy evaluations in Latin America (1990-2020).

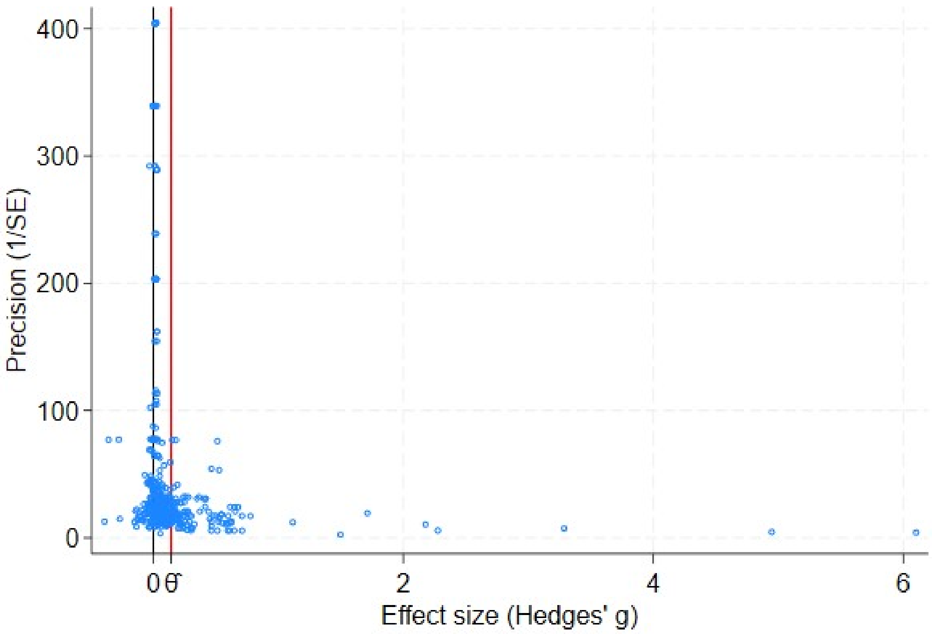
consistent positive and statistically significant effects across all outcomes, with formality impacts of 0.117 and monthly earnings effects of 0.114. Social protection programmes show large earnings effects of 1.264, although this estimate relies on limited observations. Information campaigns show modest but statistically significant formality effects of 0.015.

The outlier analysis following Havránek et al. (2020) criteria identifies 52 extreme observations (11.7% of the SMD sample) with effect sizes outside the $[-1, 1]$ range or precision measures beyond reasonable bounds. Removing these outliers reduces the overall mean effect from 0.155 to 0.115 and substantially decreases variability (the standard deviation falls from 0.464 to 0.176), indicating that the main results are not driven by extreme studies.

4.3 Publication Bias

We assess publication bias through funnel plot inspection and formal statistical tests. Figures 2 and 3 show funnel plots for pooled outcomes, with and without outliers. Both plots exhibit clear asymmetry, with studies concentrated in the bottom-right quadrant (large positive effects with low precision) and sparse representation in the bottom-left quadrant (small or negative effects with large standard errors). Approximately 73% of the estimates fall in the right half of the funnel plot, compared to the 50% expected under symmetry. This pattern shows systematic under-representation of studies finding smaller effects when precision is low, consistent with publication selection favouring studies that report larger positive impacts. The asymmetry persists after outliers are excluded.

Figure 2: Funnel plot: Hedges' g with pooled estimate



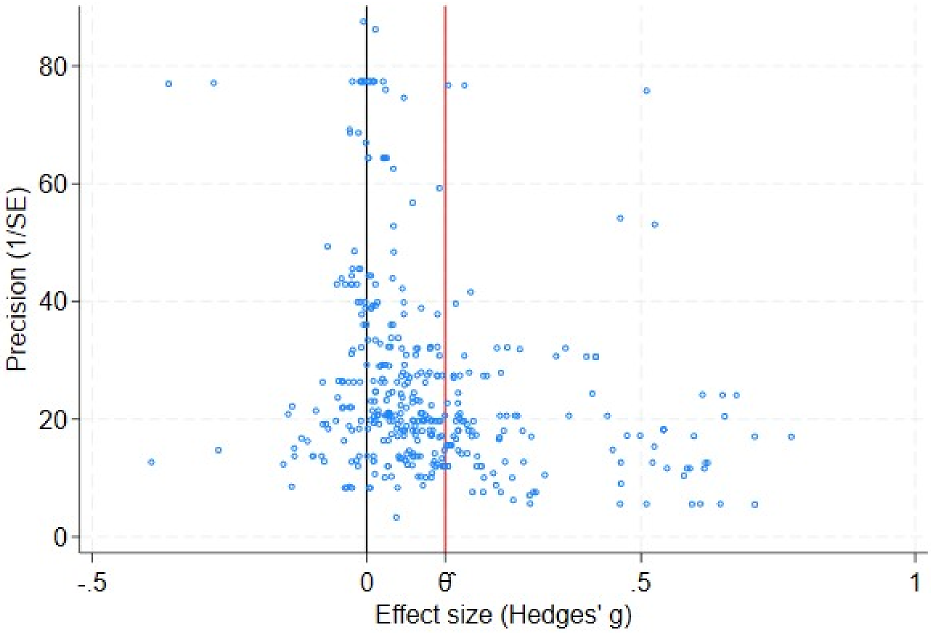
Notes: θ (red line) shows pooled random-effects estimate.

The Egger test provides formal confirmation of these visual patterns. For the pooled sample (N=444), both random effects and fixed effects models strongly reject the null hypothesis of no publication bias (Appendix B Table B.2). Publication bias appears in all outcome categories, with particularly strong evidence for monthly earnings.

Table 4 presents bias-corrected estimates using the multiple approaches described in Section 3.6. Each row represents a different method to estimate the FAT-PET model (Equation 3). The constant term provides the precision-effect test (PET), representing the bias-corrected average effect. The standard error coefficient provides the funnel asymmetry test (FAT), which tests for publication bias.

The results show sensitivity to methodological choices. The weighted least squares method following Stanley (2008) generates a bias-corrected effect of 0.004, representing a 97% reduction from the uncorrected pool estimate of 0.143. The standard error coefficient of 2.412 is statistically significant at the 1% level, providing strong evidence of publication bias. When accounting for study-level clustering through ordinary least squares with clustered standard errors, the bias-corrected estimate becomes −0.117, although statistically insignificant. The standard error coefficient remains large at 5.267 and significant at the 10% level, suggesting publication selection may entirely account for observed positive effects once within-study dependencies are addressed.

Figure 3: Funnel plot: Hedges' g with pooled estimate, outliers excluded



Notes: θ (red line) shows pooled random-effects estimate. Outliers excluded based on Havránek et al. (2020) criteria.

The fixed effects specification produces a bias-corrected estimate of -0.246 with a standard error coefficient of 7.767, significant at the 5% level. The instrumental variable approach, using sample size as an instrument for standard errors to address potential endogeneity, results in a bias-corrected estimate of 0.013 with a standard error coefficient of 2.752, significant at the 5% level. The precision-weighted approach produces a bias-corrected effect of -0.011 and a standard error coefficient of 3.219, significant at the 1% level.

All five approaches reject the null hypothesis of no publication bias at conventional significance levels through the FAT test, while none of the bias-corrected estimates from the PET test differ significantly from zero. The wide range of bias-corrected estimates (from -0.246 to 0.013) reflects the uncertainty of the model about the correct specification for publication bias correction.

Following Stanley and Doucouliagos (2014), when PET rejects the null of a zero effect, the PEESE estimator should be used, as it better approximates the quadratic relationship between effect sizes and their standard errors. Since our PET estimates do not reject zero in most specifications, we present PEESE primarily for comparison. The PEESE-corrected estimate of 0.071, significant at the 5% level with clustered standard errors, suggests effects approximately 50% smaller (0.071 versus 0.143) than the uncorrected random-effects pooling indicates. The PEESE standard error squared coefficient of 18.927 confirms the presence of publication bias, although with less precision than the linear FAT-PET specifications.

Table 4: FAT-PET-PEESE Analysis: All Outcomes Pooled

Model	**Constant (PET)**	**SE Coefficient (FAT)**
WLS (Stanley 2008)	0.004	2.412***
	(0.003)	(0.276)
OLS Clustered	-0.117	5.267*
	(0.113)	(2.895)
Fixed Effects	-0.246	7.767**
	(0.195)	(3.785)
IV (Sample Size)	0.013	2.752**
	(0.026)	(1.108)
Precision (1/SE)	-0.011	3.219***
	(0.018)	(1.146)
PEESE Analysis		
PEESE Clustered	0.071**	18.927
	(0.032)	(11.962)

Notes: N=444 from 79 studies. The table presents publication bias assessment following Equation 3 (rows 1–5) and Equation 4 (row 6). PET (Precision Effect Test): The constant term provides the bias-corrected average effect, representing the estimated effect when precision is perfect (SE=0). FAT (Funnel Asymmetry Test): The SE coefficient tests whether effect sizes systematically relate to their standard errors; a significant coefficient indicates publication bias. WLS implements inverse variance weighting following Stanley (2008). OLS Clustered uses study-level clustered standard errors. Fixed Effects includes study-level fixed effects. IV instruments standard errors with sample size. Precision weights by 1/SE. PEESE uses variance (SE^2) instead of SE and is recommended when PET rejects zero effect (Stanley and Doucouliagos, 2014). Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bias patterns vary by outcome. Formality outcomes show PEESE-corrected effects of 0.087, significant at the 1% level (Appendix B Table B.3). Hours worked exhibit mixed signals with a PEESE-corrected estimate of 0.022, not statistically significant. Monthly earnings display the strongest bias, with PEESE correction resulting in -0.097, significant at the 10% level, suggesting publication selection may entirely account for observed positive effects in this outcome category. Detailed outcome-specific results and funnel plots appear in Appendix B (Tables B.3 to B.5, Figures B.1 to B.3).

These findings indicate that conclusions about formalisation policy effectiveness depend critically on addressing publication selection. The substantial reductions in effect sizes after bias correction suggest the uncorrected literature overstates policy impacts, particularly for earnings outcomes.

4.4 Conditional Effects from Meta-Regression

We now examine which programme features are associated with larger or smaller effects after correcting for publication bias and controlling for study characteristics. The meta-regression simultaneously estimates the independent association of each characteristic while holding others constant.

4.4.1 Specification

The meta-regression model follows Equation 6:

$$g_i = \alpha + \beta \times SE(g_i) + \gamma X_i + \varepsilon_i$$

where g_i is Hedges' g for estimate i, $SE(g_i)$ corrects for publication bias, X_i represents study characteristics (programme type, target population, evaluation design, duration, outcome category, and contextual factors), and ε_i is the error term. The subscript i indexes the 444 estimates with SMD data from 79 studies. Standard errors cluster at the study level to account for dependencies between multiple estimates from the same evaluation.

We estimate both random effects and fixed effects specifications. Random effects allows true effects to vary across studies. Fixed effects assumes a common underlying effect. All models control for programme characteristics, participant demographics, evaluation design, and contextual factors. The baseline category for programme types is combined interventions that pair payroll tax reductions with administrative simplification of registration procedures (such as Brazil's SIMPLES programme, which reduced tax burdens while streamlining the registration process). The coefficients represent differences relative to this baseline.

4.4.2 Programme Type Effects

Table 5 presents meta-regression results with correction for publication bias. We focus first on programme type effects, then examine other dimensions of heterogeneity.

For formality outcomes (columns 3 to 4), active labour market policies show robust positive effects under both specifications: 0.238 ($p < 0.05$) for random effects and 0.169 ($p < 0.01$) for fixed effects. This makes ALMPs the most consistently effective policy tool relative to the baseline category. The convergence among the estimation methods reinforces this finding.

Labour inspections show the largest point estimates for formality: 0.320 ($p < 0.10$) for random effects and 0.146 ($p < 0.01$) for fixed effects, although the random effects estimate shows considerable uncertainty. The single observation in the SMD sample limits the interpretation. Information campaigns show positive formality effects under random effects of 0.226 ($p < 0.10$), although the fixed effects estimate of 0.019 is not statistically significant. This suggests that information-based approaches may produce modest improvements in formality, but the evidence is less robust than for ALMPs.

Payroll tax reductions alone show positive but statistically insignificant effects on formality under random effects of 0.186, while under fixed effects, the coefficient is significant but modest in magnitude (0.095, $p < 0.01$). The coefficients suggest payroll tax

