IDB WORKING PAPER SERIES Nº IDB-WP-01478

Can a Pay-for-Performance Program Help the Vulnerable find Jobs during a Pandemic?

Experimental Evidence from *Empleate* in Colombia

Maria Fernanda Gómez
Carolina González-Velosa

Inter-American Development Bank
Labor Markets and Social Security Division

July 2023

Can a Pay-for-Performance Program Help the Vulnerable find Jobs during a Pandemic?

Experimental Evidence from *Empleate* in Colombia

Maria Fernanda Gómez
Carolina González-Velosa

Inter-American Development Bank
Labor Markets and Social Security Division

July 2023

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Gómez, Maria Fernanda.

Can a pay-for-performance program help the vulnerable find jobs during a pandemic?: experimental evidence from Empleate in Colombia / Maria Fernanda Gómez, Carolina González-Velosa.

p. cm. — (IDB Working Paper Series ; 1478)

Includes bibliographical references.

1. Labor market-Colombia. 2. Labor supply-Colombia. 3. Unemployed-Government policy-Colombia. 4. Incentives in industry-Colombia. 5. Coronavirus infections-Social aspects-Colombia. I. González-Velosa, Carolina. II. Inter-American Development Bank. Labor Markets Division. III. Title. IV. Series.

IDB-WP-1478

http://www.iadb.org

Can a Pay-for-Performance Program Help the Vulnerable find Jobs during a Pandemic?

Experimental Evidence from *Empleate* in Colombia

Maria Fernanda Gómez

Carolina González-Velosa[1]

July 2023

[1] This paper was written as part of the project "Paying for Jobs for the Vulnerable during the COVID-19 crisis" (RG-E1722) and financed with ordinary capital of the Inter-American Development Bank. We are deeply indebted to Adriana Kugler, for her leading role in the design and implementation of this evaluation. We also thank David Kaplan, Mariano Bosch and an anonymous referee for useful comments and Lucia Carrillo for research assistance. We are profoundly grateful to Laura Hincapie for her support during the implementation of the evaluation and her insightful comments and suggestions and to Lida Moreno, Bibiana Taboada, Juan Camilo Giraldo, Yineth Rodriguez and Camilo Coronado, from Prosperidad Social, for making this evaluation possible and acknowledging its independence. All errors are our own.

Abstract

During a period of COVID-19-induced job losses and mobility restrictions, the government of Colombia's launched Empléate, an innovative Pay-for-performance (P4P) program that targeted impoverished and vulnerable workers. Empléate operated at a national scale and had a novel financial arrangement: in contrast to traditional programs wherein service providers are remunerated based on their activities, service providers in Empleate only received payments based on successful placement of participants into formal employment. They were also granted premia for sustaining participants in formal jobs months for 3 to 6 months after insertion. This article presents the results of a randomized impact evaluation of Empleate conducted between September 2020 to April 2021. The Intention-to-Treat (ITT) estimates show that program participants were 9% more likely to secure a formal job five to eight months post-treatment. Impacts were larger among men and among individuals with work experience in sectors less affected by the pandemic, with the impacts rising to 22% and 17% respectively. There is no evidence of impacts among women and among individuals without secondary education. These ITT impacts likely underestimate real Average Treatment Effects (ATE) due to issues of imperfect compliance. Complementary analysis from survey data suggests creaming, underscoring the importance of ensuring an adequate allocation of financial risk on P4P contracts. Nevertheless, many design features are promising and positive impacts are noteworthy considering the adverse repercussions in Colombia's formal labor market inflicted by the pandemic.

JEL CODES: H43, H57, I38, J24,

KEYWORDS: Labor Markets, Employment Programs, Results-Based-Financing, Active Labor Market Policies, Poverty

Executive Summary

- This paper presents experimental estimates of the impact of *Empléate*, a pay-for-performance (P4P) employment program in Colombia that targeted poor and vulnerable workers from September 2020 to April 2021, at a time when there were severe mobility restrictions and massive job losses in the country due the COVID-19 Pandemic.

- The program was administered by Prosperidad Social, the government agency in charge of social protection programs in Colombia. Prosperidad Social contracted service providers to provide a flexible combination of services that could include intermediation, job placement, case management, training services, and transportation allowances. As this was a P4P program, payments to service providers were conditional on formal employment of participants. Specifically, Prosperidad Social paid to service providers if and only if beneficiaries were placed in a formal job right after the intervention. The program also paid additional premia if the participant remained on the job after three months and after five months, respectively. In addition, the program paid premia for placement of specific demographic groups (e.g., women aged 40 or more, disadvantaged youth who participated in the program *Jóvenes en Acción* and people with disabilities). These outcomes were verified through the auditing of contracts and/or social security administrative records. This P4P scheme was an innovation in service delivery in Colombia since employment programs typically make payments to service providers to fund inputs and implementation costs.

- This paper analyses data from service providers contracted to implement the program, including employment and training agencies, as well as large private firms with structured training programs.

- Applicants to the program were randomly assigned to receive services from each of the nine service providers or to be in a control group that received no services. Given the expectations of excess demand, program administrators agreed to randomly allocate eligible participants to treatment and control groups.

- As is often the case of voluntary programs, however, there was less than full compliance. A large number of workers assigned to treatment did not take up the treatment. The number of these "non-takers" was large partly due to delays in the operation of the program which deferred provision of services, intense screening from program providers, and difficulties participating because of mandatory lockdowns due to the pandemic. Yet, non-compliance was mostly confined to the treatment group. Contamination among the control group was low, with few participants assigned to the control group getting access to services.

- We use administrative data (social security and social assistance) to collect socioeconomic information and tracking changes in the formal labor market of beneficiaries. ITT estimates of the effects of the program show that being randomly assigned to the treatment group has no effects on formal employment in the short run (two to five months after program application) but has an impact on formal employment in the medium run (five to eight months after program application). This suggests that the financial incentives given to service providers to retain beneficiaries in formal jobs for at least five months were particularly effective. Specifically, compared to those who were randomly assigned to the control group, participants assigned to the program were 9% more likely to have a formal job five to eight months after the treatment. This impact increases to 22% in the case of people with work experience in sectors that were less affected by the Covid-19 pandemic[2]. It is also higher among men, for whom the impact is 17%. There is no evidence of employment effects for women and people without secondary education. Given the low compliance levels of treated individuals, these estimated impacts likely underestimate average treatment effects.

- ITT estimates also show that being randomly assigned to the treatment group has no effects either on the monthly average wage in the short run or in the medium run. However, people receiving services from firms increased their monthly average wage in the medium run (five to eight months after program application) by US$ 5.4 (or 20,394 Colombian Pesos)[3] or 14% compared to those randomly assigned to the control group. Also, people with work experience in sectors that were less affected by the pandemic raised their monthly average wage in the same period by US$6.4 (or 24,013 Colombian Pesos) [4] or 17% compared to those from the control group. These were the only groups of interest that showed statistically significant wage effects.

- Moreover, ITT estimates show that, conditional on having a formal job, being assigned to the program increased the probability of working in a large firm (with more than twenty workers) by 3%. These effects were larger for men (8%), people with work experience in sectors that were less affected by the pandemic (8%), people without a college degree (5%) and those 30 years of age or older (4%)[5]. When comparing impacts between different types of service providers, larger effects were observed among the subsample of beneficiaries that received services from private firms.

- ITT estimates show no effects among the three demographic groups for which there was additional premia paid for placement (i.e., women aged 40 or more, disadvantaged youth who

[2] Sectors different from: services, commerce, hotels, restaurants, entertainment, and recreation.

[3] Value in dollars is estimated using the official exchange rate on September 1st,2020 (i.e., 3,745 Colombian Pesos per dollar).

[4] Value in dollars is estimated using the official exchange rate on September 1st,2020 (i.e., 3,745 Colombian Pesos per dollar).

[5] Even though private providers were in different parts of the country, there may be positive selection if applicants to private firms were generally more motivated or better prepared.

participated in the program *Jóvenes en Acción* and people with disabilities). These results highlight that additional financial incentives that are too small may not be sufficient to promote employment among workers who face difficulties entering the labor market.

- We complement the impact evaluation with an analysis of follow up survey of a random sample of eligible individuals for the program (assigned to the treatment and control groups). This data provides suggestive evidence of positive selection (creaming): beneficiaries were more likely to be male, had a higher educational attainment and a longer labor market history than average eligible participants. This is likely due to the strong incentives created by program. Given that 100% of the payment was contingent on placement and retention, the risk transfer to providers may have been too large. Moreover, the implementation schedule was tight, reducing the capacity to supply intensive services (indeed, our data shows that even though service providers could choose from a wide array of services that included training and skill certification, in most cases the intervention was limited to intermediation services and job-search assistance workshops). In this context, positive selection (creaming) may have been efficient, ensuring that services were delivered only to participants who were easier to employ, such as men with higher levels of educational attainment. This issue underscores the importance of ensuring an adequate allocation of financial risk on performance-based contracts.

- Nevertheless, many design features of *Empléate* are promising and ITT estimates indicate it had positive effects: individuals with low levels of education were able to keep formal jobs months after the intervention. Given the elevated levels of informality and instability in Colombia´s labor market and given the dramatic impact of the pandemic, this is a commendable achievement.

1. Introduction

This paper presents the results of the experimental evaluation of the *Empléate* program, which was implemented in Colombia in September of 2020. The program was introduced in the middle of a deep recession faced by Colombia following the COVID-19 outbreak.Mobility constraints to control the pandemic reduced demand sharply and shut down large parts of the economy. Financial constraints created by the loss of jobs and income, also reduced demand. As the crisis unfolded globally, international trade and capital flows halted. As a result, in 2020 output shrank by 7% and unemployment and inactivity rose sharply. In April 2020, 4.5 million jobs were lost with respect to February and the unemployment doubled, reaching 21%, and was even higher among women and youth. The number of people leaving the labor force also increased dramatically, especially among women with young children who assumed childcare responsibilities due to school closures.

Amidst this crisis, the Department of Social Prosperity (Prosperidad Social), the government agency which administers programs to alleviate poverty in the country, decided to implement *Empléate*, a pilot program targeted to the poor and vulnerable, which provided intermediation, training, and job-search assistance services to find formal employment. *Empléate* had two salient characteristics. First, it was a pay-for-performance (P4P) program in which payments to service providers were contingent on formal job placement and on retention in formal jobs for at least 3 and 5 months, respectively. Second, *Empléate* allowed a great deal of flexibility in terms of the services the providers could give workers. The services included interviewing assistance, resume preparation assistance, help with wardrobe for interviews, soft skills training, technical training, transportation, and post-placement assistance. According to a recent meta-analysis of employment programs around the world, these two characteristics improve the likelihood of success of employment programs (Kluve et al, 2016).

Given the over-subscription expected for this program, Prosperidad Social agreed to randomly assign individuals to the treatment. Interested individuals were asked to fill up an entry survey. Once their eligibility criteria (i.e., a high poverty level, lack of formal employment in the last four months, Colombian nationality, and to have not been a beneficiary of any other social or employment programs) was verified, applicants were randomly assigned to the treatment. Many individuals assigned to the treatment did not end up participating in the program due to administrative delays, problems with the operation of the program, the fact that the number of people that applied to the program and were randomly assigned was much higher than the number of people that service providers could attend, and further verification of vacancy requirements by the service provider. By contrast, there were little non-compliance in the control group: as most of those assigned to the control group did not participate in the program.

This paper shows Intention-to-Treat (ITT) estimated effects of *Empléate* on the labor market outcomes of individuals randomly assigned to the program from September 2020 to April 2021, using administrative data from social security records. Non-compliance from the treatment group implies that ITT estimates will be downward biased. ITT estimates show a positive impact of 9% on the probability of being in a formal job from five to eight months after the intervention. These results appear to be driven by impacts among workers with work experience in sectors that were less affected by the pandemic (22%), men (17%) and among workers without college (10%). Also, there is a positive impact of 14% on monthly average wage five to eight months after program application, for those participating with private firms as service providers, and an impact of 17% for workers with experience in sectors that were less affected by the pandemic. There are also positive impacts on firm size, conditional on being employed. Individuals assigned to treatment had a 3% higher probability of working in a firm of more than 20 workers. This result also appears to be driven by the impact on men (8%), workers with experience in sectors less affected by Covid-19 (8%), individuals older than 30 years of age (5%), and workers with no college degree (4%). There is no evidence that *Empléate* improved labor market outcomes of women and individuals younger than 30, nor in the three groups with additional premia (women aged 40 or more, disadvantaged youth who participated in the program *Jóvenes en Acción* and people with disabilities). All of these ITT estimates are likely to be lower bound estimates of the effects of the program, given the low compliance among those offered treatment.

The rest of the paper proceeds as follows. Section 2 provides a literature review of P4P programs. Section 3 describes the *Empléate* program. Section 4 describes the randomization and the methodology used to estimate intention-to-treat effects. Section 5 describes the administrative and survey data we use in the analysis. Section 6 presents estimates of overall impacts on several employment outcomes, such as employment and wages, as well as estimates of heterogeneous effects and robustness checks. Section 7 presents an analysis using complementary information from a follow up survey that informs of the underlying mechanisms driving the impacts. Section 8 concludes.

2. Literature Review

Governments of developed and developing countries engage in a variety of Active Labor Market Policies (ALMP) to increase opportunities in the labor market for specific demographic groups. Some of these ALMPs include: i) training programs, that seek to improve skills, ii) wage subsidies that seek to rise the labor demand of firms, iii) job search, and intermediation assistance that seek to improve the matching of firms and workers, and iv) public work programs that seek to increase labor demand. Interest in the effectiveness of ALMP typically increase during economic downturns, such as the global financial crisis or, more recently, the pandemic. Other global trends, such as automation and emigration, which structurally affect the demand and supply of skills in the labor market, have also increased the attention to this type of policies (McKenzie, 2017).

The impact evaluation literature of ALMPs is abundant. Evidence of causal effects has been available for decades, particularly in developed countries. In recent years, the evidence of experimental evaluations has risen dramatically. Evidence from developed countries has also increased notably. This has allowed for systematic reviews that summarize experimental and non-experimental evidence in developing countries (for instance, Card et al, 2017, Kluve et al 2017, Yeyati, 2019, McKenzie, 2017).

Generally, these reviews find mixed evidence on the effectiveness of ALMPs. Average impacts vary widely across groups, type of service and the time frame in which the impact is evaluated. The economic context in which the programs are implemented, and design features are also important in determining their effectiveness.

The literature shows that, frequently, average impacts are small or null. For instance, Kluve et al (2017) reviews a large body of evidence, 113 counterfactual impact evaluations of youth employment programs in developed and developing countries, using meta-analysis methods that address issues stemming from publication bias and selective reporting. They find that just one third of the evaluations have a positive impact on labor market outcomes, such as employment or earnings. And McKenzie (2017), in a smaller review of 22 experimental evaluations in developing countries, finds that impacts are generally modest. A typical intervention leading to a 2 or 3 percent increase in employment that often is not statistically different from zero. And in a systematic review of over 200 recent impact evaluations of ALMP, Card et al. (2017), find that average impacts are small or zero in the short term, but become larger or more positive 2 or 3 years after program completion.

With respect to the type of intervention, Card et al (2017) find that the effectiveness depends on the type of intervention and, in some cases, on the time profile. Programs that emphasize human capital accumulation (e.g.., training and private sector employment) have larger effects in the medium and longer runs whereas job search assistance programs that emphasize "work first" tend to have similar impacts in the short and long run. Public sector employment subsidies have small or null impacts at all time horizons. Regarding heterogeneity across participant groups, the authors find that female participants and long term unemployed tend to experience larger effects. However, impacts vary across demographic groups: disadvantaged participants appear to benefit more from work first programs whereas training and private sector employment subsidies tend to have larger average effects for the long term unemployed. Thus, there appears to be benefits to matching specific participant groups to specific types of programs. This conclusion is consistent with Kluve et al (2016), who do not find evidence of certain types of programs systematically outperforming others. They find, instead, evidence suggesting that programs that can adapt, from a broad portfolio of interventions, different services to different types of beneficiaries, tend to have better results.

A very important and related finding stems from the systematic review by Kluve et al, (2016), who show that programs that integrate multiple interventions are more likely to have a positive impact. Thus, there is no single specific combination that always works: programs that profile beneficiaries and combine varied interventions to respond to the diverse needs of a heterogeneous group of participants are more successful. Also important is the evidence about incentives systems for service providers. Kluve et al (2016) find that programs that make payments to service providers based on their performance have a greater chance of positive outcomes. They are, however, unable to identify the specific type of contractual arrangement that is more successful.

Thus, Kluve et al (2016) find evidence in favor of key features of P4P employment programs, which provide incentives to service providers and allow for flexibility in the intervention. Some authors have also assessed possible unintended consequences of P4P designs, such as "creaming" or "cherry-picking," in which providers have perverse incentives to restrict services only to more-employable participants. A well-studied intervention is Job Training Partnership Act (JTPA), implemented by the U.S. Department of Labor (DOL) in 1982 (Hawkins, 1982) which distributed funds to state programs and pre-defined performance outcomes. A common criticism of the JTPA is that training centers engaged in self-selecting their trainees ("creaming") to avoid the hardest to employ. In addition, Courty and Marschke (1998) found a moral-hazard problem where managers of participating agencies manipulate performance outcomes to maximize their awards. Heckman et al. (1997) found, however, that these concerns regarding "creaming" in the JTPA were exaggerated. Anderson et al. (1993) found evidence of creaming in the JTPA based on participating agencies from Tennessee and showed that participants who were poorly educated or in poor health were systematically deemed ineligible. The potential effects of creaming by Tennessee agencies indicate that the 71% placement rate would fall to about 62% if participants were randomly enrolled. On the other hand, Heckman and Smith (1995) find no evidence of self-selection and selection by agency staff. Similarly, while agency staff may attempt to engage in creaming, Bell and Orr (2002) use the JTPA and several other job-training programs in their evaluation and find those staff ratings of trainees explain less than 10% of the variation in earnings and welfare benefits after participation in the program.

Colombia has experience implementing this type of P4P contracted-out employment programs. The *Jóvenes en Acción* (JeA) program, which started implementation in 2002, had a P4P component that paid part of the cost of offering training to the providers up-front and the remaining the trainees were placed in an internship. Attanasio, Kugler and Meghir (2011) and Kugler et al. (2022) conducted an experimental evaluation of this program and show that the JeA program had both short- and long-term benefits for participants that more than covered the costs of the program.

More recently, Colombia has invested in a novel type of results-based contracts: social impact bonds. These are a special type of P4P contracts in which governments enter into agreements with service providers and investors to pay for the delivery of social outcomes. Typically, investors provide funding for the intervention upfront, allocating resources to service providers to cover their operating costs. If the outcomes agreed upon are achieved and verified by an external agency, the government makes payments for these outcomes to the investors. Financial risks are, therefore, not assumed by the government nor the service providers. They are, instead assumed by social investors who, in principle, may receive a moderate return on investment. Since 2017 Colombia has launched four social impact bonds targeting employment support to the vulnerable. Quasi-experimental evidence from a recent study indicates that the first of these impact bonds -*Empleando Futuro* – had large and persistent positive effects. Three months after the intervention, participants had a probability of being formally employed that was 12 percentage points larger than that of the control group. This effect is 8 percentage points a year and three months later (Chaparro et al, 2020).

Thus, there is evidence that several design features can improve the effectiveness of ALMPs, such as PMP arrangements. The literature also examines how the business cycle may affect the effectiveness of ALMP, albeit with conflicting results. Card et al. (2017) find that ALMP are more likely to show positive impacts during recessionary periods, especially through fleeting shocks. They also find that human capital programs are more cyclically sensitive. Yeyati et al (2019), on the other hand, reach a different conclusion. In their systematic review of 102 interventions, they find the effectiveness programs positively correlate with per capita GDP growth and negatively with the unemployment rate during implementation.

A new strand of literature examines the persistence of the impacts of ALMP after an economic crisis such as the one derived from the COVID-19 pandemic. Barrera et al. (2021) analyze if a pre-pandemic job training program in Cali - Colombia targeting low-income youth sustains its benefits through the pandemic. They find that the crisis washed away all the benefits of the program on employment and earnings, and that there does not seem to be a recovery on labor outcomes of trainees.

Some reasons the authors identify that may explain these are the significantly high magnitude of the economic shock of the pandemic, the short duration of the training program, and the fact that the services sector was the most affected by the pandemic as well as the most common sector for training in the program. Similarly, Hanushek et al (2017) finds that when the skills provided by training programs are too specific, they may be inadequate to adjust to economic shocks.