Table 5: Meta-regression results: SE publication bias correction

	All Outcomes		Formality		Earnings		Hours Worked	
	RE	FE	RE	FE	RE	FE	RE	FE
(i) Programme type (base: payroll + simplification)								
ALMPs	0.290	0.158***	0.238**	0.169***	0.352	0.105***	−0.203	−0.203
	(0.204)	(0.028)	(0.109)	(0.028)	(0.313)	(0.017)	(0.207)	(0.207)
Social protection	0.699***	0.140***	0.033	−0.021	1.258***	0.588***		
	(0.210)	(0.030)	(0.117)	(0.032)	(0.347)	(0.031)		
Labour inspections	0.499	0.149***	0.320*	0.146***				
	(0.347)	(0.052)	(0.183)	(0.052)				
Sensitization	0.204	0.023	0.226*	0.019				
	(0.233)	(0.042)	(0.127)	(0.042)				
Payroll tax	0.186	0.085***	0.186	0.095***				
	(0.236)	(0.028)	(0.137)	(0.030)				
(ii) Effect duration (base: long-term)								
Medium-term	−0.097**	0.000	−0.046	0.003	−0.317***	−0.085***	−0.010	−0.010
	(0.042)	(0.002)	(0.029)	(0.002)	(0.105)	(0.010)	(0.033)	(0.033)
Short-term	0.061	0.006***	−0.054*	0.006***	−0.194	−0.037***	0.017	0.017
	(0.044)	(0.002)	(0.029)	(0.002)	(0.130)	(0.010)	(0.040)	(0.040)
(iii) Outcome category (base: hours worked)								
Formality outcome	0.073	0.056***						
	(0.052)	(0.007)						
Earnings outcome	0.121**	0.053***						
	(0.054)	(0.007)						
(iv) Target group (base: men, pooled samples)								
Female	0.012	0.008***	0.007	0.012***	−0.058	−0.002	−0.002	−0.002
	(0.041)	(0.002)	(0.028)	(0.002)	(0.099)	(0.003)	(0.026)	(0.026)
Gender pooled	0.056	0.010***	0.011	0.012***	0.131	0.006**	−0.010	−0.010
	(0.038)	(0.001)	(0.026)	(0.002)	(0.093)	(0.003)	(0.030)	(0.030)
Youth(18-24)	−0.024	−0.011**	0.007	−0.028***	0.042	0.125***	0.012	0.012
	(0.039)	(0.004)	(0.025)	(0.005)	(0.108)	(0.014)	(0.048)	(0.048)
(v) Programme duration (base: long duration)								
Short duration	−0.040	−0.013*	−0.021	0.014	−0.129	−0.039***		
	(0.056)	(0.007)	(0.041)	(0.010)	(0.132)	(0.013)		
Medium duration	−0.265***	−0.009	−0.010	0.045***	−0.656***	−0.177***	−0.173	−0.173
	(0.052)	(0.007)	(0.039)	(0.010)	(0.126)	(0.015)	(0.147)	(0.147)
(vi) Evaluation design (base: non-experimental)								
Experimental design	−0.059	−0.064***	−0.090***	−0.070***	−0.324**	−0.138***	0.014	0.014
	(0.045)	(0.004)	(0.030)	(0.005)	(0.144)	(0.011)	(0.082)	(0.082)
(vii) Contextual factors								
GDP growth	0.023***	0.005***	0.004	0.007***	0.026*	0.008***	0.013	0.013
	(0.004)	(0.000)	(0.003)	(0.001)	(0.014)	(0.001)	(0.013)	(0.013)
(viii) Publication bias correction								
Standard Error (SE)	6.639***	3.435***	1.974***	3.265***	11.588***	5.388***	−0.314	−0.314
	(0.503)	(0.092)	(0.397)	(0.116)	(1.156)	(0.211)	(1.398)	(1.398)
Constant	−0.512**	−0.143***	−0.142	−0.109***	−0.356	−0.078***	0.306	0.306
	(0.227)	(0.030)	(0.125)	(0.031)	(0.335)	(0.023)	(0.261)	(0.261)
Pub. bias correction	SE	SE	SE	SE	SE	SE	SE	SE
Observations	392	392	233	233	119	119	40	40

Notes: Meta-regression with SE publication bias correction. RE = Random Effects, FE = Fixed Effects. Full specification includes all covariates and country fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

policies alone do not substantially outperform the combined baseline approach. Social protection programmes show no significant formality effects: 0.033 for random effects and -0.021 for fixed effects.

However, for earnings outcomes (columns 5 to 6), social protection programmes show strong positive effects under both specifications: 1.258 ($p < 0.01$) for random effects and 0.588 ($p < 0.01$) for fixed effects. This substantially exceeds impacts on formality itself. ALMPs show positive earnings impacts of 0.352 for random effects and 0.105 ($p < 0.01$) for fixed effects, although the random effects estimate does not reach statistical significance.

Hours worked outcomes (columns 7 to 8) show limited evidence of programme effectiveness differences. The small sample (N=40) and the absence of observations for most categories of programmes prevent reliable conclusions.

4.4.3 Effect Measurement Timing and Programme Duration

The timing of effect measurement and the duration of the programme show distinct patterns that require careful interpretation. These represent different dimensions: effect measurement timing refers to when outcomes are measured relative to programme completion (follow-up period), while programme duration refers to how long the intervention itself lasted.

The timing of the effect measurement shows important patterns. For formality outcomes, short-term effects (measured 0 to 12 months after programme completion) show negative associations relative to long-term measurement: -0.054 ($p < 0.10$) for random effects and 0.006 ($p < 0.01$) for fixed effects. Medium-term effects show smaller and statistically insignificant associations. This indicates that formality impacts take time to materialise. The longer follow-up periods tend to report larger effects.

For earnings, the pattern is more pronounced. Medium-term effects show large negative coefficients relative to long-term measurement: -0.317 ($p < 0.01$) for random effects and -0.085 ($p < 0.01$) for fixed effects. Short-term effects also show negative associations. This indicates earnings impacts also strengthen considerably over time.

These temporal patterns likely reflect several mechanisms. Skills accumulate and employers observe productivity over time. Job search takes time to find good matches. The stigma of unemployment or informality fades. Networks develop through formal employment experience. These processes suggest that short-term evaluations substantially underestimate programme impacts, particularly for earnings outcomes.

Programme duration shows strong associations with earnings outcomes. Medium-duration interventions show large negative coefficients relative to long-duration programmes: -0.656 ($p < 0.01$) for random effects and -0.177 ($p < 0.01$) for fixed effects. Short-duration programmes also show negative associations. This implies that sustained programme implementation is particularly important for achieving earnings gains.

For formality outcomes, programme duration shows mixed patterns with no clear relationship. The coefficients for short-duration and medium-duration programmes are small and statistically insignificant.

4.4.4 Demographic Patterns

Gender patterns in formality outcomes remain modest. Women-specific estimates show small positive effects: 0.007 for random effects and 0.012 ($p < 0.01$) for fixed effects. Gender-pooled estimates display similarly small positive associations. The programmes report similar formality effects for participants in women and men.

Youth-targeted estimates show coefficients of 0.007 for random effects and -0.028 ($p < 0.01$) for fixed effects. This indicates that youth-focused programmes report smaller formality effects relative to programmes targeting broader age groups, not that youth programmes are ineffective in absolute terms. The positive association in the sign-and-significance analysis (Appendix B) confirms that youth programmes produce positive effects; the meta-regression coefficient indicates that these effects are somewhat smaller than for general-population programmes.

For earnings results, youth-targeted estimates show positive associations of 0.042 for random effects and 0.125 ($p < 0.01$) for fixed effects, while female-specific estimates show small negative coefficients. This indicates systematic differences in how programmes affect earnings across demographic groups.

4.4.5 Methodological and Contextual Factors

Experimental evaluation designs show strong negative associations with effect sizes across all outcomes. For formality, the coefficient is -0.090 ($p < 0.01$) for random effects and -0.070 ($p < 0.01$) for fixed effects. For earnings, experimental designs show -0.324 ($p < 0.05$) for random effects and -0.138 ($p < 0.01$) for fixed effects. This likely reflects lower attrition bias and reduced selection on unobservables in experimental designs. Randomised experiments prevent systematic sorting of high-ability participants into treatment, which can inflate quasi-experimental estimates when unobserved characteristics correlate with both treatment assignment and outcomes. Experimental studies report effect sizes approximately 0.06 to 0.09 standard deviations smaller than comparable quasi-experimental evaluations.

GDP growth shows positive associations with effectiveness in the pooled sample: 0.023 ($p < 0.01$). For earnings specifically, the coefficient is 0.026 ($p < 0.10$) for random effects and 0.008 ($p < 0.01$) for fixed effects. This suggests that formalisation policies may be more effective in boosting earnings when implemented during periods of economic expansion.

The publication bias correction term (standard error) shows strong positive and statistically significant coefficients across all outcome categories, confirming the systematic overestimation of effects in the uncorrected literature. For formality, the coefficient is 1.974 ($p < 0.01$) for random effects and 3.265 ($p < 0.01$) for fixed effects. For earnings, the publication bias term reaches 11.588 ($p < 0.01$) for random effects, indicating a particularly severe publication bias in the earnings literature.

4.4.6 Synthesis

Conditional effect estimates provide several policy-relevant insights. Active labour market policies emerge as the most effective tool for achieving formality outcomes. The effect size of approximately 0.24 standard deviations falls between conventional benchmarks for small (0.2) and medium (0.5) effects. For formality outcomes, where the underlying variable measures the probability of formal employment, this corresponds to an increase of roughly 12 to 14 percentage points in the likelihood of being formally employed. Social protection programmes with activation components show large earnings effects (1.26 standard deviations) but limited formality impacts, suggesting that these interventions primarily affect income rather than transitions to employment status. Labour inspections and information campaigns show positive formality effects, although the evidence is less robust given limited observations.

Programme benefits strengthen over time. Long-term effects exceed short- and medium-term impacts across outcome categories. This pattern is most pronounced for earnings, where medium-term effects are 0.32 standard deviations smaller than long-term effects. For formality, short-term evaluations report effects 0.05 standard deviations smaller than long-term evaluations. These temporal patterns underscore the importance of extended follow-up periods in impact evaluations.

The consistent pattern of smaller effects after correction for publication bias emphasises the need for realistic expectations about the impact of programmes. Uncorrected pooled estimates overstate true effects considerably. Bias correction reduces formality estimates by roughly one-third and earnings estimates by half or more. This has important implications for programme planning and resource allocation. Robustness checks using alternative publication bias correction methods and outcome-specific meta-regressions appear in Appendix B (Tables B.8 to B.11). These alternative specifications produce qualitatively similar results, with the relative ranking of programme effectiveness and the direction of key associations remaining consistent across methods. The descriptive heterogeneity analysis by programme type, demographic targeting, and implementation features also appears in Appendix B.

5 Discussion and conclusions

This meta-analysis provides the first comprehensive synthesis of formalisation policy evidence for Latin America, analysing 79 impact evaluations with 527 estimates from interventions implemented between 1990 and 2020. We address gaps in the literature by including Spanish-language studies and unpublished reports, examining interventions targeting both worker and firm formalisation, and calculating standardised mean differences for 444 estimates while controlling for publication bias.

Main Findings

Three main findings emerge. First, formalisation policies predominantly produce positive effects, with adverse outcomes being rare (5.1% of estimates). However, systematic publication bias substantially reduces effect sizes when corrected. Uncorrected pooled estimates substantially overstate true effects. After correction, formality effects fall by approximately one-third, while earnings effects are reduced by half or more depending on the correction method.

Second, programme effectiveness varies substantially across intervention types. Active labour market policies show the most consistent positive effects on formality, producing improvements that fall between small and medium effect sizes by conventional standards. In practical terms, these effects correspond to increases of 12 to 14 percentage points in the probability of formal employment. Information campaigns also produce positive effects, though more modest in magnitude and with weaker statistical confidence. Labour inspections show promising point estimates but rely on limited evidence. For earnings outcomes, social protection programmes with activation components produce the largest effectssubstantially larger than formality impactsbut show negligible effects on transitions to formal employment. This divergence indicates that these programmes primarily affect income through direct transfers and improved job quality rather than by helping workers move into formal positions.

Third, the benefits of the programme strengthen over time. Long-term effects exceed short- and medium-term impacts across outcomes. For earnings, medium-term effects are 0.32 standard deviations smaller than long-term effects ($p < 0.01$). Programme duration matters: medium-duration interventions produce effects 0.27 standard deviations smaller than long-duration programmes ($p < 0.01$). Multi-component programmes outperform single-component interventions (0.173 versus 0.101 standard deviations).

Mechanisms and Interpretation

The observed patterns suggest different operating mechanisms across intervention types. Active labour market policies work through human capital accumulation, reduced search frictions, and lower hiring costs. Results show that these mechanisms translate more directly into formality outcomes than earnings gains. Information campaigns address knowledge gaps about registration procedures and formalisation benefits, explaining their modest but consistent formality effects. Labour inspections operate through deterrence, raising the expected costs of non-compliance.

Social protection programmes with activation components show large earnings effects without corresponding formality gains. Income support allows for a more effective job search and may finance investments in self-employment or skill acquisition. These mechanisms increase earnings in both formal and informal work, explaining why earnings impacts substantially exceed formality effects. The temporary nature of these programmes (typically 3 to 8 months) may limit formal employment transitions while allowing income improvements.

The heterogeneity in the effectiveness of the programme is aligned with integrated theoretical perspectives that recognise multiple segments of the informal sector (Fields, 2023; Bertranou, 2025). The effectiveness of ALMPs' effectiveness for formality supports human capital and employability mechanisms emphasised in dualist perspectives. The positive effects of information campaigns are consistent with exit theories that highlight knowledge barriers. The persistence of informality despite positive programme effects reflects structuralist insights about formal-informal links that economic growth alone cannot break.

The temporal strengthening of the effects likely reflects several mechanisms. Skills accumulate, and employers observe productivity over time. Job search takes time to find good matches. The stigma of unemployment or informality fades. Networks develop through formal employment experience. Multi-component programmes show larger effects potentially because they address multiple binding constraints simultaneously, creating complementarities between skills and job search support.

Our demographic findings differ from Escudero et al. (2019), who found stronger effects for women and youth. We observe modest gender differences (0.012 standard deviations for female-specific estimates, not significant) and negative associations for youth-targeted interventions (-0.028 standard deviations, $p < 0.01$ for formality). These divergences reflect our broader policy scope beyond ALMPs, explicit publication bias corrections, and a focus on formality rather than employment. The negative youth coefficient indicates that youth programmes report smaller formality effects than programmes targeting broader age groups, not that youth programmes are ineffective in absolute terms.

Experimental designs consistently report smaller effects than quasi-experimental studies (0.09 standard deviations smaller for formality, $p < 0.01$). This likely reflects lower attrition bias and reduced selection on unobservables in experimental designs. Randomised experiments prevent systematic sorting of high-ability participants into treatment, which can inflate quasi-experimental estimates when unobserved characteristics correlate with both treatment assignment and outcomes. GDP growth shows positive associations with effectiveness (0.023 standard deviations per percentage point, $p < 0.01$), confirming that formalisation policies work better during economic expansions.

Policy Implications

Several practical recommendations emerge. First, prioritise active labour market policies combining training with apprenticeships or labour intermediation for formality outcomes. These show consistent positive effects with substantial supporting evidence. Second, consider information campaigns as low-cost complements where knowledge gaps constitute barriers. Cost-effectiveness considerations favour information campaigns despite their modest effect sizes. These low-cost interventions may offer superior value relative to resource-intensive ALMPs or social protection programmes, although rigorous cost-benefit analysis remains scarce. Third, use social protection programmes with activation components to support vulnerable populations, recognising that these primarily affect earnings rather than formality status.

Fourth, invest in longer-term programmes. The benefits of the programme are strengthened over time, and sustained interventions are more effective than brief ones. Fifth, plan evaluation periods extending beyond 24 months to capture full impact trajectories. Short-term evaluations substantially underestimate programme impacts, particularly in terms of earnings. Sixth, maintain realistic expectations. After correcting for publication bias, even the most effective interventions produce modest formality effectsmeaningful improvements, but not transformative ones. Policymakers should therefore view these programmes as useful components of a broader formalisation strategy rather than standalone solutions. No single intervention will dramatically reduce informality; sustained effort across multiple policy domains remains essential.