We contribute to this literature with a randomized control trial of a P4P employment program in Colombia, *Empleate*, that provided very strong monetary incentives to service providers by conditioning all the payments to employment outcomes. Thus, unlike many P4P programs (e.g., Jovenes en Accion), service providers received no upfront payments and assumed all the financial risk of the program. Possible unintended consequences, such as creaming, are discussed. We also contribute to the literature that examines the impact of downturns on the effectiveness of ALMPs; *Empleate* was implemented in 2020, amidst the largest economic downturn in Colombia´s recent history.

3. Colombia´s P4P employment programs and *Empléate*

P4P contracting models have been implemented in employment programs in the US and the UK since the late 1980's. More recently, OECD countries like Australia, the UK, Netherlands, Sweden, and Italy have adopted models of large-scale outcomes-based contracting (OECD,2022). As discussed in the previous section, Colombia also has experience implementing employment programs with P4P components and has achieved positive results.

In 2017, with the launch of *Empleando Futuro*, Prosperidad Social started an agenda to introduce innovative outcomes-based financing components in employment interventions. *Empleando Futuro* was the first contracted social impact bond in a developing country for which there was government outcome funding (Brookings, 2017). Since then, Prosperidad Social created an outcomes fund to reduce the transaction costs of implementing several new P4P initiatives. *Empléate* was one of them.

In *Empléate*, Prosperidad Social established a financing arrangement in which nine service providers received payments contingent on the verified achievement of results[6]. By tying funding to outcomes, Prosperidad Social intended to align the incentives of service providers results and, also, to give service providers flexibility in the design and delivery of the interventions. Service providers did not receive upfront payments at the beginning of the program. They only received payments based on outcomes with the rates shown in Table 1. Thus, for each beneficiary that got a formal job at the end of the intervention, they received approximately USD $405. If a beneficiary still had a formal job three or five months after the intervention, they received an additional USD

[6] There were 12 services providers participating in the program. We only evaluate results from participants serviced by 9, as the other 3 providers started program before the random assignment was feasible.

$149 and USD $68, respectively. There were extra payments for employment results among beneficiaries with disabilities, women aged 40 or more, and graduates from *Jóvenes en Acción*, a program targeted to vulnerable youth that provides subsidies for tertiary education and vocational training.

Table 1. Payments for outcomes in *Empléate* Program

	Payment per beneficiary (USD)
Beneficiary gets a formal job at the end of the intervention	$405
Beneficiary remains in a formal job 3months after the intervention	$149
Beneficiary remains in a formal job 5months after the intervention	$68
Premium if beneficiary has a disability	$68
Premium if beneficiary is a women aged 40 or more	$68
Premium if beneficiary participated in the *Jóvenes en Acción* program	$68

Notes: Payments in dollars are estimated using the official exchange rate in September 1st,2020 (i.e., 3745 Colombian Pesos per dollar).

Participants in *Empléate* had to meet at least one of the following criteria to be eligible: i. be poor or vulnerable according to Colombia's proxy mean test, SISBEN; ii. be a graduate from the *Jóvenes en Acción* program, a cash transfer program targeted to vulnerable youth conditional on education and training; iii. be a member of the Red Unidos, an antipoverty strategy targeted to poor families that provides psychosocial support and preferential access to social services. Moreover, they should not have had a formal job in the last 4 months, as verified by social security registries (PILA), nor participated in employment programs in the last three years. Finally, they had to be Colombian citizens and 18 years old or older at the time they participated in the program.

Service providers could receive basic intermediation services (such as registration in the

Public Employment Service, job -search workshops or aptitude testing) or more “specialized” services, such as occupational training or childcare support. Data from a follow up random survey, shows that 99% of beneficiaries received basic intermediation services and only 14.8% received specialized services. The most frequent basic services were: registration in the Public Employment Service Information System (78.9% of beneficiaries), psychological and technical testing (33.6%), and workshops providing job-search tools (32.1%). Specialized services were way less frequent and, when present, consisted of: resume polishing and document management (13.3%), technical training (8.4%), support during the selection process, such as. home visits, specific labor skills tests (5.7%), and complementary support to provide work attire, transportation, and childcare (5.6%). (see Appendix 1 for more details).

We evaluate the program *Empléate* from September 2020 to April 2021[7] when it operated in 14 regions of the country[8]. More than 70% of the participants lived in Bogota or the northern regions (i.e., departments of Atlántico, Córdoba, and Bogota). Participants were on average 30 years old, and 77% were women. The vast majority were ranked as poor according to the proxy means test (91.4% belong to two lowest socioeconomic strata), 19.7% had completed higher education, 25.5% had a job before their application to the program and their average income was COP 588,810 (which is less than one minimum wage). When applying to the program, 65.6% of beneficiaries expected to have a salary equivalent to one minimum wage and 30.4% a salary of between one and two minimum wages. Most (58.2%) aspired to having a stable work in a company.

There are important differences in the design of *Empleate* and *JeA* which, as discussed in the previous section, had positive short run and long run impacts according to an experimental evaluation by Attanasio, Kugler and Meghir (2011) and Kugler et al. (2022). Whereas in JeA, a

[7] The program operated until November 2021, but we only consider cohorts until April 2021.
[8] Service providers were based in the following sates or departments: Antioquia, Atlántico, Cundinamarca, Valle del Cauca, Meta, Risaralda, Santander, Norte de Santander, Córdoba, Huila, Bolivia, Magdalena, Nariño y Bogotá D.C.

fraction of the payment was made upfront and the other conditional on the internship, in Empleate all payments to service providers were contingent on employment outcomes. Another difference is the combination of services: JeA program focused more on human capital accumulation, offering technical and soft skills through classroom and on the job training, but it did not offer basic placement services, retention services and tools for remote work. On the contrary, Empléate was more focused on job assistance and placement, rather than training. Both programs offered a monetary stipend to cover for transportation and childcare costs.

4 Methodology

4.1. Randomized Experiment

Random assignment was applied to a group of individuals who, according to data provided by Prosperidad Social, met the eligibility criteria and expressed interest in the program. This process had several stages. First, Prosperidad Social and service providers invited potential participants through multiple channels (e.g., social media, press, radio, local advertising, flyers, etc.). Second, interested participants filled out a short survey and Prosperidad Social verified eligibility criteria (i.e., poverty level, lack of formal employment in the last four months, nationality, and not having been a beneficiary of other employment program) using administrative data. Eligible participants were, then, randomly allocated to treatment or control groups and the contact information of the treated individuals was sent to the service providers.

The program operated on a rolling schedule, with program administrators permanently summoning participants and providing services. There were twenty-two randomizations from September 30 to April 30. However, the program continued its operations until November 2021. Randomizations were done at the level of the service provider since individuals applied to the program through each of the service-providers operating in their location and they could be assigned to different vacancies within the same provider.

There were twelve service providers but only nine participated in this study. Service providers included firms, private and non-profit organizations as well as regional institutions in charge of the administration of a wide array of public subsidies and social services including training and employment agencies.[9]

Each randomization was done using a random number generator in Stata, which ranked individuals within each service provider. The top 85th percentile of applicants according to the ranking were assigned to the treatment group and the remaining 15th percentile were assigned to the control group.[10] For each randomization, we verified there were no systematic differences in baseline characteristics, including gender, age, level of education, and housing conditions.

For several reasons, the treated group was a non-randomly selected subset of those assigned to the treatment. First, as it is often the case in voluntary programs, not all of those expressing interest eventually participated in *Empléate*. Also, due to administrative difficulties, sometimes there were delays between the enrollment and service provision, reducing program participation of those assigned to the treatment. More importantly, service providers screened among those assigned to treatment and selected those that had a greater chance to be employed at the end of the program. In fact, many service providers had pre-identified potential vacancies and selected participants that met specific vacancy requirements (e.g., age or specific education degrees) among those assigned to the treatment.

Given that this screening process was prevalent, a vast number of those assigned to the

[9] These providers included: Corporación Gestión Empresarial, Fundacion Colombia Incluyente, Contrapensimeta, Fenalco Meta, Combarranquilla, Corporación Minuto de Dios, Andes BPO, Cajasan, Academia Sinú, Audio VisualesSurcolombiana and Industrial Cacaotera de Huila and Comfenalco Valle.

[10] A majority was assigned to the treatment group as requested by the program administrators. Given that the program operated under difficult and uncertain conditions (severe lockdowns amidst the pandemic), and program outreach was challenging, they wanted to reduce the risk of having insufficient participation.

treatment, did not participate in the program. Table 2 below classifies the sample according to random assignment (Zi) and program participation (Di). Between September 2020 and April 2021, 21,928 eligible individuals were randomly assigned to the treatment and 3,842 were randomly assigned as controls. There was unfortunately a severe compliance problem confined to the treatment group. Of those assigned to the treatment, a vast majority, 20,513, did not participate in the program and only 1,415 (6%) did. On the contrary, compliance was very high among those assigned to the control group: very few (51 individuals) participated in the program, and 3.791 (99%) did not. The fact that compliance was so high among the control group, with very little contamination, is reassuring.

Table 2. Participants by lottery assignment and program participation.

	Randomly Assigned to Control Zi=0	Randomly Assigned to Treatment Zi=1	Total
Did Not Participate Di=0	Group A: 3,791 (99% of control group)	Groups B: 20,513	24,304
Participated Di=1	Group C: 51	Group D: 1.415 (6% of treatment group)	1,466
Total	3,842	21.928	25,770

Given that Group D, the group that receives the assigned treatment, is a non-random selection of those that were assigned to the treatment, comparisons between this group and the control group would be misleading. It is very likely that the selection of those assigned to treatment into Group D is positive: those who were selected by service providers due to their employment attributes may have earned more even without participating in the program. Below, we describe the identification strategy we use to address the bias due to non-compliance.

4.2. Identification Strategy

Given random assignment, we estimate the causal effects of the *Empléate* program using the following regression model:

$$Y_i = \alpha + \rho \text{Assigned}_i + \theta_i + \eta_i \tag{1}$$

where the Assigned_i indicator takes the value of 1 if the person is randomly assigned to the program and zero if they were assigned to the control group. The outcome variable, Y_i, is an indicator of: (i) formal employment two to five months after the intervention; (ii) formal employment five to eight months after the intervention; (iii) monthly average wage two to five months after the intervention; (iv) monthly average wage five to eight months after the intervention; (v) and formal employment in a large firm (more than twenty workers) after the intervention. Since the randomization occurred at the level of the service provider, we include service-provider fixed effects θ_i.