Finally, allocate resources for a rigorous evaluation of the intervention types understudied. The evidence is primarily concentrated on active labour market policies (79% of estimates). Rigorous evaluations of payroll tax reductions, labour inspections, and simplification measures remain scarce despite suggestive evidence of effectiveness.

Limitations and Future Research

Our analysis faces several limitations. Excluding studies without complete statistical information may introduce selection bias, although including unpublished studies and evaluations in Spanish mitigates this concern. The predominance of ALMP evaluations reflects historical priorities rather than necessarily indicating superior effectiveness. High heterogeneity ($I^2 > 99\%$) indicates that unobserved factors such as implementation quality, local labour market tightness, informal sector structure, enforcement capacity, and complementary policies play crucial roles that our covariates cannot fully capture. This reinforces the need for context-sensitive policy adaptation rather than one-size-fits-all approaches. Pooling estimates across different formality definitions may mask important distinctions between dimensions of formalisation.

Future research should address several priorities. First, conduct rigorous evaluations of understudied intervention types, particularly payroll tax reductions, labour inspections, and simplification measures. Second, examine combined policy approaches to understand which combinations work best and whether effects are additive or synergistic. Third, investigate mechanisms through which interventions affect outcomes. Understanding why interventions work enables better adaptation across contexts. Fourth, assess implementation quality and its relationship to effectiveness. Implementation quality likely matters substantially but remains understudied. Finally, examine firm-level formalisation more extensively, as evidence concentrates on worker outcomes.

Concluding Remarks

Informality remains a defining challenge for Latin American labour markets. Our meta-analysis demonstrates that formalisation policies produce positive outcomes, but effect sizes are modest and vary across contexts and intervention types. The substantial publication bias we document underscores the need for realistic expectations. After bias correction, the effects remain positive, but smaller than the uncorrected estimates suggest.

Active labour market policies show consistent evidence of the effectiveness of formality outcomes. Social protection programmes produce large earnings effects but limited formality transitions. Information campaigns and labour inspections show promise, but require more rigorous evaluation. No single intervention provides a complete solution. Formalisation represents a gradual process that requires sustained effort in multiple policy domains, combining targeted interventions with broader macroeconomic policy, regulatory reform, and institutional strengthening.

The high heterogeneity in our results suggests substantial scope for understanding what makes interventions work in specific contexts. Context-specific factors, implementation quality, and complementary policies are likely to play crucial roles. Future progress depends on continued investment in rigorous impact evaluation, particularly for understudied intervention types, combined with efforts to understand mechanisms and implementation processes. The current evidence supports cautious optimism: formalisation policies work, but realistic expectations about impact magnitudes and attention to implementation remain essential for success.

References

Abel, M., Carranza, E.and Geronimo, K., and Ortega, M. (2022). Can temporary wage incentives increase formal employment? Experimental evidence from Mexico. Discussion Paper 15740, Institute of Labor Economics (IZA), Bonn.

Aedo, C. and Pizarro, M. (2004). Rentabilidad económica del programa de capacitación laboral de jóvenes Chile Joven. Report, INACAP and Mideplan, Santiago de Chile.

Alcázar, L. and Huerta, S. (2023). Mejorando la empleabilidad de mujeres urbanas vulnerables en tiempos de pandemia en el Perú: evaluación experimental del componente de capacitación virtual en un programa de empleo temporal. Documento de Investigación 127, Grupo de Análisis para el Desarrollo.

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3-4):323–338.

Alzúa, M., Cruces, G., and López, C. (2016). Long-run effects of youth training programs: Experimental evidence from Argentina. *Economic Inquiry*, 54(4):1839–1859.

Alzuá, M. L. and Brassiolo, P. (2006). The impact of training policies in Argentina: An evaluation of Proyecto Joven. Working Paper 15/06, Inter-American Development Bank, Washington DC.

Amarante, V., Burdín, G., Ferrando, M.and Manacorda, M., and Vernengo, A.; Vigorito, A. (2009). Informe final de la evaluación de impacto del PANES. Working paper, Ministerio de Desarrollo Social - Universidad de la República, Montevideo.

Aparicio, G. (2014). Does formality improve firm performance? Evidence from a quasi-experiment in Mexico. Available at https://sites.bu.edu/neudc/files/2014/10/paper$_2$83.*pdf*.

Attanasio, O., Guarín, A., Medina, C., and Meghir, C. (2017). Vocational training for disadvantaged youth in Colombia: A long-term follow-up. *American Economic Journal: Applied Economics*, 9(2):131–143.

Attanasio, O., Kugler, A., and Meghir, C. (2011). Subsidizing vocational training for disadvantaged youth in Colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 3(3):188–220.

Barrera-Osorio, F., Kugler, A., and Silliman, M. (2023). Hard and soft skills in vocational training: Experimental evidence from Colombia. *The World Bank Economic Review*, 37(3):409–436.

Barron, M. (2020). Business training programs and microenterprise formalization in Perú. *Cogent Economics & Finance*, 8:1791546.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217.

Bernal, R., Eslava, M., Meléndez, M., and Pinzón, A. (2017). Switching from payroll taxes to corporate income taxes: Firms employment and wages after the 2012 Colombian tax reform. *Economia*, 18(1):41–74.

Berniell, L. and de la Mata, D. (2017). Starting on the right track? The effects of rst job experience on short and long term labor market outcomes. Working paper 2017/26, CAF.

Bertranou, F. (2025). Informalidad laboral: evolución conceptual, medición y enfoque para la transición hacia la formalización. Nota técnica, Organización Internacional del Trabajo, Santiago de Chile. Oficina de la OIT para el Cono Sur de América Latina.

Blanchard, P., Brum, M., Carrasco, P., Parada, C., and Perazzo, I. (2025). Employment effects of a social and labour inclusion programme. *Labour Economics*, 94:102717.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2021). *Introduction to Meta-Analysis*. John Wiley & Sons.

Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1):1–32.

Bruhn, M. (2013). A tale of two species: Revisiting the effect of registration reform on informal business owners in Mexico. *Journal of Development Economics*, 103:275–283.

Bruhn, M. and McKenzie, D. (2014). Entry regulation and the formalization of microenterprises in developing countries. *The World Bank Research Observer*, 29(2):186–201.

Cabrera, J. M., Cid, A., and Bernatzky, M. (2016). The effect of one-on-one assistance on the compliance with labor regulation. a field experiment in extremely vulnerable settings. Working Paper 84639, Munich Personal RePEc Archive (MPRA).

Calderón-Mejía, V. and Marinescu, I. E. (2012). The impact of Colombia's pension and health insurance systems on informality. Discussion Paper 6439, Institute of Labor Economics (IZA).

Calero, C., Gonzalez Diez, V., Soares, Y., Kluve, J., and Corseuil, C. (2017). Can arts-based interventions enhance labor market outcomes among youth? Evidence from a randomized trial in Rio de Janeiro. *Labour Economics*, 45:131–142.

Camargo, J., Lima, L., Riva, F., and Souza, A. P. (2021). Technical education, non-cognitive skills and labor market outcomes: Experimental evidence from Brazil. *IZA Journal of Labor Economics*, 10(1):1–34.

Card, D., Ibarrarán, P., Regalia, F., Rosas-Shady, D., and Soares, Y. (2011). The labor market impacts of youth training in the Dominican Republic. *Journal of Labor Economics*, 29(2):267300.

Card, D., Kluve, J., and Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *The Economic Journal*, 120(548):452–477.

Card, D., Kluve, J., and Weber, A. (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3):894–931.

Carvalho, C. C., Corbi, R., and De-Losso, R. (2021). Estimating the employment and educational effects of vocational training: the role of school quality. Working Paper 2021-10, University of São Paulo (FEA-USP), São Paulo.

Chaparro, J., García, G., and Cardona, M. (2020). Evaluación de impacto de EMPLEANDO FUTURO. Primer bono de impacto social en América Latina. Evaluación de impacto, SIBs.CO.

Corseuil, C. H., Foguel, M. N., and Gonzaga, G. (2019). Apprenticeship as a stepping stone to better jobs: Evidence from Brazilan matched employer-employee data. *Labour Economics*, 57:177–194.

Costa, G. (2022). *Interventions on human capital formation among vulnerable populations: experimental evidence from two large-scale programs in Brazil.* PhD thesis, Doctoral dissertation. Sao Paulo Getulio Vargas Foundation.

Da Mata, D., Oliveira, R., and Silva, D. (2025). Who benefits from job training programs? Evidence from a high-dosage program in Brazil. *Journal of Development Economics*, 175:103476.

De Andrade, G. H., Bruhn, M., and McKenzie, D. (2016). A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms. *The World Bank Economic Review*, 30(1):24–54.

de Barros, R., Biderman, C., Costa, G. W., Lima, L., and Souza, A. (2023). Can at-risk youth be rescued? Experimental evidence of a nationwide youth empowerment program. Available at SSRN: https://ssrn.com/abstract=5223527.

De Mel, S., McKenzie, D., and Woodruff, C. (2013). The demand for, and consequences of, formalization among informal firms in Sri Lanka. *American Economic Journal: Applied Economics*, 5(2):122–150.

De Soto, H. (1989). *The Other Path: The Invisible Revolution in the Third World*. Harper & Row, New York.

Delajara, M., Freije-Rodriguez, S., and Soloaga, I. (2013). Evaluation of training for the unemployed in Mexico: learning by comparing methods. Working Paper 55210, Munich Personal RePEc Archive (MPRA).

Departamento Nacional de Planeación (2008). Subprograma Jóvenes en Acción: Consultoría para la evaluación de impacto del Subprograma Jóvenes en Acción. Evaluación de Políticas Públicas 9, Departamento Nacional de Planeación, Dirección de Evaluación de Políticas Públicas, Bogotá, D.C.

Díaz, J. and Rosas, D. (2016). Impact evaluation of the job youth training program Projoven. Working Paper 693, Inter-American Development Bank, Washington DC.

Díaz, J. J., Chacaltana, J., Rigolini, J., and Ruiz, C. (2018). Pathways to formalization: going beyond the formality dichotomy. Discussion Paper 11750, Institute of Labor Economics (IZA).

Díaz, J. J. and Jaramillo, M. (2006). An evaluation of the Peruvian "youth labor training program"–Projoven. Working paper 10/06, Inter-American Development Bank.

Djankov, S. (2009). The regulation of entry: A survey. *The World Bank Research Observer*, 24(2):183–203.

Doerr, A. and Novella, R. (2024). The long-term effects of job training on labor market and skills outcomes in Chile. *Labour Economics*, 91:102619.

Durand, G. (2018). Bolivia and the program to support employment: an impact evaluation of its conditional cash transfer component. Master's thesis, School of Economics and Management, Lund University.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109):629–634.

Escudero, V. (2018). Workfare programmes and their delivery system: Effectiveness of construyendo Perú. Working paper 39, International Labour Organization (ILO).

Escudero, V., Kluve, J., López Mourelo, E., and Pignatti, C. (2019). Active labour market programmes in Latin America and the Caribbean: Evidence from a meta analysis. *The Journal of Development Studies*, 55(12):2644–2661.

Fajnzylber, P., Maloney, W. F., and Montes-Rojas, G. V. (2011). Does formality improve micro-firm performance? Evidence from the Brazilan SIMPLES program. *Journal of Development Economics*, 94(2):262–276.

Feld, B. (2024). Direct and spillover effects of enforcing labor standards: Evidence from Argentina. *Journal of Human Resources*, 59(5):1503–1534.

Fields, G. S. (1990). Labour market modelling and the urban informal sector: Theory and evidence. In Fields, G. S. and Pieters, J., editors, *Employment and Development: How Work Can Lead From and Into Poverty*, pages 145–172. Oxford University Press.

Fields, G. S. (2009). Segmented labor market models in developing countries. In Kincaid, H. and Ross, D., editors, *The Oxford Handbook of Philosophy of Economics*, pages 476–508. Oxford University Press.

Fields, G. S. (2023). *The Job Ladder: Transforming Informal Work and Livelihoods in Developing Countries*. Oxford University Press.

Floridi, A., Demena, B. A., and Wagner, N. (2020). Shedding light on the shadows of informality: A meta-analysis of formalization interventions targeted at informal firms. *Labour Economics*, 67(101925).

Floridi, A., Demena, B. A., and Wagner, N. (2021). The bright side of formalization policies. meta-analysis of the benefits of policy-induced versus self-induced formalization. *Applied Economics Letters*, 28(20):1807–1812.

Galdo, J. and Chong, A. (2012). Does the quality of public-sponsored training programmes matter? Evidence from bidding processes data. *Labour Economics*, 19:970–998.

Galiani, S., Meléndez, M., and Ahumada, C. N. (2017). On the effect of the costs of operating formally: New experimental evidence. *Labour Economics*, 45:143–157.

Garsous, G., Corderi, D., and Velasco, M. (2015). Tax incentives and job creation in the tourism industry of Brazil. Working paper 644, Inter-American Development Bank.

Gasparini, L., Haimovich, F., and Olivieri, S. (2009). Labor informality bias of a poverty-alleviation program in Argentina. *Journal of Applied Economics*, 12(2):181–205.

Gómez, M. F. and González-Velosa, C. (2023). Can a pay-for-performance program help the vulnerable find jobs during a pandemic?: Experimental evidence from empleate in Colombia. Working paper 1478, Inter-American Development Bank.

Gruber, J. (1997). The incidence of payroll taxation: evidence from Chile. *Journal of Labour Economics*, 15(S3):72–101.

Havránek, T., Stanley, T., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W., Rost, K., and Aert, R. (2020). Reporting guidelines for meta-analysis in economics. *Journal of Economic Surveys*, 34.

Hernani-Limarino, W. L. and Villarroel, P. M. (2015). Capacitación laboral y empleabilidad Evidencia de Mi Primer Empleo Digno. *Revista Latinoamericana de Desarrollo Económico*, 13:35–76.

Higgings, J. P., Li, T., and Deeks, J. J. (2023). Choosing effect measure and computing estimates of effect. In *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*, chapter 6. Cochrane.

Ibarrarán, P., Kluve, J., Ripani, L., and Rosas-Shady, D. (2019). Experimental evidence on the long-term effects of a youth training program. *ILR Review*, 72(1):185–222.

Ibarrarán, P., Ripani, L., Taboada, B., Villa, J., and Garcia, B. (2014). Life skills, employability and training for disadvantaged youth: Evidence from a randomized evaluation design. *IZA Journal of Labor & Development*, 3(10).

ILO (2024). *2024-2030 Strategy for the Promotion of Formalization in Latin America and the Caribbean*. ILO Regional Office for Latin América and the Caribbean, Geneve.

Irsova, Z., Doucouliagos, H., Havranek, T., and Stanley, T. D. (2024). Meta-analysis of social science research: A practitioner's guide. *Journal of Economic Surveys*, 38(5):1547–1566.

Jessen, J. and Kluve, J. (2021). The effectiveness of interventions to reduce informality in low- and middle-income countries. *World Development*, 138(105256).

Katz, L. F. (1996). Wage subsidies for the disadvantaged. Working paper 5679, National Bureau of Economic Research.

Kluve, J. (2010). The effectiveness of European active labor market programs. *Labour Economics*, 17(6):904–918.

Kluve, J., Puerto, S., Robalino, D., Romero, J., Rother, F., Stöterau, J., Weidenkaff, F., and Witte, M. (2019). Do youth employment programs improve labor market outcomes? A quantitative review. *World Development*, 114:237–253.

Kugler, A., Kugler, M., and Prada, L. O. H. (2017). Do payroll tax breaks stimulate formality? Evidence from Colombias reform. Working paper 23308, National Bureau of Economic Research.

Kugler, A., Kugler, M., Saavedra, J. E., and Herrera-Prada, L. O. (2022). Long-term educational consequences of vocational training in colombia: Impacts on young trainees and their relatives. *Journal of Human Resources*, 57(1):178–216.

Layard, P. R. G., Layard, R., Nickell, S. J., and Jackman, R. (2005). *Unemployment: macroeconomic performance and the labour market.* Oxford University Press.

Lenz, A.-K. (2017). *Studies on entrepreneurship and formalization in Brazil.* PhD thesis, The School of Public Business and Administration (EBAPE)-Getulio Vargas Foundation, Rio de Janeiro.

Lerman, R. (2019). Do firms benefit from apprenticeship investments? *IZA World of Labor*.

Levy, S. (2008). *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico.* Brookings Institution Press, Washington, D.C.

Llerena Pinto, C., Llerena Pinto, A., and Llerena Pinto, F. (2013). Evaluación de impacto del programa mi primer empleo en el Ecuador. Evaluación de impacto, ECONÓMICA CIC Centro de Investigación, Analítica y Desarrollo de Sistemas Informáticos para las Ciencias Sociales y Administración.

López Mourelo, E. and Escudero, V. (2017). Effectiveness of active labour market tools in conditional cash tranfers programmes:evidence for Argentina. *World Development*, 94:422–447.

Macroconsult, S. A. (2012). Evaluación de impacto del programa construyendo Perú. Working paper, Ministerio de Economía y Finanzas. Unidad de Coordinación de Préstamos Sectoriales (UCPS), Lima.

Madalozzo, R. C. and Bortoluzzo, A. B. (2011). The impact of tax exemptions on labor registration: The case of Brazilan domestic workers. Working Papers 232/2011, Instituto de Ensino e Pesquisa (Insper).

Maloney, W. F. (2004). Informality revisited. *World Development*, 32(7):1159–1178.

Marimon, R. and Zilibotti, F. (1999). Unemployment vs. mismatch of talents: Reconsidering unemployment benefits. *The Economic Journal*, 109(455):266–291.

Mera, M., Camisassa, J., Baliña, J., and Méndez, E. L. (2023). Evaluación de cursos para personas emprendedoras en la ciudad autónoma de Buenos Aires. Documento de Trabajo 217, CIPPEC.

Monteiro, J. C. and Assunção, J. J. (2012). Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilan microenterprises. *Journal of Development Economics*, 99(1):105–115.

Morales, L. F. and Medina, C. (2017). Assessing the effect of payroll taxes on formal employment: The case of the 2012 tax reform in Colombia. *Economía*, 18(1):75–124.

Morales-González, K., Ávila, W. M. d., and Cruz-Almanza, S. d. l. (2019). Evaluación del servicio público de empleo: sus efectos en la inserción laboral formal en el área Metropolitana de Barranquilla, Colombia. *Lecturas de Economía*, (91):211–239.

Moser, C. O. N. (1978). Informal sector or petty commodity production: Dualism or independence in urban development. *World Development*, 6(9-10):1041–1064.

Nogueira Puentes, R. (2018). Las políticas activas del mercado de trabajo: revisión de las experiencias de evaluación y una estimación del impacto del programa "Uruguay Trabaja". Master's thesis, Udelar. FCEA.

Novella, R., Rucci, G., Vazquez, C., and Kaplan, D. (2017). Training vouchers and labour market outcomes in Chile. *Labour*, 32(2):243–260.

Novella, R. and Valencia, H. (2022). Active labor market policies in a context of high informality: The effect of pae in Bolivia. *The Journal of Development Studies*, 58(12):2583–2603.

Ohaco, M. (2023). Registración del servicio doméstico: Evaluación de impacto de una política de difusión dirigida a empleadores. *Economica*, 69(1).

OConnell, S. D., Mation, L. F., Basto, J. B. T., and Dutz, M. A. (2019). Making public job training work: Evidence from decentralized targeting using firm input. Available at https://www.stephenoconnell.org/files/OM_MakingJobTrainingWork.pdf.

Pereda, P., Narita, R., Rocha, F., Diaz, M., Almeida, E., Borges, B. P. d. S., and Matsunaga, L. (2022). Formalization of female micro-entrepreneurs in Brazil. Available at SSRN: https://ssrn.com/abstract=4000220.

Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., Mason, A. D., and Saavedra-Chanduvi, J. (2007). *Informality: Exit and Exclusion.* World Bank, Washington, D.C.

Peña, H. E. (2010). *Impact Evaluation of a Job-Training Programme for Disadvantaged Youths: The Case of Projoven.* PhD thesis, Maastricht University, Maastricht, The Netherlands.

Picchio, M. (2020). Meta-analysis. In Zimmermann, K. F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–29. Springer International Publishing.

Pignatti, C. (2016). Do public employment services improve employment outcomes? Evidence from Colombia. Working Paper 10, International Labour Organization ILO, Geneva.

Pignatti, C. (2020). Compliance with Labour Legislation in Informal Labour Markets. Working paper. Available at SSRN: https://ssrn.com/abstract=4995256.

Piza, C. (2018). Out of the shadows? Revisiting the impact of the Brazilan SIMPLES program on firms formalization rates. *Journal of Development Economics*, 134:125–132.

Ponczek, V. and Ulyssea, G. (2022). Enforcement of labour regulation and the labour market effects of trade: Evidence from Brazil. *The Economic Journal*, 132(641):361–390.

Portes, A., Castells, M., and Benton, L. A., editors (1989). *The Informal Economy: Studies in Advanced and Less Developed Countries*. Johns Hopkins University Press, Baltimore, MD.

Ranis, G. and Stewart, F. (1999). V-goods and the role of the urban informal sector in development. *Economic Development and Cultural Change*, 47(2):259–288.

Reis, M. (2015). Vocational training and labor market outcomes in Brazil. *The BE Journal of Economic Analysis & Policy*, 15(1):377–405.

Rocha, R., Ulyssea, G., and Rachter, L. (2018). Do lower taxes reduce informality? Evidence from Brazil. *Journal of Development Economics*, 134:28–49.

Ronconi, L. (2010). Enforcement and compliance with labor regulations in Argentina. *ILR Review*, 63(4):719–736.

Ronconi, L. and Colina, J. (2011). Simplification of labor registration in Argentina: Achievements and pending issues. Working Paper 277, Inter-American Development Bank.

Ronconi, L., Sanguinetti, J., Fachelli, S., Casazza, V., and Franceschelli, I. (2006). Poverty and employability effects of workfare programs in Argentina. Available at SSRN: https://ssrn.com/abstract=3173205.

Rosas-Shady, D. (2006). Impact evaluation of Projoven youth labor training program in Perú. Working Paper 04/06, Inter-American Development Bank.

Salazar-Xirinachs, J. M. and Chacaltana, J. (2018). *Políticas de Formalización en América Latina: Avances y Desafíos. Lima: OIT, Oficina Regional para América Latina y el Caribe*. ILO Regional Office for Latin América and the Caribbean, Lima.

Santa María, M., Olivera, M., Acosta, P., Vasquez, T., and Rodriguez, A. (2009). Evaluación de impacto programa jóvenes con futuro. Working Paper 44, Fedesarrollo–Centro de Investigación Económica y Social.

Schmid, G., Speckesser, S., and Hilbert, C. (2001). Does active labour market policy matter? An aggregate impact analysis for germany. In *Labour market policy and unemployment: An Evaluation of Active Measures in France, Germany, the Netherlands, Spain and Sweden*, pages 77–114. Edward Elgar Cheltenham.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, 70(1):103–127.

Stanley, T. D. and Doucouliagos, H. (2011). *Meta-Regression Analysis in Economics and Business*. Routledge.

Stanley, T. D. and Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1):60–78.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from Brazil. *American Economic Review*, 108(8):2015–2047.

Valdivia, M. (2015). Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Perú. *Journal of Development Economics*, 113:33–51.

Van Gameren, E. (2010). Evaluación de impacto del programa de apoyo al empleo. Evaluación de impacto, Secretaría de Trabajo y Previsión Social (STPS), Ciudad de México.

Viollaz, M. (2018a). Are labour inspections effective when labour regulations vary according to the size of the firm? Evidence from Perú. *ILR Review*, 157(2):213–242.

Viollaz, M. (2018b). Enforcement of labour market regulations: heterogeneous compliance and adjustment across gender. *IZA Journal of Labour Policy*, 7(1):1–28.

Zucco, C., Lenz, A.-K., Goldszmidt, R., and Valdivia, M. (2020). Face-to-face vs. virtual assistance to entrepreneurs: Evidence from a field experiment in Brazil. *Economics Letters*, 188(108922).

Zucco, C., Lenz, A.-K., Goldszmidt, R., and Valdivia, M. (2023). Do political preferences affect policy learning and uptake? Evidence from a field experiment with informal entrepreneurs. *The Journal of Politics*, 85(3):1003–1016.

Appendices

Appendix A

Sample Selection and Construction Process

Paper-level screening first identified 45 studies that did not meet our fundamental inclusion criteria, removing 150 estimates. The largest exclusion category was social protection programmes without activation components (67 estimates), followed by studies that did not present estimations for probability of being formal (38 estimates), studies that did not evaluate formalisation policy effectiveness as a primary objective (30 estimates), and studies employing non-experimental identification methods (15 estimates). This initial screening left 112 relevant studies containing 1,810 estimates.

Estimate-level refinement applied granular selection criteria to these remaining estimates, excluding 1,302 for various methodological and definitional reasons. The largest exclusion categories reflected our analytical focus: 247 estimates related to cohort or call variations not incorporated, 247 estimates for outcome variables outside our scope, 180 estimates for employment outcomes we decided not to analyse, and 166 estimates where only combined treatments were available when disaggregated effects were required. Methodological exclusions included 110 estimates lacking standard errors, 95 estimates representing duplicate formality definitions, and various other technical exclusions totalling 267 estimates.

Data construction adjustments subsequently created 19 additional pooled gender estimates for five studies that reported separate male and female effects but lacked overall estimates. This refinement process resulted in 527 estimates from 79 studies, ensuring methodological consistency while capturing the full breadth of formalisation policy evaluations in Latin American.

Data Construction and Technical Adjustments

We made several technical adjustments during data construction to ensure consistency and avoid double-counting across studies. For studies presenting multiple treatment variations targeting the same outcome, we prioritised aggregate treatment effects when available. Nine studies reported both individual and aggregate treatment estimates (such as different types of training programmes or sensitisation mechanisms), and we consistently selected the aggregate estimate to maintain comparability while avoiding overrepresentation of individual studies. Six studies did not provide aggregate estimates for multiple treatments and therefore contribute multiple estimates per outcome category. These include studies examining different training modalities or varied sensitisation approaches within the same evaluation framework.

Table A.1: Study and estimate selection process

Selection Stage	**Studies**	**Estimates**
Initial search results	152	1,960
Paper-level exclusions:	*-45*	*-150*
Does not evaluate formalization policy effectiveness	-13	
Does not present estimations for probability of being formal	-11	
Non-experimental identification	-13	
Social protection without activation component	-8	
Remaining after paper exclusions	**107**	**1,810**
Estimate-level exclusions from relevant papers:		*-1,302*
Cohort/call not incorporated		-247
Outcome variable not considered		-247
Employment outcome not considered		-180
Only combined treatment is considered		-166
No standard errors		-110
Duplicate formality (definition not considered)		-95
Age categories not used		-57
Informal employment outcome not considered		-43
Duplicate earnings (definition not considered)		-42
Estimates not considered		-29
Only pooled definition of firm considered		-25
Matching method not selected		-18
Paper analyzed in new version		-12
Subgroup not considered		-12
Only pooled definition of worker considered		-8
Estimate with non-experimental identification		-6
Control group not considered		-3
Duplicate employment (definition not considered)		-2
Estimate-level inclusions from relevant papers:		*19*
Pooled gender estimates created		19
Final analytical sample	**79**	**527**

Notes: The table shows the complete selection process from 1,960 initial estimates to 527 final estimates. Data construction adjustments created pooled gender estimates for studies reporting only gender-disaggregated results. Paper-level exclusions removed entire studies that did not meet inclusion criteria. Estimate-level exclusions removed specific estimates within retained studies for methodological and definitional reasons. When all estimates from a study are not considered, the study is not considered.

One study, Pignatti (2016), presented estimates for four different control groups. We excluded the "employers" control group as it was not comparable with other labour intermediation programmes in our sample, retaining only the three control groups representing job seekers.

Three studies contributed estimates for both firm-level and worker-level outcomes within the same intervention, reflecting the dual nature of formalisation processes. These studies examined both employer registration and employee social security coverage, providing complementary perspectives on policy effectiveness. We retained both types of estimates as they measure distinct aspects of formalisation.

Our sample includes eleven programmes that have been evaluated by multiple independent studies, representing 27 total evaluations. These include large-scale national programmes such as Colombia's *Jóvenes en Acción* (evaluated by four separate studies) and Peru's *Projoven programme* (assessed in five different evaluations). This overlap strengthens our evidence base by providing multiple perspectives on the same interventions while allowing us to examine consistency across different evaluation approaches and time periods.