We also estimate the following specification with covariates to control for any remaining imbalances between the treatment and control group:

$$Y_i = \alpha + \rho \text{Assigned}_i + \theta_i + \beta X_i + \eta_i \tag{2}$$

where the vector X_i in this specification includes baseline controls including gender, age, level of education (dummies for elementary, high school, technical education, and college, in which no schooling is the excluded category), and marital status. To improve precision of the estimates and control for any baseline imbalances, the model also includes pre-treatment characteristics: (i) the baseline score in Colombia´s official proxy means test to target social spending (SISBEN), which is an indicator of poverty and socioeconomic status, (ii) labor history indicators that measure the number of months with formal employment and employment in a large firm 3 years before program participation and (iii) an indicator that shows if the person had a formal job 6 months before the program. The sensitivity of the estimates to these three variables is examined. We also control for regional fixed effects.

We estimate fully saturated models for different groups of workers to examine potential heterogeneous effects across demographic characteristics (gender, age and educational attainment) and types of service providers (private firms, and employment and training agencies). Also, we estimated separate models for the groups with special premia in the program: women older than 40 years old, graduates of the program *Jóvenes en Acción*, and people with disabilities. Finally, we run separate models based on the pre-treatment experience to identify potential differences in the impacts across sectors that were more, or less, affected by the pandemic.

The coefficients ρ in these regressions can be interpreted as an intention-to-treat (ITT) effects since there was not full take up among those assigned to treatment. Given the low take-up of treatment, ITT estimates can be interpreted as a lower bound of the average treatment effect (ATE).

Administrative delays likely means that those who made it into the *Empleate* program were those who could not get jobs elsewhere, which means that there was negative selection. On the other hand, given the strong incentives for service providers from the P4P program, it is possible (and consistent with evidence provided below) that creaming of applicants and positive selection occurred. However, the downward bias from ITT parameters and the administrative delays probably outweigh the upward bias from creaming.

5. Data Description

5.1. Baseline Data

a. Initial Survey

Individuals interested in applying to the program were asked to fill a short baseline survey with 33 questions in total. This survey collected basic information on name, location, type of identification document, date of birth, department (or state) and municipality, urban or rural location, gender, marital status, ethnicity, number of children, level of education, current labor situation, work experience, and type of support required to obtain a job. While 19,889 completed this baseline questionnaire of the 25,770 total individuals randomized, there were 5,881 who did not complete this survey.

b. SISBEN III and SISBEN IV

We merged the baseline information with information from the unified vulnerability assessment and identification system for social programs collected by the National Planning Department, known as SISBEN (Sistema de Selección de Beneficiarios para Programas Sociales) in Colombia. SISBEN III was collected between 2013 and 2018 and the SISBEN IV from 2018 to the present. The data collected by SISBEN includes detailed information about the total number of people living in the household, number of children, marital status, level of education, employment, salary, as well as living conditions in the household (including access to clean water, electricity and sanitation). We use SISBEN data to complement baseline information (gender, marital status, age, level of education) for the 5,881 individuals that did not complete the baseline survey. We also use the score in SISBEN as a measure of poverty and socioeconomic status.

c. PILA

We merge the baseline information with administrative information from the Colombian social security records, which comes from the Integrated Social Security Form (known as PILA in Spanish) to examine the impacts of *Empléate* on formal employment after participation in the program. These records provide information on formal employment (as measured by contributions to social security) and firm size. Indicators of formal employment take the value of 1 if the person reports social security contributions 2 to 5 months or 5 to 8 months after applying to the program, and zero if the person signed up for the program and does not make contributions during these periods. Historic information, three years prior to program participation, is also available and can be used to include exogenous covariates in the estimations.

d. Follow-up Survey

A follow-up survey was collected by the surveying company SEI between August and September of 2021 to a random sample of individuals assigned to treatment and control. Among those assigned to the treatment, the survey oversampled those who participated in the program (1,615 were surveyed in the treatment group and 1,615 in the control group). The survey included information on sociodemographic characteristics, level of education and employment history. Importantly, the survey included questions about the services provided by the *Empléate* program and the characteristics and participant perceptions of the services. The survey includes questions about how the participants learned about the program; when and how they applied; what their expectations were; what services they received, and whether they were placed into a job through the program. Due to the small sample size, data from this survey is not meant to be used in the estimations but rather, to inform our study on the mechanisms that may explain the impacts from the program. For example, the survey is informative about the type and number of services received by the beneficiaries. The survey was also used to analyze employment characteristics of the treatment and control group after the intervention.

5.2. Descriptive Statistics

Table 3 shows descriptive statistics of baseline characteristics of observations in the control group (Column 1) and the treatment group (Column 2), respectively. The bottom row shows the number of observations: eighty-five percent of the sample was assigned to the treatment. Column 3 reports the estimated difference in baseline characteristics between treatment and control groups and Column 4 shows p-values for a difference in means t-test. Since randomization was stratified within service providers, we estimate these differences in baseline characteristics using service providers fixed effects.

Overall, the two samples are balanced, indicating that the randomization worked well. The only exceptions are the indicators for higher education and age. The control group has a slightly higher probability (0.017 percentage points) of having a higher education degree at a 5% significance level. The control group is also slightly younger, being on average 0.36 years younger than the treatment group at a 10% significance level.

Table 3 shows that 24% of those in the treatment and control groups are male. The average age is 36 years old and 29% are married. Forty four percent in the treatment group and 45.7% in the control group have a high school degree. Overall, 37% have a technical education degree, and only 8.5% of the control group and 9% of the treatment group have a college education. Given that everyone in the sample qualifies for social programs provided by the government, it includes only households classified as poor or vulnerable according to SISBEN, Colombia's proxy mean tests. The average SISBEN score for the control group as well as the treatment group is 21.6 points.[11] A small share of approximately 6% beneficiaries had participated in the *Jóvenes en Acción* program.

[11] SISBEN cutoff scores for program participation were: (i) for main cities: 0 to 41.74; (ii) for the rest of the urban population: 0 to 45.47; and (iii) for rural areas: 0 to 36.83.

We also use administrative data from social security records in PILA to build variables that describe formal labor history. Table 3 shows that about a third of the sample worked in a formal job in the six months before the program. Around 86% worked at least once in a large formal firm (with more than 20 employees) in the three years before the treatment. And on average they contributed only 7 months to social security in the three years before the treatment. Yet, Columns 3 and 4 show that these labor histories are very similar among the treatment and control groups.

Table 3. Descriptive statistics by random assignment

Variable	**Randomly assigned to control, Zi=0**	**Randomly assigned to treatment, Zi=1**	**Difference in means (1)-(2)**	**P-value difference in means T-test (1)-(2)**
Male	0.243	0.245	-0.001	0.845
	[0.007]	[0.003]		
Age	35.916	36.275	-0.359*	0.081*
	[0.188]	[0.080]		
Married	0.298	0.295	0.003	0.726
	[0.007]	[0.003]		
Primary education or less	0.040	0.043	-0.003	0.414
	[0.003]	[0.001]		
High school education	0.457	0.440	0.017**	0.045*
	[0.008]	[0.003]		
Technical Education	0.372	0.379	-0.008	0.370
	[0.008]	[0.003]		
College education or more	0.085	0.090	-0.005	0.309
	[0.004]	[0.002]		
SISBEN score	21.671	21.597	0.074	0.740
	[0.205]	[0.086]		
Former participant of Jóvenes en Acción program	0.060	0.057	0.003	0.507
	[0.004]	[0.002]		
Employed 6 months before treatment	0.293	0.280	0.013	0.213
	[0.010]	[0.004]		
Big formal employer (with more than 20 employees 3 years before treatment	0.874	0.881	-0.007	0.356
	[0.007]	[0.003]		
# of months contributing to social security before treatment (3 years before treatment)	6.907	6.680	0.226	0.162
	[0.151]	[0.062]		
N	3,842	21,928		

6. Impact of *Empléate*

Table 5 shows ITT estimates of the effects of the program on employment outcomes which are likely lower bounds of the average effect of the treatment on the treated. Column (1) shows estimates of participating in the program on the probability of having a formal job, measured as contributions to social security in administrative data, two to five months after the treatment. Column (2) also reports the impact on the probability of formal employment, five to eight months after the treatment. Column (3) and (4) show, respectively, the impact on monthly average wage two to five months and five eight months after program application. Column (5) shows the impact on the probability of working in a big firm after program application. The table has 12 panels, the first of which shows estimates for all the sample. The remaining panels show estimates for the samples of women, men, individuals younger than 30 years old, individuals aged 30 or more, women aged 40 or more, people with high school, people without high school, people with college, people without college, beneficiaries that received services from training and employment agencies, beneficiaries that received services from private firms with training programs, people with work experience in sectors that were most affected by the pandemic (tourism, service and commerce), and those with work experience in other sectors.

There are no short-run effects (i.e., two to five months after the treatment), which could be explained by locking-in effects or delays in social security registration. Impacts are only observed on the probability of being employed five to eight months after the treatment. This may suggest that the financial incentives that were given to service providers to retain beneficiaries in formal jobs five months after the treatment had a significant impact. Program participants were 2.3 percentage points more likely to have formal employment five to eight months after the treatment, which is equivalent to a 9% increase with respect to the control group baseline. These effects were driven by greater impacts on people with work experience in sector less affected by Covid-19 (22%), men (17%), people with high school and no college (10%), and those receiving services from private firms (12%).

Also, the program shows no effects on wages. Only people who were randomly assigned to the treatment group that participated with private firms as service providers, increased their monthly average wage five to eight months after program application by COP 20,394 (US$ 5.4)[12] or 14% compared to those randomly assigned to the control group. And people with work experience in sectors less affected by the pandemic raised their monthly average wage in this period by COP 24,013 (US $ 6.4) [13] or 17% compared to the control group. These effects are statistically significant at a 10% level.

There are also effects of the program on the probability of working in a big firm (more than twenty workers). Program participants were 2.6 percentage points more likely to work in a big firm from 1 to 8 months after the program, which is 3.2% greater than the control group baseline. The effects were driven mostly by men (8%), people with work experience in sectors less affected by the pandemic (8%), people without college (4%) and people older than 30 (5%).