Handling Interaction Terms and Gender Pooling

We implemented two additional data construction procedures to maximise the analytical value of available estimates. First, six studies reported results using interaction terms rather than separate subgroup estimates. Following the approach recommended by Irsova et al. (2024), we calculated the total treatment effects for specific demographic groups by computing partial derivatives of the treatment effect. For gender interactions, for instance, we derived effects as $\beta_{male} = \beta_{treatment}$ and $\beta_{female} = \beta_{treatment} + \beta_{interaction}$, with standard errors adjusted using the delta method: $SE_{female} = \sqrt{SE^2_{treatment} + SE^2_{interaction}}$, assuming zero covariance between coefficients due to the typical unavailability of covariance estimates (Irsova et al., 2024; **?**). This approach allowed us to derive gender-specific and age-specific effects from studies by Alzuá and Brassiolo (2006), Barrera-Osorio et al. (2023), López Mourelo and Escudero (2017), and Doerr and Novella (2024).

Second, we created pooled gender estimates for study-outcome combinations that reported separate male and female effects but lacked an overall estimate. Following Borenstein et al. (2021), we employed inverse-variance weighting to combine mutually exclusive subgroups within studies:

$$\hat{\beta}_{pooled} = \frac{w_m \hat{\beta}_m + w_f \hat{\beta}_f}{w_m + w_f} \tag{7}$$

where weights are defined as $w = 1/SE^2$, and pooled standard errors as $SE_{pooled} = 1/\sqrt{w_m + w_f}$. This inverse-variance approach is preferred over simple sample-size weighting as it better accounts for variance driven by study design and outcome measurement precision (Borenstein et al., 2021). This procedure generated 19 additional estimates in five studies, allowing us to include studies with gender-disaggregated results in analyses that examine the overall effectiveness of the programme.

Sample and Control Group Observations for SMD Construction

The calculation of standardised mean differences requires precise treatment and control group sample sizes for each estimate. However, most studies in our sample report regression-based results where the analytical sample may differ substantially from the initial design sample due to attrition, missing covariates, or subgroup-specific analyses. We therefore constructed treatment and control group sizes based on the actual regression samples rather than design samples to ensure accurate weighting in our meta-analytical models.

For studies reporting overall treatment and control group distributions, we applied these proportions to subgroup analyses when demographic breakdowns were not explicitly provided. When studies presented descriptive statistics showing the demographic composition of treatment and control groups separately (e.g., 40% of the treatment group and 45% of the control group were women), we maintained these differential distributions in our sample size calculations. In cases where only general demographic proportions were reported (e.g., 70% of the total sample were women), we assumed equal representation between treatment and control groups while acknowledging that this can introduce measurement error.

This approach was particularly relevant for quasi-experimental studies, where the concept of distinct treatment and control groups is less straightforward than in randomised trials. In difference-in-differences, regression discontinuity, and instrumental variable designs, we constructed treatment and control samples based on the analytical framework of each study, typically defining the control group as observations contributing to the baseline comparison in the regression specification. While this methodology introduces some imprecision, particularly for subgroup analyses, we considered it preferable to using aggregate sample sizes that would not reflect the actual precision of demographic-specific estimates.

Our approach ensures that the weights assigned to each estimate in our meta-analysis correspond to the statistical precision of the underlying regression rather than potentially misleading design sample sizes, thereby improving the accuracy of our standardised mean difference calculations and subsequent meta-analytical results.

Appendix B

This appendix provides supplementary analyses referenced in Section 4. We present outcome-specific publication bias diagnostics, a preliminary sign-and-significance analysis, descriptive heterogeneity patterns, and robustness checks for the meta-regression results.

Outcome Variable Definitions

Table B.1 presents the complete list of formality, earnings, and hours worked definitions used across all studies in our sample, along with programme types analysed. This demonstrates the heterogeneity in outcome measurement that motivates our use of standardised mean differences.

Table B.1: Outcome Definitions Across Studies

Author (Year)	**Programme**	**Formality**	**Hours**	**Earnings**
Abel et al. (2022)	HS	(E) Social security registration	–	Monthly - formal, level
Aedo and Pizarro (2004)	T + A	(E) Work contract	–	Monthly, level
Alcázar and Huerta (2023)	T	(E) Formal salaried worker	–	Monthly, level
Alzuá and Brassiolo (2006)	T + A	(E) Employer pays social security	–	Monthly, level
Alzúa et al. (2016)	T + A	(E) SIPA registration	–	Monthly - formal, level
Amarante et al. (2009)	SP	(E) Social security access	Weekly, level	Monthly, log
Aparicio (2014)*	S	(F) Government registration	–	–
Attanasio et al. (2011)*	T + A	(E) Health, pension, injury benefits	Weekly, level	Monthly - formal, level
Attanasio et al. (2017)	T + A	(E) PILA registration	–	Monthly - formal, level
Barrera-Osorio et al. (2023)	T	(E) Formal contract	Weekly, level	Monthly, level
Barron (2020)	T	(F) Tax ID number	–	–
Bernal et al. (2017)	PTR	–	–	Monthly average wage, log (firm level)
Berniell and de la Mata (2017)	HS	(E) SIPA registration	–	Monthly - formal, level
Blanchard et al. (2025)*	SP	(E) Employer has paid licence	Weekly, level	–
Bruhn (2013)	S	(F) Government registration / (E) Signed contract	–	–
Cabrera et al. (2016)	SE	(F) Legal work permit	–	–
Calderón-Mejía and Marinescu (2012)*	S	(E) Health and pension contributions	–	–
Calero et al. (2017)	T	(E) Official work card	Weekly, level	Monthly, level
Camargo et al. (2021)	T	(E) Official work card	–	Monthly, level

Continued on next page

Table B.1 – continued from previous page

Author (Year)	Programme	Formality	Hours	Earnings
Card et al. (2011)	T + A	(E) Employer health insurance	Weekly, level	Monthly, level
Carvalho et al. (2021)	T	(E) Formal sector worker	–	–
Chaparro et al. (2020)	T + LI	(E) PILA registration	–	–
Corseuil et al. (2019)	T + A	(E) RAIS registration	–	–
Costa (2022)	T	(E) RAIS registration	–	–
Da Mata et al. (2025)	T + A	(E) RAIS registration	–	Monthly - formal, level
De Andrade et al. (2016)	SE/LIN	(F) Government registration	–	–
de Barros et al. (2023)	T	(E) RAIS registration	–	–
Delajara et al. (2013)	T + LI	(E) IMSS registration	–	Monthly - formal, level
Departamento Nacional de Planeación (2008)*	T + A	(E) Formal sector employment	Weekly, level	Monthly, log
Díaz et al. (2018)*	PTR + S	(F) Municipal licenses / (E) Electronic Payroll Registry	–	–
Díaz and Jaramillo (2006)	T + A	(E) Formal contract/insurance/social security	Weekly, level	Monthly, level
Díaz and Rosas (2016)*	T + A	(E) Retirement pension	Weekly, log	Monthly, log
Doerr and Novella (2024)	T	(E) Social security/unemployment insurance	Hours worked	Monthly, level
Durand (2018)	A	(E) Healthcare access	–	Monthly, log
Escudero (2018)	SP	(E) Pension system affiliation	Weekly, log	–
Fajnzylber et al. (2011)	PTR + S	(F) Tax registration	–	–
Feld (2024)	SE + LIN	(E) Pension contributions	Weekly, log	Monthly, log
Galdo and Chong (2012)	T + A	(E) Formal contract	–	Monthly, level

Continued on next page

Table B.1 – continued from previous page

Author (Year)	Programme	Formality	Hours	Earnings
Galiani et al. (2017)	SE + PTR	(F) License	–	–
Garsous et al. (2015)	PTR	–	–	Monthly average wage, log (municipal)
Gasparini et al. (2009)	SP	(E) SIJP registration	–	–
Gómez and González-Velosa (2023)	T + LI	(E) Social security contributions	–	Monthly - formal, level
Hernani-Limarino and Villarroel (2015)	T + A	(E) Social protection compliance	–	Monthly, level
Ibarrarán et al. (2019)*	T + A	(E) Employer health insurance	–	Monthly, log
Ibarrarán et al. (2014)*	T + A	(E) Employer health insurance	–	Monthly, log
Kugler et al. (2017)*	PTR	(E) Health and pension contribution	–	–
Kugler et al. (2022)	T + A	(E) Health/pension coverage	–	–
Lenz (2017)	SE	(F) MEI registration	–	–
Llerena Pinto et al. (2013)	A	(E) Formal sector employment	–	–
López Mourelo and Escudero (2017)	T + LI	(E) Formal sector employment	Weekly, level	–
Macroconsult (2012)	SP	(E) Legal entity/contract	Weekly, level	Monthly, level
Madalozzo and Bortoluzzo (2011)	PTR	(F) Registered employment	–	–
Mera et al. (2023)	T	(F) Tax authority registration	–	–
Monteiro and Assunção (2012)	PTR + S	(F) Official license	–	–
Morales and Medina (2017)	PTR	–	–	Monthly average wage (firm level)
Morales-González et al. (2019)	LI	(E) Social security benefits	–	–
Nogueira Puentes (2018)	SP	(E) BPS contributions	–	–
Novella and Valencia (2022)	HS + LI	(E) Social security contributions	–	Monthly, log

Continued on next page

Table B.1 – continued from previous page

Author (Year)	Programme	Formality	Hours	Earnings
Novella et al. (2017)	T	(E) Unemployment insurance registration	–	Monthly, log
OConnell et al. (2019)	T	(E) RAIS registration	–	Monthly, level
Ohaco (2023)	SE	(F) Domestic worker registration	–	–
Peña (2010)	T + A	(E) Youth training contract	–	Monthly, log
Pereda et al. (2022)	PTR + S	(F) MEI registration	–	–
Pignatti (2020)*	LIN	(E) Health social security	Weekly, level	Monthly, level
Pignatti (2016)	LI	(E) Social security coverage	–	–
Piza (2018)	PTR + S	(F) Municipal/state license	–	–
Reis (2015)	T	(E) Formal sector employment	–	Monthly, level
Rocha et al. (2018)	PTR + S	(F) RAIS registration	–	–
Ronconi and Colina (2011)	S	(E) Pension contributions	–	–
Ronconi et al. (2006)	SP	(E) Social security coverage	–	Monthly, level
Rosas-Shady (2006)	T + A	(E) Formal sector employment	Weekly, level	Monthly, level
Santa María et al. (2009)	T + A	(E) Social security/pension affiliation	–	Monthly, log
Ponczek and Ulyssea (2022)	LIN	–	–	Monthly average, level (micro-region)
Valdivia (2015)	T	(F) Tax registration	–	–
Van Gameren (2010)	LI/T/A	(E) Labour benefits	–	Monthly, level
Viollaz (2018a,b)*	LIN	(E) Employment benefits	–	–

Zucco et al. (2020, 2023)	SE	(F) MEI formalization	–	–

Notes: (E) = Employee formality; (F) = Firm formality. Programme type abbreviations: HS = Hiring subsidies; T = Training; A = Apprenticeship or Internship; SP = Social protection; S = Simplification; SE = Sensitization; PTR = Payroll tax reduction; LI = Labour intermediation; LIN = Labour inspections. Dashes (–) indicate outcome not measured in the study. Preferred estimate was selected based on priority in studies with multiple definitions, marked with a * following - 1) registration with authorities, 2) employment benefits, 3) pension benefits, 4) social security benefits, 5) signed contract.

Outcome-Specific Publication Bias Analysis

We complement the pooled publication bias analysis in Section 4.3 with outcome-specific assessments. Table B.2 presents Egger test results for each outcome category. All outcomes show evidence of publication bias, with particularly strong evidence for monthly earnings (coefficient: 11.237, $p < 0.01$).

Table B.2: Summary of Egger Test Results by Outcome

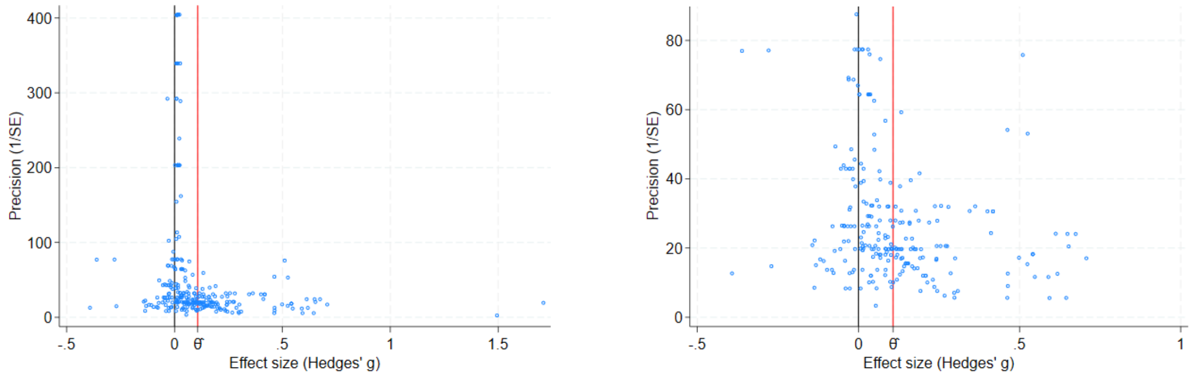
Outcome	Test	Coefficient	Interpretation
All Outcomes	RE	4.921 ***	Pub. Bias
		(0.413)	
	FE	2.412 ***	Pub. Bias
		(0.055)	
Formality	RE	1.766 ***	Pub. Bias
		(0.349)	
	FE	2.332 ***	Pub. Bias
		(0.071)	
Hours Worked	RE	1.396 **	Pub. Bias
		(0.576)	
	FE	1.651 ***	Pub. Bias
		(0.397)	
Monthly Earnings	RE	11.237 ***	Pub. Bias
		(1.109)	
	FE	3.186 ***	Pub. Bias
		(0.100)	

Notes: Egger's test is a linear regression of the intervention effect estimates on their standard errors weighted by their inverse variance. RE refers to Random Effects, FE refers to Fixed Effects. Pub. Bias refers to Publication Bias. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figures B.1 to B.3 display funnel plots for each outcome category. Visual inspection confirms the asymmetry patterns identified in the pooled analysis, with studies concentrated in regions indicating positive effects with low precision across all outcome types.

Tables B.3 to B.5 present detailed FAT-PET-PEESE bias correction estimates for each outcome category. Bias correction substantially reduces effect sizes across all outcomes, with the largest reductions for earnings outcomes. For formality outcomes, the PEESE-corrected estimate is 0.087 ($p < 0.01$). Hours worked shows a PEESE-corrected estimate of 0.022, though this does not reach statistical significance. Monthly earnings displays the strongest bias, with PEESE correction yielding -0.097, suggesting publication selection may entirely account for observed positive effects in this outcome category.

Figure B.1: Funnel plots for formality outcomes

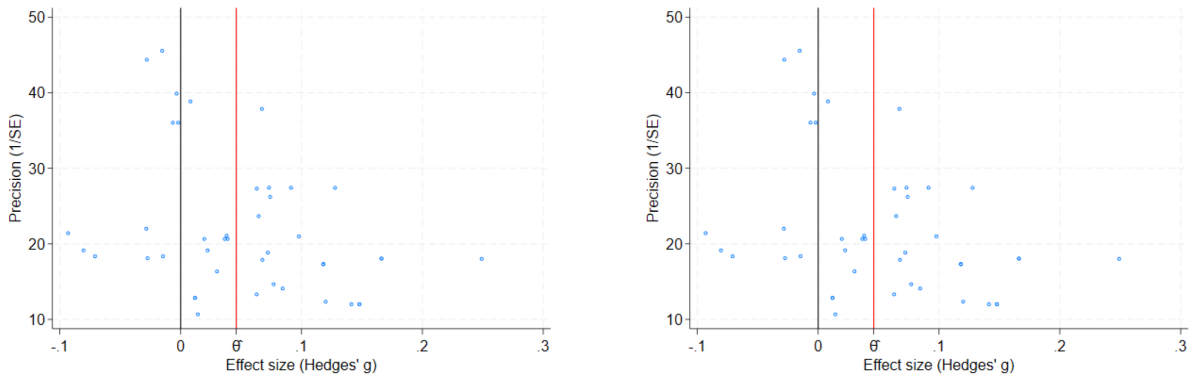


(a) With pooled estimate

(b) Outliers excluded

Notes: θ (red dashed lines) show pooled random-effects estimates. N=269.

Figure B.2: Funnel plots for hours worked outcomes



(a) With pooled estimate

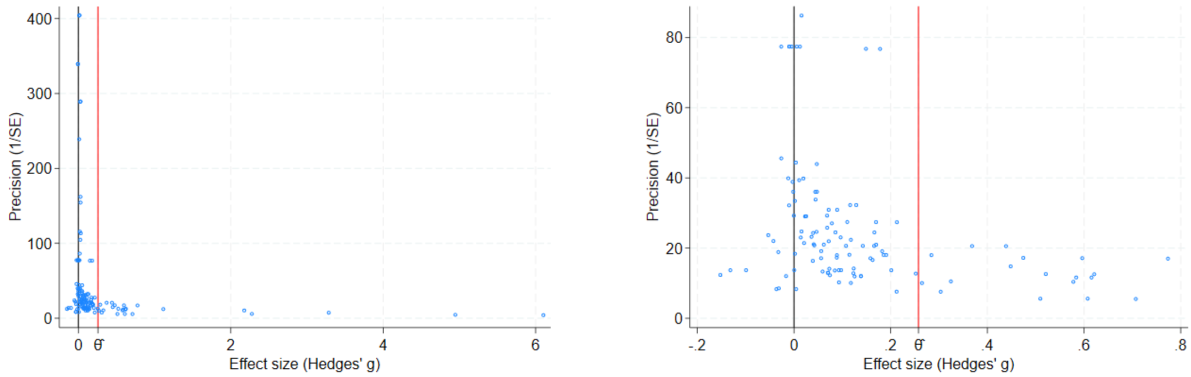
(b) Outliers excluded

Notes: θ (red dashed lines) show pooled random-effects estimates. N=44.

Sign and Significance Analysis

Before presenting our main standardised mean difference meta-analysis in Section 4, we conducted a preliminary sign and significance analysis following Escudero et al. (2019), Card et al. (2018), and Jessen and Kluve (2021). This complementary method examines patterns in the direction and statistical significance of treatment effects without requiring complete sample size information. The approach requires only the reported sign and statistical significance of each estimate, making it applicable to the full sample of 527 estimates. This analysis provides a descriptive overview of patterns in the literature but has important limitations: it discards information about effect magnitudes, treats all statistically significant positive effects equally regardless of size, and does not correct for publication bias. These limitations motivate our primary focus on standardised mean difference meta-analysis with explicit bias correction in the main text.

Figure B.3: Funnel plots for monthly earnings outcomes



(a) With pooled estimate

(b) Outliers excluded

Notes: θ (red dashed lines) show pooled random-effects estimates. N=131.

For each estimate, we extracted information on effect direction (positive or negative) and statistical significance based on conventional thresholds (typically $p < 0.05$ or $p < 0.10$). Our sample presents 287 (54.5%) statistically significant positive, 213 (40.4%) statistically insignificant, and 27 (5.1%) statistically significant negative effects. Given the low proportion of significant negative effects, we follow Escudero et al. (2019) in creating a binary outcome variable coded as 1 if the estimate is statistically significant positive and 0 otherwise.

Table B.6 presents results from linear probability models examining the relationship between programme effectiveness (binary indicator for positive and significant effects) and study characteristics. Explanatory variables are added sequentially across eight specifications, with standard errors clustered at the study level. Models 7 to 8 include country fixed effects.

The analysis reveals that most study characteristics are not significantly correlated with the likelihood of obtaining positive significant results. Three patterns emerge with statistical support. First, youth-targeted programmes show robust positive associations in specifications with country fixed effects (Models 7 to 8), with coefficients of 0.176 ($p < 0.05$) and 0.181 ($p < 0.05$). However, youth coefficients are small and statistically insignificant in specifications without country controls, indicating the relationship is sensitive to model specification. Second, experimental evaluation designs show consistently large negative associations across Models 6 to 8, with coefficients ranging from -0.301 to -0.349 (all $p < 0.01$). This likely reflects stricter methodological standards in experimental studies rather than worse policy performance. Third, information campaigns show positive coefficients in Models 6 to 8, ranging from 0.505 to 0.543 (all $p < 0.10$).

Most other factors show no significant relationships. Programme type differences between ALMPs, social protection, payroll tax reductions, and simplification measures

Table B.3: FAT-PET-PEESE Analysis: Formality Outcomes

Model	Constant (PET)	SE Coeff (FAT)	Interpretation
WLS (Stanley 2005)	0.006	2.332***	Pub. Bias
	(0.003)	(0.389)	
OLS Clustered	0.018	2.024***	Pub. Bias
	(0.035)	(0.597)	
Fixed Effects	-0.058	3.584***	Pub. Bias
	(0.046)	(0.945)	
IV (Sample Size)	0.032	1.741**	Pub. Bias
	(0.034)	(0.752)	
Weighted (1/SE)	0.007	2.234***	Pub. Bias
	(0.009)	(0.483)	
PEESE Analysis			
PEESE Clustered	0.087	6.871***	
	(0.025)	(1.493)	

Notes: N=444 from 59 studies. PET (Precision Effect Test) provides bias-corrected effects through the constant term. FAT (Funnel Asymmetry Test) tests the SE coefficient for publication bias. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

are statistically insignificant across specifications. Formality and earnings outcomes show positive but insignificant coefficients relative to hours worked. Both effect duration (short versus medium versus long-term measurement timing) and programme duration show inconsistent patterns with no statistical significance. Women-focused estimates and gender-pooled estimates show small negative coefficients that rarely reach significance. GDP growth shows a positive but insignificant association (0.014).

The inclusion of country fixed effects in Models 7 to 8 substantially increases the ^{2} from 0.130 to 0.223 to 0.230, indicating that country-level factors explain considerable variation in results beyond observable characteristics.

Table B.7 presents results using study-level weights where each study receives equal weight regardless of the number of estimates it contributes. Results are qualitatively similar to the main specification.

This preliminary analysis identifies experimental designs as systematically reporting more conservative findings and shows that country-level factors explain substantial variation. However, the approach discards information about effect magnitudes and does not correct for publication bias. These limitations motivate our main standardised mean difference meta-analysis presented in Section 4.

Table B.4: FAT-PET-PEESE Analysis: Hours Worked Outcomes

Model	Constant (PET)	SE Coeff (FAT)	Interpretation
WLS (Stanley 2005)	-0.031	1.651***	Pub. Bias
	(0.023)	(0.546)	
OLS Clustered	-0.011	1.232*	No bias
	(0.041)	(0.723)	
Fixed Effects	0.010	0.836	No bias
	(0.056)	(1.062)	
IV (Sample Size)	-0.012	1.255*	No bias
	(0.036)	(0.700)	
Weighted (1/SE)	-0.021	1.417**	Pub. Bias
	(0.032)	(0.595)	
PEESE Analysis			
PEESE Clustered	0.022	10.178	
	(0.030)	(7.378)	

Notes: N=444 from 59 studies. PET (Precision Effect Test) provides bias-corrected effects through the constant term. FAT (Funnel Asymmetry Test) tests the SE coefficient for publication bias. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Heterogeneity by Programme Type and Implementation Features

The high heterogeneity in the pooled analysis ($I^2 = 99.9\%$) indicates substantial variation in treatment effects across studies. We examine effect sizes across key characteristics to understand sources of this variation. The results in this section are simple weighted averages within each subgroup. They are not meta-regression estimates, which appear in Section 4.4 and control for multiple characteristics simultaneously while correcting for publication bias.

Programme Type Heterogeneity Figure B.4 presents effect sizes for each programme category with sufficient observations. ALMPs show the most consistent effectiveness. With 366 observations, random effects meta-analysis shows an effect size of 0.112 ($p < 0.01$). Between-study heterogeneity is high ($I^2 = 99.6\%$), but the estimate is statistically significant, indicating robust positive effects across diverse implementation contexts. All three estimation methods converge around similar point estimates.

Social protection programmes with activation components show the largest point estimates (RE: 0.394, $p < 0.05$). Precision is low, with wide confidence intervals reflecting substantial variation across seven contributing studies. Heterogeneity is high ($I^2 = 99.98\%$). The fixed effects estimate approaches zero, underscoring sensitivity to assumptions about between-study variation.

Information campaigns show modest but consistent effects (RE: 0.015, $p < 0.01$).

Table B.5: FAT-PET-PEESE Analysis: Monthly Earnings Outcomes

Model	Constant (PET)	SE Coeff (FAT)	Interpretation
WLS (Stanley 2005)	-0.002	3.186 ***	Pub. Bias
	(0.005)	(0.482)	
OLS Clustered	-0.410	11.850 *	No bias
	(0.263)	(6.315)	
Fixed Effects	-0.634	15.776 **	Pub. Bias
	(0.327)	(5.726)	
IV (Sample Size)	-0.025	5.114 **	Pub. Bias
	(0.027)	(2.606)	
Weighted (1/SE)	-0.051	5.564 *	No bias
	(0.047)	(2.920)	
PEESE Analysis			
PEESE Clustered	-0.097	70.313 ***	
	(0.051)	(24.673)	

Notes: N=444 from 59 studies. PET (Precision Effect Test) provides bias-corrected effects through the constant term. FAT (Funnel Asymmetry Test) tests the SE coefficient for publication bias. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Heterogeneity is relatively low ($I^2 = 33.2\%$), suggesting information-based approaches produce small but predictable improvements across contexts. All three estimation methods cluster tightly together.

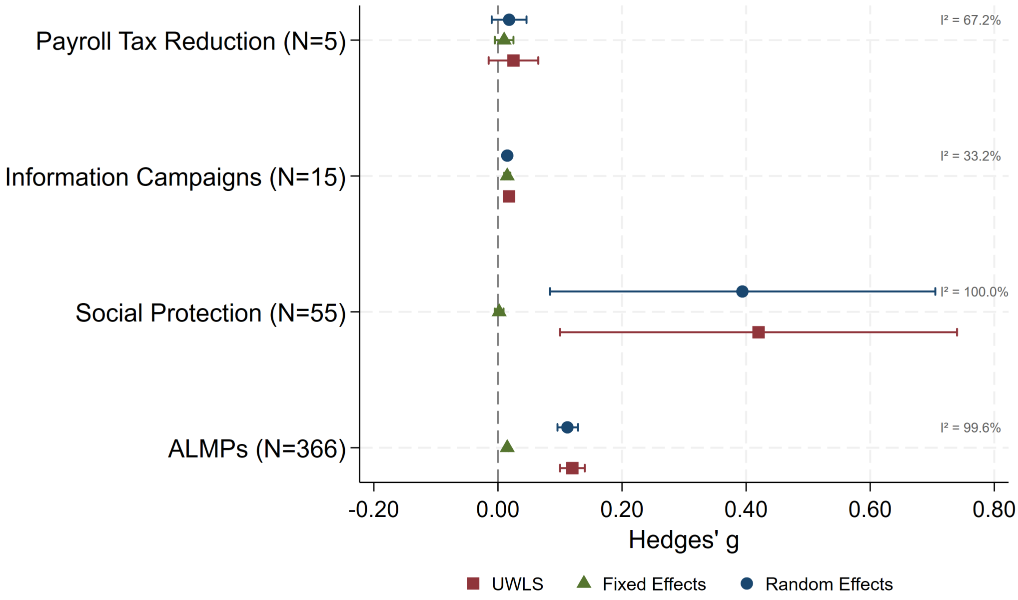
Evidence on payroll tax reductions remains limited. With only five observations from four studies, the random effects estimate (0.018) does not reach statistical significance. The single observation for labour inspections and the absence of simplification estimates reflect evidence gaps.

Demographic Targeting Figure B.5 summarises results across participant characteristics. Male participants show larger random effects estimates (0.180, $p < 0.01$) compared to females (0.110, $p < 0.01$), though confidence intervals overlap substantially. Both groups display I^2 values above 99%, indicating high heterogeneity. The pooled gender estimate (0.143, $p < 0.01$) falls between the gender-specific estimates.

Youth-focused interventions using strict ILO definitions (<30 years) show effect sizes of 0.122 ($p < 0.01$). Broader criteria ($\leq$35 years) show similar effects of 0.118 ($p < 0.01$). Confidence intervals are narrow, and consistency across both youth definitions suggests robust effects for younger workers. Programmes targeting general populations show larger point estimates (0.169, $p < 0.01$) but substantially wider confidence intervals, reflecting greater variation in effects.