Finally, the program does not show statistically significant effects on any of the employment outcomes analyzed for the groups that had differential incentives for placement and retention such as women older than 40 years old, graduates of the program *Jóvenes en Acción,* and people with disabilities.

[12] Value in dollars is estimated using the official exchange rate in September 1st,2020 (i.e., 3745 COP per dollar).
[13] Value in dollars is estimated using the official exchange rate in September 1st,2020 (i.e., 3745 COP per dollar).

Table 5. ITT estimates of the effect of the program on employment outcomes

	(1) Employed in the formal sector 2-5 months after the treatment	(2) Employed in the formal sector 5-8 months after the treatment	(3) Monthly average wage in the formal sector 2-5 months after the treatment	(4) Monthly average wage in the formal sector 5-8 months after the treatment	(5) Employed in a big firm (more than 20 workers)
A. All					
ITT estimate	0.006	0.023**	2,271	11,784	0.026*
	(0.010)	(0.012)	(5,684)	(7,911)	(0.015)
Observations	12,198	9,001	12,198	9,001	4,764
Control group mean	0.299	0.260	124,147	137,070	0.808
B. Women					
ITT estimate	0.006	0.015	-278	6,371	0.012
	(0.012)	(0.013)	(6,509)	(8,909)	(0.017)
Observations	8,783	6,638	8,783	6,638	3,215
Control group mean	0.283	0.248	119,637	132,107	0.829
C. Men					
ITT estimate	0.009	0.050*	9,992	25,818	0.058**
	(0.021)	(0.025)	(11,591)	16,986)	(0.028)
Observations	3,415	2,363	3,415	2,363	1,549
Control group mean	0.340	0.296	136,010	150,973	0.762
D. 30 or younger					
ITT estimate	0.006	0.024	9,773	10,258	0.012
	(0.016)	(0.016)	(8,623)	(10,839)	(0.021)
Observations	4,931	4,797	4,931	4,797	2,049
Control group mean	0.337	0.269	130,540	142,666	0.845
E. Older than 30					
ITT estimate	0.009	0.026	-2,066	13,521	0.039*
	(0.013)	(0.017)	(7,531)	(11,622)	(0.021)
Observations	7,267	4,204	(7,267)	(4,204)	2,715
Control group mean	0.271	0.250	119,440	130,323	0.778

F. Women older than 40					
ITT estimate	0.024	0.044	1,239	12,296	0.046
	(0.019)	(0.029)	(10,308)	(17,857)	(0.032)
Observations	3,387	1,362	3,387	1,362	1,146
Control group mean	0.230	0.206	105,822	111,070	0.779
G. People with high school					
ITT estimate	0.008	0.027**	2,152	12,559	0.020
	(0.011)	(0.012)	(5,870)	(8,159)	(0.015)
Observations	11,486	8,585	11,486	8,585	4,543
Control group mean	0.302	0.259	126,156	138,015	0.814
I. People without high school					
ITT estimate	-0.012	-0.119	1,119	-45,172	0.340***
	(0.058)	(0.087)	(36,018)	(51,170)	(0.115)
Observations	364	186	364	186	122
Control group mean	0.224	0.345	81,756	169,369	0.565
J. People with college					
ITT estimate	0.029	0.009	4,308	5,030	-0.022
	(0.032)	(0.036)	(19,784)	(29,255)	(0.052)
Observations	1,474	1,089	1,474	1,089	653
Control group mean	0.340	0.288	166,740	174,413	0.776
K. People without college					
ITT estimate	0.007	0.027**	2,959	11,748	0.035**
	(0.011)	(0.013)	(6,005)	(8,264)	(0.015)
Observations	10,376	7,682	10,376	7,682	4,012
Control group mean	0.295	0.258	119,222	133,895	0.810
L. People receiving services from training and employment agencies					
ITT estimate	-0.009	0.013	2,650	-15	0.028
	(0.014)	(0.016)	(7,750)	(10,206)	(0.022)
Observations	5,848	4,615	5,848	4,615	2,027
Control group mean	0.265	0.246	103,890	130,800	0.810

M. People receiving services from private firms					
ITT estimate	0.018	0.032*	778	20,394*	0.025
	(0.015)	(0.017)	(8,285)	(12,226)	(0.020)
Observations	6,350	4,386	6,350	4,386	2,737
Control group mean	0.330	0.275	142,335	143,631	0.806
N. Graduates from Jóvenes en Acción					
ITT estimate	-0.000	0.052	10,678	47,266	0.062
	(0.038)	(0.042)	(20,570)	(28,769)	(0.047)
Observations	907	711	907	711	390
Control group mean	0.398	0.294	173,040	146,028	0.763
O. People with disabilities					
ITT estimate	-0.032	0.181	24,389	85,997	-0.062
	(0.127)	(0.139)	(74,552)	(83,848)	(0.234)
Observations	132	111	132	111	38
Control group mean	0.308	0.077	115,418	34,943	0.600
P. People with experience in sectors most affected by $Covid19^1$					
ITT estimate	-0.007	0.006	3,998	-1,327	-0.008
	(0.018)	(0.020)	(10,032)	(13,198)	(0.025)
Observations	3,958	3,307	3,958	3,307	1,562
Control group mean	0.318	0.270	123,163	142,000	0.864
Q. People with experience in sector less affected by $Covid19^2$					
ITT estimate	0.028	0.057***	3,817	24,013*	0.059**
	(0.018)	(0.019)	(9,604)	(13,334)	(0.025)
Observations	4,170	3,467	4,170	3,467	1,756
Control group mean	0.320	0.260	139,945	144,771	0.754

Note: ITT estimates of the effect of being selected to participate on the program on formal employment, as measured by social security records, and on working on a big firm (more than 20 workers). All estimates are derived from regressions that include age, gender, marital status, level of education, SISBEN score, formality six months before program participation, having worked on a big firm three years before program participation and total number of registers in PILA three years before program participation, service provider fixed effects and regional fixed effects. Robust standard errors *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

1. Sectors most affected by Covid-19 include services, commerce, hotels, restaurants, entertainment, and recreation.
2. Sectors less affected by Covid-19 include construction, education, health, scientific and technical activities, manufacturing, agriculture, financial activities, transport, water distribution, communications, realtor activities, mining, vehicle reparation, among others.

Table 6, Table 7, and Table 8 present robustness tests of the stability of the ITT estimated effects when adding and removing regressors from the main specification for employment, monthly wage, and firm size, respectively. Specifications in Column A only include socioeconomic controls; those in Column B add number of records in PILA three years before program application; those in Column C add employment in a big firm three years before program application; those in Colum D add formality six months before program application; Column E add service provider fixed effects; and Column F include regional fixed effects. Column F is our preferred specification.

Table 6, Table 7, and Table 8 show that results found in Table 5 are robust to different specifications. There is no statistically significant effect of the program on short-run formal employment (two to five months after program participation) in any of the six panels. The statistically significant effect of the program on medium-run formal employment (five to eight months after program participation) is robust to different specifications (Panels B, C, D, E and F). Additionally, there is no statistically significant effect of the program on short-run monthly average wage (two to five months after program participation) in any specification. Medium-run monthly average wage shows a statistically significant effect only on panel B, but not in the rest of the panels. Also, the statistically significant effect of the program on the probability of working in a big firm is robust to all specifications (Panel A to D).

Table 6. Robustness tests for ITT estimates of the effect of the program on formal employment

	(1) Employed in the formal sector 2-5 months after the treatment						**(2) Employed in theformal sector 5-8 months after the treatment**					
	(A)	(B)	(C)	(D)	(E)	(F)	(A)	(B)	(C)	(D)	(E)	(F)
Eligible	-0.003 (0.007)	0.004 (0.006)	0.006 (0.010)	0.006 (0.010)	0.007 (0.010)	0.006 (0.010)	0.012 (0.008)	0.021*** (0.007)	0.023* (0.012)	0.023** (0.012)	0.023** (0.012)	0.023** (0.012)
# of observations	23,910	23,910	12,199	12,199	12,199	12,198	18,015	18,015	9,002	9,002	9,002	9,001
Mean control	0.193	0.193	0.299	0.299	0.299	0.299	0.167	0.167	0.260	0.260	0.260	0.260
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of records in PILA 3 years before treatment	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Employed in a big firm 3 years before treatment	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Formality six months before treatment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Service provider fixed effects	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Regional fixed effects	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Note: ITT estimates of the effect of being selected to participate on the program on formal employment, as measured by social security records. Socioeconomic controls include age, gender, marital status, level of education and SISBEN score. Robust standard errors *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table 7. Robustness tests for ITT estimates of the effect of the program on monthly wage

	(3) Monthly average wage in the formal sector 2-5 months after the treatment						**(4) Monthly average wage in the formal sector 5-8 months after the treatment**					
	(A)	(B)	(C)	(D)	(E)	(F)	(A)	(B)	(C)	(D)	(E)	(F)
Eligible	-1,569 (3,627)	1,824 (3,295)	1,924 (5,695)	2,142 (5,694)	2,333 (5,684)	2,271 (5,684)	4,562 (4,999)	9,698** (4,614)	11,534 (7,918)	11,676 (7,918)	11,575 (7,906)	11,784 (7,911)
# of observations	23,910	23,910	12,199	12,199	12,199	12,198	18,015	18,015	9,002	9,002	9,002	9,001
Mean control	77,926	77,926	124,147	124,147	124,147	124,147	86,989	86,989	137,070	137,070	137,070	137,070
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of records in PILA 3 years before treatment	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Employed in a big firm 3 years before treatment	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Formality six months before treatment	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Service provider fixed effects	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Regional fixed effects	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Note: ITT estimates of the effect of being selected to participate on the program on monthly average wage. Socioeconomic controls include age, gender, marital status, level of education and SISBEN score. Robust standard errors *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table 8. Robustness tests for ITT estimates of the effect of the program on firm size

	(5) Employed in a big firm (more than 20 workers)					
	(A)	(B)	(C)	(D)	(E)	(F)
Eligible	0.023*	0.025*	0.027*	0.026*	0.026*	0.026*
	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
# of observations	6,028	6,028	4,765	4,765	4,765	4,764
Mean control	0.808	0.808	0.808	0.808	0.808	0.808
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of records in PILA 3 years before treatment	No	Yes	Yes	Yes	Yes	Yes
Employed in a big firm 3 years before treatment	No	No	Yes	Yes	Yes	Yes
Formality six months before treatment	No	No	No	Yes	Yes	Yes
Service provider fixed effects	No	No	No	No	Yes	Yes
Regional fixed effects	No	No	No	No	No	Yes

Note: ITT estimates of the effect of being selected to participate on the program on working on a big firm (more than 20 workers). Socioeconomic controls include age, gender, marital status, level of education and SISBEN score. Robust standard errors *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

7. Program Design and Implementation

The impact evaluation was complemented with a descriptive analysis using a follow up survey and interviews with program managers. This helped us understand the mechanisms underlying the observed impacts of the *Empléate* as well as implementation challenges.