Implementation Features Figure B.6 examines implementation dimensions. The upper panel shows effect measurement timing (follow-up period after programme comple-

Figure B.4: Meta-Analytic Effect Sizes by Programme Type



Notes: Meta-analytic effect sizes with 95% confidence intervals by programme type, comparing three estimation methods: unrestricted weighted least squares (UWLS), fixed effects (FE), and random effects (RE). Effect sizes represent standardised mean differences (Hedges' g). Sample sizes shown in parentheses on y-axis. I^2 statistics indicate between-study heterogeneity, shown for random effects estimates. Labour inspections (N=1) and simplification measures (N=0) excluded due to insufficient observations.

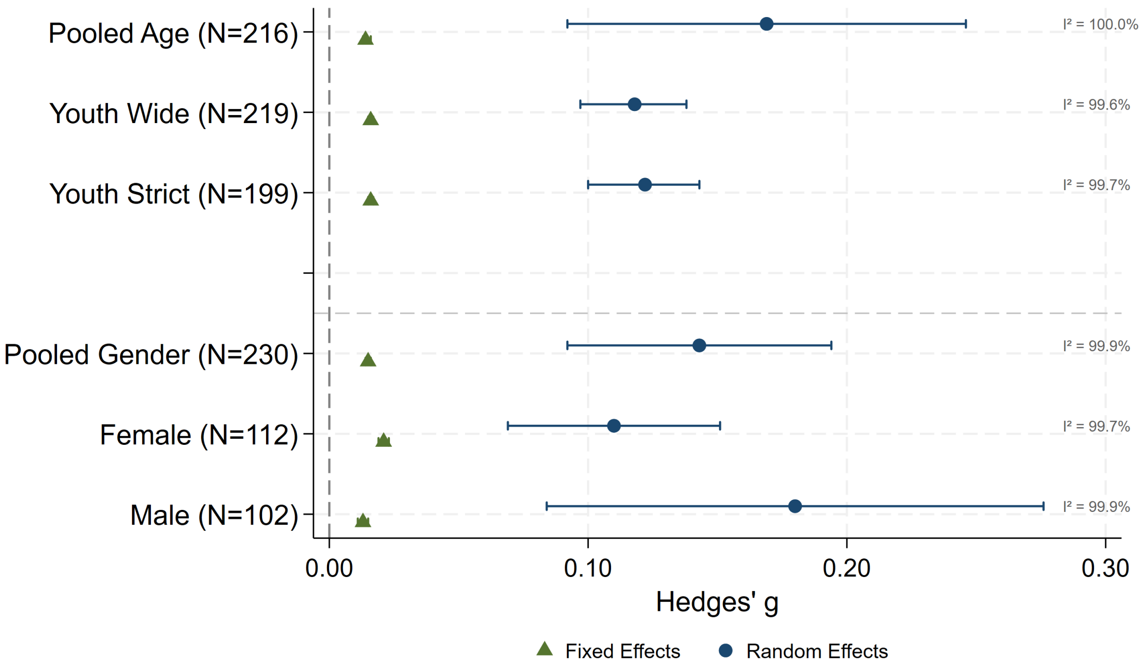
tion). Short-term measurement ($\leq$12 months after completion) shows the largest random effects estimates (0.196, $p < 0.01$). Medium-term measurement follows (0.110, $p < 0.01$). Long-term measurement shows the smallest impacts (0.079, $p < 0.01$). Confidence intervals do not overlap, suggesting these differences are meaningful.

This pattern requires careful interpretation. It does not mean short programmes are more effective than long programmes. Rather, it reflects when outcomes are measured. Short-term measurements may capture immediate compliance responses or programme participation effects that fade over time. Long-term measurements may reflect more sustainable but smaller behavioural changes. Alternatively, different types of interventions may be measured at different time horizons. The meta-regression analysis in Section 4.4 addresses this by controlling for programme type, duration, and other characteristics simultaneously, revealing that effects actually strengthen over time when other factors are held constant.

The lower panel compares programme complexity. Multi-component programmes show larger random effects estimates (0.173, $p < 0.01$) than single-component interventions (0.101, $p < 0.01$), suggesting potential complementarities between policy tools. Both categories display high heterogeneity ($I^2 > 99\%$). The difference between multi-component and single-component approaches is statistically significant.

Fixed effects and random effects estimates differ substantially across all categories, indicating considerable between-study heterogeneity. Despite this, point estimates are

Figure B.5: Meta-Analytic Effect Sizes by Target Population



Notes: Meta-analytic effect sizes with 95% confidence intervals by demographic targeting, comparing fixed effects (FE) and random effects (RE) methods. Effect sizes represent standardised mean differences (Hedges' g). Sample sizes shown in parentheses on y-axis. I^2 statistics indicate between-study heterogeneity for random effects estimates. Upper panel shows gender-specific results; lower panel shows age-targeted results. Youth Strict uses ILO definition (<30 years); Youth Wide includes participants $\leq$35 years.

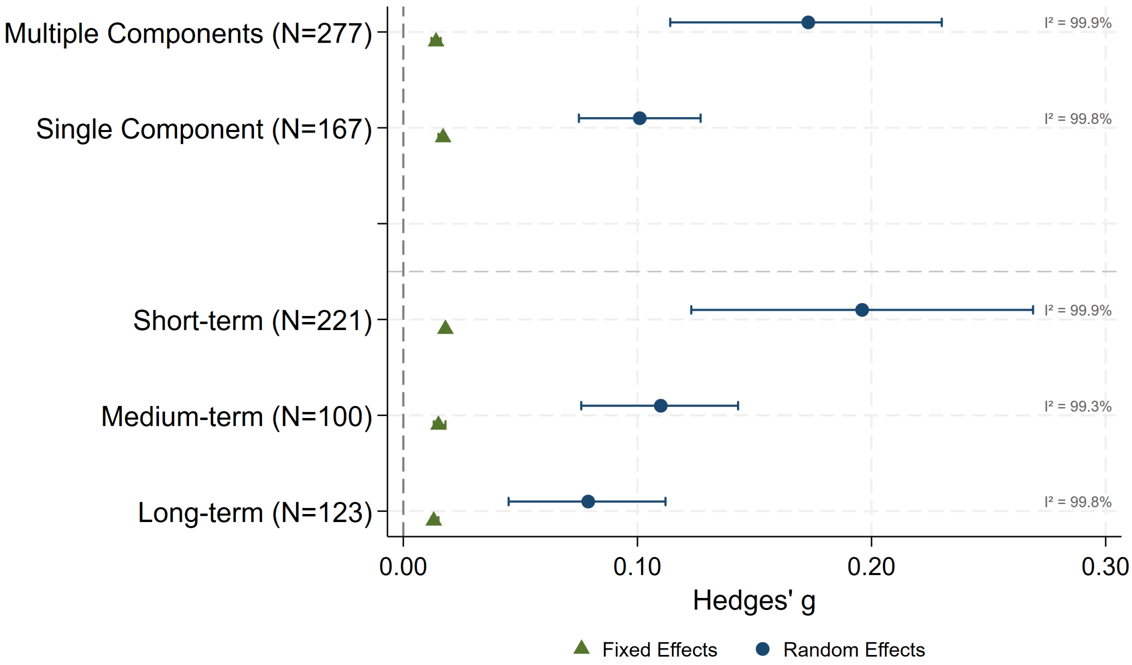
consistently positive across subgroups, and statistical significance holds in most cases. This suggests formalisation policies produce genuine positive effects, though effect magnitudes vary considerably across contexts.

Robustness Checks and Alternative Specifications

Table B.8 presents meta-regression results using PEESE publication bias correction with variance rather than standard error terms. The specification shows qualitatively similar patterns across programme types compared to the main results in Table 5, though with generally smaller effect magnitudes. The relative ranking of programme effectiveness remains consistent. Social protection programmes show the largest effects over the baseline category across both correction methods.

Tables B.9 to B.11 provide detailed meta-regression results for each outcome category separately. These confirm the heterogeneous impact patterns identified in the pooled model presented in Section 4.4.

Figure B.6: Effect Measurement Timing and Programme Complexity



Notes: Meta-analytic effect sizes with 95% confidence intervals by implementation features, comparing fixed effects (FE) and random effects (RE) methods. Effect sizes represent standardised mean differences (Hedges' g). Sample sizes shown in parentheses on y-axis. I^2 statistics indicate between-study heterogeneity for random effects estimates. Upper panel shows effect measurement timing (follow-up period): Short-term ($\leq$12 months after programme completion), Medium-term (12 to 24 months), Long-term (>24 months). Lower panel shows programme complexity: Single Component programmes implement one intervention type; Multiple Components combine different types.

Table B.6: Meta-regression results: Baseline models for total sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(i) Programme type (base: payroll + simplification)								
ALMPs	0.178	0.203	0.237	0.186	0.123	0.251	0.090	0.032
	(0.245)	(0.249)	(0.249)	(0.255)	(0.256)	(0.244)	(0.262)	(0.279)
Social protection	−0.145	−0.126	−0.081	−0.084	−0.191	−0.168	−0.247	−0.270
	(0.270)	(0.275)	(0.276)	(0.276)	(0.283)	(0.273)	(0.295)	(0.298)
Labour inspections	0.055	0.091	0.134	0.140	0.081	0.055	−0.210	−0.253
	(0.259)	(0.269)	(0.272)	(0.270)	(0.281)	(0.273)	(0.301)	(0.317)
Sensitization	0.287	0.283	0.283	0.284	0.209	0.543*	0.525*	0.505*
	(0.267)	(0.268)	(0.269)	(0.266)	(0.268)	(0.282)	(0.286)	(0.301)
Payroll tax	0.292	0.296	0.314	0.311	0.292	0.329	0.058	−0.033
	(0.267)	(0.266)	(0.266)	(0.262)	(0.266)	(0.257)	(0.281)	(0.291)
Information campaign	0.200	0.221	0.225	0.231	0.227	0.189	−0.201	−0.280
	(0.368)	(0.381)	(0.383)	(0.383)	(0.380)	(0.361)	(0.362)	(0.366)
(ii) Effect duration (base: short-term)								
Medium-term		0.016	0.029	0.019	0.004	−0.042	−0.070	−0.054
		(0.095)	(0.095)	(0.094)	(0.095)	(0.091)	(0.090)	(0.087)
Short-term		0.065	0.073	0.080	0.062	−0.058	−0.156*	−0.136
		(0.096)	(0.096)	(0.094)	(0.095)	(0.099)	(0.088)	(0.082)
(iii) Outcome category (base: hours worked, other outcomes)								
Formality outcome			0.097	0.114	0.154	0.126	0.136	0.151
			(0.111)	(0.109)	(0.108)	(0.101)	(0.099)	(0.096)
Earnings outcome			0.025	0.042	0.064	0.063	0.078	0.078
			(0.098)	(0.099)	(0.098)	(0.094)	(0.091)	(0.087)
(iv) Target group (base: men, older workers, general population)								
Female				−0.072	−0.075	−0.077	−0.080	−0.079
				(0.093)	(0.093)	(0.094)	(0.094)	(0.097)
Gender pooled				−0.064	−0.075	−0.094*	−0.063	−0.059
				(0.060)	(0.059)	(0.056)	(0.057)	(0.058)
Youth(18-24)				0.087	0.061	0.100	0.176**	0.181**
				(0.105)	(0.100)	(0.090)	(0.069)	(0.072)
(v) Programme duration (base: long duration)								
Short duration					−0.133	0.014	0.153	0.127
					(0.120)	(0.098)	(0.112)	(0.118)
Medium duration					0.001	0.106	0.089	0.034
					(0.118)	(0.103)	(0.112)	(0.102)
(vi) Evaluation design (base: non-experimental)								
Experimental design						−0.301***	−0.349***	−0.314***
						(0.095)	(0.091)	(0.090)
(vii) Contextual factors								
GDP growth								0.014
								(0.011)
Constant	0.400*	0.340	0.235	0.271	0.389	0.400	0.664**	0.637**
	(0.238)	(0.257)	(0.286)	(0.292)	(0.314)	(0.287)	(0.309)	(0.317)
Country effects	*No*	*No*	*No*	*No*	*No*	*No*	*Yes*	*Yes*
R^2	0.046	0.049	0.055	0.064	0.078	0.130	0.223	0.230
Observations	527	527	527	527	527	527	527	518

Notes: Dependent variable is binary indicator for positive and significant effect. Standard errors in parentheses clustered at study level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.7: Meta-regression results: Baseline models for total sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(i) Programme type (base: payroll + simplification)								
ALMPs	0.205	0.204	0.222	0.254	0.217	0.280	0.114	0.084
	(0.211)	(0.215)	(0.219)	(0.223)	(0.226)	(0.217)	(0.230)	(0.244)
Social protection	−0.050	−0.052	−0.022	−0.046	−0.111	−0.097	−0.176	−0.177
	(0.245)	(0.245)	(0.250)	(0.251)	(0.258)	(0.247)	(0.292)	(0.287)
Labour inspections	−0.020	−0.025	−0.001	−0.010	−0.048	−0.051	−0.245	−0.256
	(0.238)	(0.249)	(0.256)	(0.253)	(0.260)	(0.257)	(0.263)	(0.270)
Sensitization	0.346	0.347	0.347	0.349	0.317	0.515*	0.581**	0.573**
	(0.238)	(0.237)	(0.237)	(0.235)	(0.239)	(0.266)	(0.275)	(0.282)
Payroll tax	0.226	0.225	0.240	0.239	0.227	0.249	0.022	−0.007
	(0.268)	(0.267)	(0.269)	(0.266)	(0.270)	(0.266)	(0.259)	(0.265)
Simplification	0.386	0.383	0.383	0.386	0.385	0.379	0.073	0.036
	(0.300)	(0.312)	(0.312)	(0.309)	(0.309)	(0.304)	(0.281)	(0.288)
(ii) Effect duration (base: short-term)								
Medium-term		−0.006	−0.007	−0.004	−0.007	−0.006	0.003	0.008
		(0.122)	(0.123)	(0.120)	(0.118)	(0.117)	(0.107)	(0.105)
Short-term		−0.006	−0.008	−0.017	−0.020	−0.049	−0.148	−0.141
		(0.103)	(0.104)	(0.104)	(0.103)	(0.110)	(0.100)	(0.098)
(iii) Outcome category (base: hours worked, other outcomes)								
Formality outcome			0.119	0.108	0.139	0.128	0.162	0.175
			(0.124)	(0.122)	(0.118)	(0.115)	(0.111)	(0.107)
Earnings outcome			0.086	0.078	0.099	0.097	0.126	0.126
			(0.122)	(0.120)	(0.119)	(0.115)	(0.108)	(0.102)
(iv) Target group (base: men, older workers, general population)								
Female				−0.053	−0.052	−0.048	−0.039	−0.031
				(0.092)	(0.092)	(0.093)	(0.095)	(0.101)
Gender pooled				−0.070	−0.076	−0.097	−0.077	−0.079
				(0.072)	(0.073)	(0.074)	(0.071)	(0.074)
Youth(18-24)				−0.090	−0.110	−0.073	0.024	0.024
				(0.099)	(0.098)	(0.101)	(0.097)	(0.099)
(v) Programme duration (base: long duration)								
Short duration					−0.079	−0.020	0.004	−0.023
					(0.107)	(0.104)	(0.104)	(0.103)
Medium duration					0.031	0.107	−0.002	−0.046
					(0.115)	(0.117)	(0.122)	(0.108)
(vi) Evaluation design (base: non-experimental)								
Experimental design						−0.197*	−0.247**	−0.213**
						(0.102)	(0.099)	(0.099)
(vii) Contextual factors								
GDP growth								0.010
								(0.011)
Constant	0.364*	0.370	0.253	0.335	0.392	0.386	0.636**	0.607**
	(0.205)	(0.230)	(0.271)	(0.282)	(0.299)	(0.282)	(0.296)	(0.303)
Country effects	*No*	*No*	*No*	*No*	*No*	*No*	*Yes*	*Yes*
R^2	0.057	0.057	0.061	0.067	0.075	0.096	0.195	0.192
Observations	527	527	527	527	527	527	527	518

Notes: Dependent variable is binary indicator for positive and significant effect. Standard errors in parentheses clustered at study level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: Meta-regression results: PEESE publication bias correction

	All Outcomes		Formality		Earnings		Hours Worked	
	RE	FE	RE	FE	RE	FE	RE	FE
(i) Programme type (base: payroll + simplification)								
ALMPs	0.183	0.124***	0.208*	0.132***	0.169	0.041**	−0.140	−0.140
	(0.188)	(0.028)	(0.109)	(0.028)	(0.244)	(0.016)	(0.205)	(0.205)
Social protection	0.598***	0.119***	0.013	−0.035	0.981***	0.488***		
	(0.194)	(0.030)	(0.118)	(0.032)	(0.274)	(0.031)		
Labour inspections	0.324	0.155***	0.275	0.153***				
	(0.320)	(0.052)	(0.184)	(0.052)				
Sensitization	0.113	0.000	0.216*	0.047				
	(0.217)	(0.042)	(0.128)	(0.042)				
Payroll tax	0.090	0.046	0.159	0.046				
	(0.218)	(0.028)	(0.138)	(0.030)				
(ii) Effect duration (base: long-term)								
Medium-term	−0.051	0.002	−0.039	0.005***	−0.160**	−0.059***	−0.017	−0.017
	(0.038)	(0.002)	(0.029)	(0.002)	(0.080)	(0.010)	(0.033)	(0.033)
Short-term	0.086**	0.007***	−0.050*	0.008***	−0.037	−0.003	0.009	0.009
	(0.041)	(0.002)	(0.030)	(0.002)	(0.101)	(0.010)	(0.039)	(0.039)
(iii) Outcome category (base: hours worked)								
Formality outcome	0.081*	0.045***						
	(0.048)	(0.007)						
Earnings outcome	0.125**	0.040***						
	(0.050)	(0.007)						
(iv) Target group (base: men, pooled samples)								
Female	0.001	0.008***	0.001	0.011***	−0.050	−0.003	0.005	0.005
	(0.038)	(0.002)	(0.028)	(0.002)	(0.079)	(0.003)	(0.025)	(0.025)
Gender pooled	0.038	0.006***	0.005	0.008***	0.108	0.000	0.001	0.001
	(0.035)	(0.001)	(0.026)	(0.002)	(0.073)	(0.003)	(0.027)	(0.027)
Youth(18-24)	0.009	−0.018***	0.014	−0.037***	0.109	0.142***	0.004	0.004
	(0.036)	(0.004)	(0.025)	(0.005)	(0.084)	(0.013)	(0.049)	(0.049)
(v) Programme duration (base: long duration)								
Short duration	−0.167***	−0.056***	−0.065*	−0.047***	−0.205**	−0.100***		
	(0.050)	(0.007)	(0.039)	(0.009)	(0.102)	(0.012)		
Medium duration	−0.263***	−0.036***	−0.021	0.005	−0.542***	−0.178***	−0.123	−0.123
	(0.048)	(0.007)	(0.040)	(0.010)	(0.100)	(0.015)	(0.143)	(0.143)
(vi) Evaluation design (base: non-experimental)								
Experimental design	−0.039	−0.070***	−0.085***	−0.080***	−0.194*	−0.114***	−0.013	−0.013
	(0.041)	(0.004)	(0.031)	(0.005)	(0.112)	(0.011)	(0.080)	(0.080)
(vii) Contextual factors								
GDP growth	0.020***	0.007***	0.004	0.008***	0.026**	0.011***	0.012	0.012
	(0.004)	(0.000)	(0.003)	(0.001)	(0.011)	(0.001)	(0.012)	(0.012)
Variance(PEESE)	29.695***	31.878***	9.408***	20.549***	60.499***	50.586***	4.498	4.497
	(2.019)	(0.910)	(1.960)	(1.203)	(4.304)	(1.586)	(13.432)	(13.432)
Constant	−0.181	−0.038	−0.044	0.011	−0.070	0.012	0.205	0.205
	(0.206)	(0.030)	(0.122)	(0.030)	(0.254)	(0.021)	(0.228)	(0.228)
Pub. bias correction	*PEESE*	*PEESE*	*PEESE*	*PEESE*	*PEESE*	*PEESE*	*PEESE*	*PEESE*
R^2								
Observations	392	392	233	233	119	119	40	40

Notes: Meta-regression with PEESE publication bias correction using variance. RE = Random Effects, FE = Fixed Effects. Full specification includes all covariates and country fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.9: Meta-regression results: Formality outcomes (Random Effects with Publication Bias Correction)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(i) Programme type (base: payroll + simplification)							
ALMPs	0.242^{*}	0.273^{**}	0.294^{**}	0.321^{**}	0.312^{**}	0.240^{*}	0.238^{**}
	(0.134)	(0.134)	(0.134)	(0.129)	(0.131)	(0.125)	(0.109)
Social protection	0.183	0.215	0.241^{*}	0.230^{*}	0.218	0.026	0.033
	(0.139)	(0.138)	(0.138)	(0.133)	(0.135)	(0.134)	(0.117)
Labour inspections	0.206	0.207	0.207	0.326	0.302	0.388^{*}	0.320^{*}
	(0.230)	(0.228)	(0.227)	(0.220)	(0.223)	(0.211)	(0.183)
Sensitization	0.192	0.192	0.196	0.314^{**}	0.299^{**}	0.306^{**}	0.226^{*}
	(0.143)	(0.142)	(0.141)	(0.139)	(0.142)	(0.135)	(0.127)
Payroll tax	0.190	0.212	0.212	0.282^{*}	0.265	0.221	0.186
	(0.172)	(0.171)	(0.171)	(0.165)	(0.167)	(0.157)	(0.137)
(vii) Publication bias correction							
SE - Publication Bias	1.759^{***}	1.871^{***}	1.949^{***}	1.952^{***}	1.958^{***}	1.858^{***}	1.974^{***}
	(0.348)	(0.351)	(0.355)	(0.345)	(0.360)	(0.352)	(0.397)
(ii) Effect duration (base: long-term)							
Medium-term		-0.011	-0.015	-0.041	-0.041	-0.072^{**}	-0.046
		(0.034)	(0.034)	(0.033)	(0.033)	(0.033)	(0.029)
Short-term		0.054^{**}	0.050^{*}	0.002	-0.002	-0.112^{***}	-0.054^{*}
		(0.027)	(0.028)	(0.029)	(0.029)	(0.032)	(0.029)
(iii) Target group (base: men, pooled samples)							
Female			0.003	-0.000	0.001	-0.004	0.007
			(0.034)	(0.032)	(0.032)	(0.030)	(0.028)
Gender pooled			0.052^{*}	0.037	0.036	0.020	0.011
			(0.030)	(0.029)	(0.029)	(0.028)	(0.026)
Youth(18-24)			0.006	0.028	0.034	0.022	0.007
			(0.027)	(0.026)	(0.027)	(0.027)	(0.025)
(iv) Evaluation design (base: non-experimental)							
Experimental design				-0.120^{***}	-0.125^{***}	-0.127^{***}	-0.090^{***}
				(0.027)	(0.030)	(0.031)	(0.030)
(v) Programme duration (base: long duration)							
Short duration					0.018	-0.001	-0.021
					(0.037)	(0.044)	(0.041)
Medium duration					0.047	-0.016	-0.010
					(0.036)	(0.042)	(0.039)
(vi) Contextual factors							
GDP growth							0.004
							(0.003)
Constant	-0.201	-0.260^{*}	-0.312^{**}	-0.249^{*}	-0.262^{*}	-0.099	-0.142
	(0.135)	(0.137)	(0.140)	(0.135)	(0.141)	(0.140)	(0.125)
Meta-analysis method	*RE*	*RE*	*RE*	*RE*	*RE*	*RE*	*RE*
Country effects	*No*	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Publication bias correction	*SE*	*SE*	*SE*	*SE*	*SE*	*SE*	*SE*
R^2							
Observations	269	269	269	269	269	269	233

Notes: Sample restricted to formality outcomes. Dependent variable is Hedges' g (standardised mean difference). Random effects meta-regression with publication bias correction using standard error. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.10: Meta-regression results: Earnings outcomes(Random Effects with Publication Bias Correction)

	(1)	(2)	(3)	(4)	(5)	(6)
(i) Programme type (base: payroll + simplification)						
ALMPs	−0.232	−0.179	0.003	0.095	0.281	0.352
	(0.347)	(0.336)	(0.337)	(0.328)	(0.294)	(0.313)
Social protection	0.501	0.481	0.575	0.782**	1.365***	1.258***
	(0.370)	(0.360)	(0.356)	(0.352)	(0.315)	(0.347)
(vii) Publication bias correction						
SE - Publication Bias	9.902***	10.775***	11.619***	11.278***	10.408***	11.588***
	(1.046)	(1.054)	(1.085)	(1.094)	(1.094)	(1.156)
(ii) Effect duration (base: long-term)						
Medium-term		−0.286**	−0.319**	−0.334***	−0.317***	−0.317***
		(0.124)	(0.125)	(0.120)	(0.105)	(0.105)
Short-term		0.047	−0.054	−0.096	−0.206*	−0.194
		(0.109)	(0.117)	(0.114)	(0.123)	(0.130)
(iii) Target group (base: men, pooled samples)						
Female			−0.053	−0.063	−0.028	−0.058
			(0.119)	(0.113)	(0.094)	(0.099)
Gender pooled			0.155	0.174*	0.115	0.131
			(0.106)	(0.102)	(0.088)	(0.093)
Youth(18-24)			−0.005	−0.021	0.180**	0.042
			(0.093)	(0.094)	(0.088)	(0.108)
(iv) Evaluation design (base: non-experimental)						
Experimental design			−0.212**	−0.179*	−0.382***	−0.324**
			(0.099)	(0.107)	(0.129)	(0.144)
(v) Programme duration (base: long duration)						
Short duration				−0.045	−0.215*	−0.129
				(0.119)	(0.120)	(0.132)
Medium duration				−0.349***	−0.615***	−0.656***
				(0.106)	(0.120)	(0.126)
(vi) Contextual factors						
GDP growth						0.026*
						(0.014)
Constant	−0.163	−0.202	−0.321	−0.268	−0.130	−0.356
	(0.343)	(0.335)	(0.346)	(0.351)	(0.314)	(0.335)
Meta-analysis method	*RE*	*RE*	*RE*	*RE*	*RE*	*RE*
Country effects	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Publication bias correction	*SE*	*SE*	*SE*	*SE*	*SE*	*SE*
R^2						
Observations	131	131	131	131	131	119

Notes: Sample restricted to earnings outcomes. Dependent variable is Hedges' g (standardised mean difference). Random effects meta-regression with publication bias correction using standard error. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.11: Meta-regression results: Hours worked (Random Effects with Publication Bias Correction))

	(1)	(2)	(3)	(4)	(5)	(6)
(i) Programme type (base: payroll + simplification)						
ALMPs	0.021	0.023	0.080***	0.100*	−0.087	−0.203
	(0.023)	(0.023)	(0.029)	(0.055)	(0.171)	(0.207)
(vii) Publication bias correction						
SE - Publication Bias	1.117*	1.079	1.598***	0.807	−0.183	−0.314
	(0.654)	(0.657)	(0.601)	(0.744)	(1.323)	(1.398)
(ii) Effect duration (base: long-term)						
Medium-term		−0.030	−0.026	−0.030	0.006	−0.010
		(0.027)	(0.022)	(0.022)	(0.029)	(0.033)
Short-term		−0.026	−0.037*	−0.050**	0.019	0.017
		(0.024)	(0.019)	(0.021)	(0.035)	(0.040)
(iii) Target group (base: men, pooled samples)						
Female			0.008	0.003	−0.000	−0.002
			(0.023)	(0.023)	(0.025)	(0.026)
Gender pooled			0.007	−0.000	−0.008	−0.010
			(0.023)	(0.023)	(0.029)	(0.030)
Youth(18-24)			−0.016	−0.002	0.006	0.012
			(0.026)	(0.030)	(0.045)	(0.048)
(iv) Evaluation design (base: non-experimental)						
Experimental design			−0.103***	−0.112***	−0.030	0.014
			(0.022)	(0.030)	(0.065)	(0.082)
(v) Programme duration (base: long duration)						
Short duration				0.099*	0.037	
				(0.058)	(0.085)	
Medium duration				0.019	−0.099	−0.173
				(0.035)	(0.124)	(0.147)
(vi) Contextual factors						
GDP growth						0.013
						(0.013)
Constant	−0.019	0.001	−0.016	0.002	0.219	0.306
	(0.028)	(0.033)	(0.037)	(0.048)	(0.236)	(0.261)
Meta-analysis method	*RE*	*RE*	*RE*	*RE*	*RE*	*RE*
Country effects	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Publication bias correction	*SE*	*SE*	*SE*	*SE*	*SE*	*SE*
R^2						
Observations	44	44	44	44	44	40

Notes: Sample restricted to hours worked outcomes. Dependent variable is Hedges' g (standardised mean difference). Random effects meta-regression with publication bias correction using standard error. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.