First, the surveys provide information about the combination of services delivered. Interestingly, even though a potential benefit of P4P contracts is that, by focusing on outcomes instead of inputs, they give providers more flexibility and a scope for innovation in service provision, there was not much variation in delivery across service providers. Although *Empléate´s* contractual arrangements gave service providers the opportunity to deliver a wide range of specialized services (such as training and certification, counseling to improve the labor profile, support for transportation, dressing or childcare), according to the survey, most beneficiaries (85%) didn´t receive them. Providers mostly delivered basic intermediation services, such as registration in the employment service information system, short job search workshops, and psycho-technical tests. (Appendix 1). Thus, the program was mainly a job placement intervention rather than one that developed or certified skills.

Administrative data also shows some evidence of creaming. When comparing individuals that were randomly assigned to the program and those that were effectively treated, the latter are 10 percentage points more likely to be male, are 10 years older, and are 9 and 5 percentage points more likely of having a technical and university degree, respectively. Beneficiaries also have a greater chance of having a recent formal job, previous experience in a formal firm and a longer history of contributions to social security. Thus, beneficiaries in general have attributes that are positively related to job readiness (Appendix 2).

As was shown in Table 5, *Empléate* had a greater effect on the employability of individuals treated by private firms than of those treated by traditional employment and training agencies. Evidence from administrative data suggests that, to some extent, these greater effects could have been driven by more intense screening by private firms: they are older, have a much longer history of formal jobs and a higher socioeconomic status (as measured by the SISBEN score) than those treated by traditional employment and training agencies[14] (Appendix 3). Also, evidence from the follow up survey shows that traditional employment and training agencies gave at least one specialized service to a higher proportion of beneficiaries compared to firms (17% vs 10%). However, when analyzing each type of specialized service, there were no statistically significant

[14] There are also differences in other attributes, such as education and gender, but the magnitudes are not very large

differences between the two types of service providers. And, when comparing the provision of basic services between firms and employment and training agencies, firms offered personalized or group interviews to a higher proportion of beneficiaries (30% vs 25%), whereas employment and training agencies offered self-employment tools workshops and job search tools workshops to a higher share of beneficiaries (19% vs 13% and 35% vs 26%) (Appendix 4). Thus, the greater impact of *Empleate* when training was provided by firms does not seem to be explained by type of services they offered, but with a more intense screening.

This evidence of creaming in *Empléate* may be revealing a perverse incentive to restrict the intake to more employable participants. Such incentives have been documented in other P4P designs and may have been prevalent in *Empléate* due to its design features. Given that 100% of the payment to service providers was contingent on placement and retention, the risk transfer to service providers may have been too large, reducing their capacity to supply the right services and creating incentives to intense creaming. This issue underscores the importance of ensuring an adequate allocation of financial risk on performance-based contracts.

However, in certain contexts, some creaming may be efficient to the extent that it ensures that services are only delivered to participants who can benefit from the program. This may have been the case in *Empléate*, to the extent that, due to mobility restrictions amidst the pandemic, the program had large implementation difficulties which delayed the operation. Service providers that had identified job opportunities weren´t able to place beneficiaries on a timely manner. They also had difficulties reaching potential beneficiaries during the lockdown. Thus, the intensity of services provided was limited and the program may have only been efficient to those with lower employability barriers.

8. Conclusion

This paper presents the findings from an experimental evaluation of a P4P employment program implemented in Colombia during the COVID-19 recession, *Empléate*. By contracting out employment services using P4P, this program adopted an innovative service delivery model, uncommon in Latin America. Service providers only received payments contingent on employment outcomes; namely, formal job placement upon program completion and formal employment for at least three and five months after the intervention.

The evaluation is based on a randomized control trial conducted from September 2020 to April 2021. Individuals who qualified for the program and expressed interest in the program were randomly assigned to a treatment or a control group and those assigned to the treatment group could receive a variety of services to help them get jobs and stay in their jobs. However, as is often the case in voluntary programs, there was less than full compliance. A very large number of workers assigned to treatment did not take up the treatment.

ITT estimates show the program had a positive impact on the probability of being formally employed five to eight months after the intervention. This impact was of 22% for people with experience in sectors less affected by the pandemic, 17% for men, and of 10% for those without college. This suggests that the financial incentives that were given to service providers to retain beneficiaries in formal jobs five months after the treatment were particularly effective. Also, that the negative effects the pandemic had on specific economic sectors traduced to the results of the program, since people with work experience on these areas did not show any impact on employment outcomes. Moreover, there was a positive effect of 14% on the monthly average wage five to eight months after program application for people that participated with private firms as service providers, and of 17% for people with work experience in sectors less affected by Covid-19. There were also positive impacts on the probability of working on a large formal sector firm for men (8%), people with work experience in sectors less affected by the pandemic (8%), people without college (5%) and individuals 30 years or older (4%). There is no evidence of employment effects for women and people without secondary education. Given the low compliance levels, these estimated impacts likely underestimate average treatment effects.

Moreover, administrative data provides suggestive evidence of creaming: beneficiaries were more likely to be male, had a higher educational attainment and a longer history on the formal labor market than average eligible participants. Creaming may have been induced by two factors. First, due to mobility restrictions amidst the pandemic, the program had large implementation difficulties which delayed its operation. Service providers that had identified job opportunities were not able to start providing services on a timely manner and had difficulties reaching potential beneficiaries. The implementation schedule was also affected by administrative delays. This reduced the possibility of providing a wider array of services. Indeed, data from a follow up survey conducted to a random sample of eligible individuals to the program (assigned to the treatment and control groups), shows that even though service providers could choose to provide a range of services that included training and skill certification, in most cases the intervention was limited to basic intermediation services such as registry in the employment service and short job-search assistance workshops. In this context, creaming may have been efficient, ensuring that this low-intensity intervention was delivered only to participants who could benefit from the services. Second, and maybe more importantly, the design of the P4P incentives was such that the service providers may have assumed too much risk. Given that 100% of the payment was contingent on placement and retention, the risk transferred to providers may have been too large, reducing their capacity to supply the right services and creating incentives for intense creaming. This issue underscores the importance of ensuring an adequate allocation of financial risk on performance-based contracts.

Nevertheless, many design features of *Empléate* are promising and it had positive effects on keeping individuals with low levels of education in formal jobs months after the intervention. Given the high levels of informality and instability in Colombia's labor market and given the dramatic impact the pandemic had on formal jobs, this is a commendable achievement.

References

Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S., Quinn, S. 2016a. "Anonymity or Distance? Experimental Evidence on Obstacles to Youth Employment Opportunities". Mimeo. Stanford University.

Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S., Quinn, S. Shilpi, F. 2016b. "All the fun of the (job) fair: Matching firms and workers in a field experiment in Ethiopia". Mimeo.

Alzúa, M.L., Katzkowicz, N. 2021. "Pay for Performance for Prenatal Care and NewbornHealth: Evidence from a Developing Country," *World Development*, 141:105-385. https://doi.org/10.1016/j.worlddev.2020.105385.

Amuedo-Dorantes, C., Mach, T. 2003. "Performance Pay and Fringe Benefits: Work Incentives or Compensating Wage Differentials?" *International Journal of Manpower*, 24: 673–698. https://doi.org/10.1108/01437720310496157.

Anderson, K.H., Burkhauser, R.V., Raymond, J.E. 1993a. "The Effect of Creaming on Placement Rates under the Job Training Partnership Act," *Industrial and Labor Relations Review*, 46: 613–624. https://doi.org/10.1177/001979399304600402.

Anderson, K.H., Burkhauser, R.V., Raymond, J.E. 1993b. "The Effect of Creaming on Placement Rates under the Job Training Partnership Act," *Industrial and Labor Relations Review*, 46: 613–624. https://doi.org/10.1177/001979399304600402.

Ashraf, N., Bandiera, O., Lee, S.S. 2014a. "Awards Unbundled: Evidence from a Natural Field Experiment," *Journal of Economic Behavior & Organization*, 100: 44–63. https://doi.org/10.1016/j.jebo.2014.01.001.

Ashraf, N., Bandiera, O., Jack, B.K. 2014b. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery," *Journal of Public Economics*, 120: 1–17. https://doi.org/10.1016/j.jpubeco.2014.06.014.

Attanasio, O, A. Kugler and C. Meghir. 2011. "Subsidizing Vocational Training forDisadvantaged Youth in Colombia: Evidence from a Randomized Trial,", *American Economic Journal: Applied Economics*, 3(3): 188-220. https://www.aeaweb.org/articles?id=10.1257/app.3.3.188.

Balafoutas, L., Dutcher, E.G., Lindner, F., Ryvkin, D. 2017. "The Optimal Allocation of Prizes in Tournaments of Heterogeneous Agents," *Economic Inquiry*, 55: 461–478. https://doi.org/10.1111/ecin.12380.

Bandiera, O., Barankay, I., Rasul, I. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *Quarterly Journal of Economics*, 120: 917–962. https://doi.org/10.1093/qje/120.3.917.

Bandiera, O., Fischer, G. 2013. "Can "Good" HR Practices be Exported? Evidence from a Field Experiment in Ghana," *IGC*, 44.

Barrera-Osorio, F., Gonzalez, K., Lagos, F., Deming, D.J. 2020. "Performance Information in Education: An Experimental Evaluation in Colombia," *Journal of Public Economics*, 186: 104185. https://doi.org/10.1016/j.jpubeco.2020.104185.

Barrera-Osorio, F., Kugler, A.D. and Silliman, M.I. 2020. "Hard and soft skills in vocational training: Experimental evidence from Colombia". NBER Working Paper No. 27548. Cambridge, MA. *National Bureau of Economic Research*.

Barrera-Osorio, F., Kugler, A., Silliman, M. 2021. "Job training through turmoil". NBER Working Paper No. 29565. *National Bureau of Economic Research.*

Barrera-Osorio, F., Raju, D. 2017. "Teacher Performance Pay: Experimental Evidence from Pakistan," *Journal of Public Economics*, 148: 75–91. https://doi.org/10.1016/j.jpubeco.2017.02.001.

Bastani, H., Goh, J., Bayati, M. 2019. "Evidence of Upcoding in Pay-for-Performance Programs," *Management Science*, 65: 1042–1060. https://doi.org/10.1287/mnsc.2017.2996.

Bell, S.H., Orr, L.L. 2002. "Screening (and Creaming?) Applicants to Job Training Programs: the AFDC Homemaker–home Health Aide Demonstrations," *Labour Economics*, 9: 279– 301. https://doi.org/10.1016/S0927-5371(02)00006-4.

Benkert, J.-M., Letina, I. 2020. "Designing Dynamic Research Contests," *American Economic Journal: Microeconomics*, 12: 270–289. https://doi.org/10.1257/mic.20180263.

Beuermann, D.W., Bottan, N.L., Hoffmann, B., Jackson, C.K. and Cossio, D.A.V. 2021. "Does Education Prevent Job Loss During Downturns? Evidence from Exogenous School Assignments and COVID-19 in Barbados". NBER Working Paper No. 29231. Camabridge, MA. *National Bureau of Economic Research.*

Bimpikis, K., Ehsani, S., Mostagir, M. 2019. "Designing Dynamic Contests," *Operations Research*, 67: 339–356. https://doi.org/10.1287/opre.2018.1823

Booth, A.L., Frank, J. 1999. "Earnings, Productivity, and Performance-Related Pay," *Journal of Labor Economics*, 17: 447–463. https://doi.org/10.1086/209927

Brehm, M., Imberman, S.A., Lovenheim, M.F. 2015. "Achievement Effects of Individual Performance Incentives in a Teacher Merit Pay Tournament," NBER Working Paper No. 21598. *National Bureau of Economic Research*. https://doi.org/10.3386/w21598.

Brownback, A., Sadoff, S. 2020. "Improving College Instruction through Incentives," *Journal of Political Economy*, 128: 2925–2972. https://doi.org/10.1086/707025.

Card, D., Kluve, J., Weber, A. 2017. "What works? A meta analysis of recent active labor market program evaluations". NBER Working Paper No. 21431. *National Bureau of Economic Research.*

Che, Y.-K., Gale, I. 2003. "Optimal Design of Research Contests," *American Economic Review*,93: 646–671. https://doi.org/10.1257/000282803322157025.

Chen, Y.P. 2009. "Cream-Skimmer or Underdog? Labor Type Selectivity, Pre-Program Wage, and Rural Labor Training Program Outcome," *IZA Discussion Papers,* 32.

Cohen, C., Kaplan, T.R., Sela, A. 2008. "Optimal Rewards in Contests," *RAND Journal of Economics*, 39: 434–451. https://doi.org/10.1111/j.0741-6261.2008.00021.x.

Courty, P., Marschke, G. 1998a. "Measuring Government Performance: Lessons from a Federal Job Training Program," SSRN Scholarly Paper No. ID 5182. *Social Science Research Network (SSRN)*.

Courty, P., Marschke, G. 1998b. "Measuring Government Performance: Lessons from a Federal Job Training Program," SSRN Scholarly Paper No. ID 5182. *Social Science Research Network (SSRN).*

Dammert, A., Galdo, J., Galdo, V. 2015. "Integrating mobile phone technologies into labor-market intermediation: a multi-treatment experimental design". IZA Journal of Labor and Development 4:11

Darden, M., McCarthy, I., Barrette, E. 2018. "Who Pays in Pay for Performance? Evidence from Hospital Pricing," NBER Working Paper No. 24304. National Bureau of Economic Research. https://doi.org/10.3386/w24304.

Duchoslav, J., Cecchi, F. 2019. "Do Incentives Matter when Working for God? The Impact of Performance-based Financing on Faith-based Healthcare in Uganda," *World Development*, 113: 309–319. https://doi.org/10.1016/j.worlddev.2018.09.011.

Ewing, B.T. 1996. "Wages and Performance-based Pay: Evidence from the NLSY," *Economics Letters*, 51: 241–246. https://doi.org/10.1016/0165-1765(95)00775-X

Farber, H.S. 2015. "Job loss in the Great Recession and its aftermath: US evidence from the displaced workers survey". NBER Working Paper No. 21216. *National Bureau of Economic Research.*

Figlio, D.N., Kenny, L.W. 2007. "Individual Teacher Incentives and Student Performance," *Journal of Public Economics*, 91: 901–914. https://doi.org/10.1016/j.jpubeco.2006.10.001.

Franklin, S. 2015. "Location, Search Costs and Youth Unemployment: A Randomized Trial of Transport Subsidies in Ethiopia", CSAE Working Paper WPS/2015-11.

Forslund, A., P. Fredriksson and J. Vikstrom. 2011. "What Active Labor Market Policy Works In a Recession?". Nordic Economic Policy Review 1: 171-207.

Fullerton, R.L., McAfee, R.P. 1999. "Auctioning Entry into Tournaments," *Journal of Political Economy*, 107: 573–605. https://doi.org/10.1086/250072.

Gneezy, U., Rustichini, A. 2000. "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115: 791–810. https://doi.org/10.1162/003355300554917.

Goltsman, M., Mukherjee, A. 2011. "Interim Performance Feedback in Multistage Tournaments: The Optimality of Partial Disclosure," *Journal of Labor Economics*, 29: 229–265. https://doi.org/10.1086/656669.

Goodman, S.F., Turner, L.J. 2013. "The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program," *Journal of Labor Economics*, 31: 409–420. https://doi.org/10.1086/668676.

Gupta, A. 2021. "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program," *American Economic Review*, 111: 1241–1283. https://doi.org/10.1257/aer.20171825.

Halac, M., Kartik, N., Liu, Q. 2017. "Contests for Experimentation," *Journal of Political Economy*, 125: 1523–1569. https://doi.org/10.1086/693040.

Hanushek, E.A., Schwerdt, G., Woessmann, L. and Zhang, L. 2017. "General education, vocational education, and labor-market outcomes over the lifecycle". Journal of Human Resources, 52(1), pp.48-87.

Hawkins, A.F. 1982. "H.R.5320 - 97th Congress (1981-1982): Job Training Partnership Act."

Heckman, J., Heinrich, C., Smith, J. 1997. "Assessing the Performance of Performance Standards in Public Bureaucracies," *The American Economic Review*, 87: 389–395.

Heckman, J.J., Smith, J.A. 2004. "The Determinants of Participation in a Social Program:Evidence from a Prototypical Job Training Program," *Journal of Labor Economics*, 22: 243– 298. https://doi.org/10.1086/381250.

Kauhanen, A. and Riukula, K. 2019. "The Costs of Job Loss and Task Usage". ETLA Working Paper No. 73. Helsinki, Finland. Etla Research.

Kluve, J. 2010. "The Effectiveness of European Active Labor Market Programs." Labour Economics 17: 904-918.

Kluve, J., Puerto, S., Robalino, D., Romero, J., Rother, F., Stoterau., Weidenkaff, F., Witte, M. 2016. "Do Youth Employment Programs Improve Labor Market Outcomes? A systematic Review". IZA Discussion Working Paper No. 10263. *Institute for the Study of Labor.*

Lechner, M., Wunsch, C. 2009. "Are Training Programs More Effective when Unemployment is

High?" Journal of Labor Economics 27 (4): 653-692.

Silliman, M. and Virtanen, H. 2019. "Labor market returns to vocational secondary education". ETLA Working Paper No. 65. Helsinki, Finland. Etla Research.

Heinrich, C.J., Choi, Y. 2007. "Performance-Based Contracting in Social Welfare Programs," *The American Review of Public Administration*, 37: 409–435. https://doi.org/10.1177/0275074006297553.

Hendricks, M.D. 2014. "Does it pay to pay teachers more? Evidence from Texas," *Journal of Public Economics*, 109: 50–63. https://doi.org/10.1016/j.jpubeco.2013.11.001.

Hofstede, G. 2006. "Dimensionalizing cultures, in: Online Readings in Psychology and Culture, Unit 2," *Center for Cross-Cultural Research*.

Hofstede, G., Minkov, M. 2013. "Values Survey Module 2013 Manual," *Geert Hofstede BV*, 17.

Khan, A.Q., Khwaja, A.I., Olken, B.A. 2019. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," *American Economic Review*, 109: 237–270. https://doi.org/10.1257/aer.20180277.

Khan, A.Q., Khwaja, A.I., Olken, B.A. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *The Quarterly Journal of Economics*, 131: 219–271. https://doi.org/10.1093/qje/qjv042.

Kireyev, P. 2020. "Markets for ideas: prize structure, entry limits, and the design of ideation contests," *The RAND Journal of Economics*, 51: 563–588. https://doi.org/10.1111/1756-2171.12325.

Koning, P., Heinrich, C.J. 2013. "Cream-Skimming, Parking and Other Intended and Unintended Effects of High-Powered, Performance-Based Contracts," *Journal of Policy Analysis and Management*, 32: 461–483. https://doi.org/10.1002/pam.21695.

Kugler, A., M. Kugler, J. Saavedra, and L. Herrera-Prada. 2022. "The Long-term Impacts and Spillovers of Training for Disadvantaged Youth," *Journal of Human Resources,* 57(1): 178-216.

Jensen, R. 2012. "Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India", Quarterly Journal of Economics 127 (2), 753–792.

Lazear, E.P. 2003. "Output-Based Pay: Incentives, Retention or Sorting?" *IZA Discussion Papers*, 761.

Lazear, E.P. 2000. "Performance Pay and Productivity," *American Economic Review*, 90: 1346–1361. https://doi.org/10.1257/aer.90.5.1346.

Lazear, E.P. 1986. "Salaries and Piece Rates," *The Journal of Business*, 59: 405–431.

Lazear, E.P., Rosen, S. 1981. “Rank-Order Tournaments as Optimum Labor Contracts,” *Journal of Political Economy*, 89: 841–864. https://doi.org/10.1086/261010.

Lemus, J., Marshall, G. 2021. “Dynamic Tournament Design: Evidence from Prediction Contests,” *Journal of Political Economy*, 129: 383–420. https://doi.org/10.1086/711762.

Lu, J. 2016. “The Performance of Performance-Based Contracting in Human Services: A Quasi-Experiment,” *Journal of Public Administration Research and Theory*, 26: 277–293. https://doi.org/10.1093/jopart/muv002.

Luo, R., Miller, G., Rozelle, S., Sylvia, S., Vera-Hernández, M. 2020. “Can Bureaucrats Really Be Paid Like Ceos? Substitution Between Incentives and Resources Among School Administrators in China,” *Journal of the European Economic Association*, 18: 165–201. https://doi.org/10.1093/jeea/jvy047.

Mas, A. 2006. “Pay, Reference Points, and Police Performance,” *The Quarterly Journal of Economics*, 121: 783–821. https://doi.org/10.1162/qjec.121.3.783.

Mbiti, I., Romero, M., Schipper, Y. 2019. “Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania,” NBER Working Paper No. 25903. *National Bureau of Economic Research*. https://doi.org/10.3386/w25903.

McKenzie, D. 2017. “How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence”, Policy Research Working Paper 8011. World Bank Group.

Mellor, J., Daly, M., Smith, M. 2017. “Does It Pay to Penalize Hospitals for Excess Readmissions? Intended and Unintended Consequences of Medicare's Hospital Readmissions Reductions Program,” *Health Econ*, 26: 1037–1051. https://doi.org/10.1002/hec.3382.

Miller, G., Babiarz, K.S. 2013. “Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs,” NBER Working Paper No. 18932. *National Bureau of Economic Research.* https://doi.org/10.3386/w18932.

Moldovanu, B., Sela, A. 2001. “The Optimal Allocation of Prizes in Contests,” *American Economic Review*, 91: 542–558. https://doi.org/10.1257/aer.91.3.542.

Morton, W.R. 2013. “Ticket to Work and Self-Sufficiency Program: Overview and Current Issues,” *Congressional Research Service*, 39.

Muralidharan, K., Sundararaman, V. 2011. “Teacher Performance Pay: Experimental Evidence from India,” *Journal of Political Economy*, 119: 39–77. https://doi.org/10.1086/659655.

Ngo, D.K.L., Bauhoff, S. 2021. “The medium-run and scale-up effects of performance-based financing: An extension of Rwanda's 2006 trial using secondary data,” *World Development*, 139: 105264. https://doi.org/10.1016/j.worlddev.2020.105264.

Norton, E.C., Li, J., Das, A., Chen, L.M. 2016. “Moneyball in Medicare,” NBER Working Paper No. 22371. *National Bureau of Economic Research*. https://doi.org/10.3386/w22371.

Olszewski, W., Siegel, R. 2020. “Performance-maximizing large contests,” *Theoretical Economics*, 15: 57–88. https://doi.org/10.3982/TE3588.

Prendergast, C. 1999. “The Provision of Incentives in Firms,” *Journal of Economic Literature*, 37: 7–63. https://doi.org/10.1257/jel.37.1.7.

Ryan, A.M., Krinsky, S., Maurer, K.A., Dimick, J.B. 2017. “Changes in Hospital Quality Associated with Hospital Value-Based Purchasing,” *New England Journal of Medicine*, 376: 2358–2366. https://doi.org/10.1056/NEJMsa1613412.

Shearer, B. 2004. “Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment,” *The Review of Economic Studies*, 71: 513–534.

Sisak, D. 2009. “Multiple-Prize Contests – the Optimal Allocation of Prizes,” *Journal of Economic Surveys*, 23: 82–114. https://doi.org/10.1111/j.1467-6419.2008.00557.x.

Stapleton, D., Livermore, G., Thornton, C., O'Day, B., Weathers, R., Harrison, K., O'Neil, S., Sama Martin, E., Wittenburg, D., Wright, D. 2008. “Ticket to Work at the Crossroads: A Solid Foundation with an Uncertain Future,” *Mathematica*.

Strack, P. 2016. “Risk-Taking in Contests: The Impact of Fund-Manager Compensation on Investor Welfare,” SSRN Scholarly Paper No. ID 2739177. *Social Science Research Network (SSRN)*. https://doi.org/10.2139/ssrn.2739177.

Taylor, C.R. 1995. “Digging for Golden Carrots: An Analysis of Research Tournaments,” *The American Economic Review*, 85: 872–890.

Xiao, J. 2016. “Asymmetric all-pay contests with heterogeneous prizes,” *Journal of Economic Theory*, 163: 178–221. https://doi.org/10.1016/j.jet.2015.12.006.

Appendix 1. Services provided in *Empléate*

Services received by beneficiaries	% from total beneficiaries
Basic Services	
Registry in the Public Employment Service Information System	78.9
Psycho-technical test	33.6
Personalized or group interview	26.9
Basic skills workshops	16.4
Self-employment tools workshops	17.1
Job search tools workshops	32.1
None	0.9
Specialized Services	
Improvement of labor profile (e.g., reconstruction of CVs; employability tools; document management support; psychosocial support)	13.3
Complementary supports (e.g., tools supply; dressing and hairdressing; transport; support for minor or third people care; support for disabilities; telecommuting and digital tools)	5.6
Training services (e.g., tailor training, technical certifications; specific skills strengthening)	8.4
Selection process Support (e.g., home visits; depth interviews; psychosocial tests; specific labor skills tests)	5.7

Appendix 2. Descriptive statistics by beneficiaries from the treatment group

Variable	**Non beneficiaries , Zi=0**	**Beneficiaries, Zi=1**	**Difference inmeans (1)-(2)**	**P-value differencein means T-test (1)-(2)**
Male	0.239	0.329	-0.090***	0.000***
	[0.003]	[0.012]		
Age	36.084	39.061	-2.977***	0.000***
	[0.082]	[0.323]		
Married	0.305	0.160	0.145***	0.000***
	[0.003]	[0.010]		
Primary education or less	0.043	0.045	-0.002	0.703
	[0.001]	[0.006]		
High school education	0.449	0.306	0.143***	0.000***
	[0.003]	[0.012]		
Technical Education	0.374	0.457	-0.083***	0.000***
	[0.003]	[0.013]		
College education or more	0.087	0.134	-0.048***	0.000***
	[0.002]	[0.009]		
SISBEN score	21.612	21.380	0.232	0.507
	[0.088]	[0.363]		
Former participant of Jóvenes en Acción program	0.056	0.083	-0.027***	0.000***
	[0.002]	[0.007]		
Employed 6 months before treatment	0.278	0.305	-0.028*	0.070*
	[0.004]	[0.015]		
Big formal employer (>20 employees) (3 years before treatment	0.879	0.899	-0.019*	0.077*
	[0.003]	[0.010]		
# of months contributing to social security before treatment (3 yearsbefore treatment)	6.511	9.143	-2.632***	0.000***
	[0.064]	[0.264]		
N	20,513	1,415		

Appendix 3. Descriptive statistics of eligible population by type of service provider

Variable	Firms, Zi=0	Training and Employment agencies Zi=1	Difference inmeans (1)-(2)	P-value differencein means T-test (1)-(2)
Male	0.225	0.261	-0.036***	0.000***
	[0.004]	[0.004]		
Age	38.370	34.237	4.134***	0.000***
	[0.109]	[0.094]		
Married	0.292	0.298	-0.006	0.291
	[0.004]	[0.004]		
Primary education or less	0.047	0.039	0.008***	0.001***
	[0.002]	[0.002]		
High school education	0.496	0.397	0.099***	0.000***
	[0.004]	[0.004]		
Technical Education	0.326	0.423	-0.096***	0.000***
	[0.004]	[0.004]		
College education or more	0.099	0.081	0.018***	0.000***
	[0.003]	[0.002]		
SISBEN score	25.236	18.491	6.745***	0.000***
	[0.106]	[0.107]		
Former participant of Jóvenes en Acción program	0.064	0.053	0.011***	0.000***
	[0.002]	[0.002]		
Employed 6 months before treatment	0.290	0.271	0.018**	0.018**
	[0.005]	[0.006]		
Big formal employer (>20 employees) (3 years before treatment	0.883	0.877	0.006	0.298
	[0.004]	[0.004]		
# of months contributing to social security before treatment (3 yearsbefore treatment)	7.969	5.596	2.372***	0.000***
	[0.089]	[0.073]		
N	12,380	13,737		

Appendix 4. Descriptive statistics of services provided to beneficiaries by type of service provider – Information from follow up survey

Variable	Firms, Zi=0	Training and Employment agencies Zi=1	Difference in means (1)-(2)	P-value differencein means T-test (1)-(2)
Received any basic service	0.995	0.990	0.005	0.381
	[0.004]	[0.004]		
Received any specialized service	0.106	0.170	-0.065***	0.004***
	[0.016]	[0.014]		
Number of basic services received	1.942	2.104	-0.162*	0.091*
	[0.071]	[0.057]		
Number of specialized services received	0.237	0.377	-0.139**	0.013**
	[0.040]	[0.034]		
Number of basic and specialized services received	2.179	2.481	-0.301**	0.019**
	[0.093]	[0.077]		
B: Registry in the Public Employment Service Information System	0.773	0.797	-0.024	0.342
	[0.022]	[0.014]		
B: Psycho-technical test	0.330	0.339	-0.009	0.758
	[0.024]	[0.017]		
B: Personalized or group interview	0.306	0.251	0.055**	0.046**
	[0.024]	[0.016]		
B: Basic skills workshop	0.142	0.175	-0.033	0.158
	[0.018]	[0.014]		
B: Self-employment tools workshop	0.127	0.192	-0.066***	0.005***
	[0.017]	[0.014]		
B: Job search tools workshop	0.264	0.349	-0.085***	0.003***
	[0.023]	[0.017]		
S: Improvement of labor profile	0.850	0.908	-0.058	0.295
	[0.057]	[0.025]		
S: Complementary supports	0.475	0.344	0.131	0.134
	[0.080]	[0.042]		
S: Training services	0.550	0.573	-0.023	0.803
	[0.080]	[0.043]		
S: Selection process Support	0.375	0.389	-0.014	0.872
	[0.078]	[0.043]		
N	379	770		

Note: Services that have a B before the description, means that they are basic services, the ones that have an S before the description means that they are specialized services.