Does Paying for Jobs for the vulnerable during the COVID-19 Crisis Pay? Experimental Evidence of a Pay-for-Performance Employment Program in Colombia

Maria Fernanda Gómez Gerena
Carolina González-Velosa
Adriana Kugler

May 2022

Executive Summary

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

- The program was administered by Prosperidad Social, the government agency in charge of social protection programs in Colombia. The program was implemented via an innovative P4P arrangement. Prosperidad Social contracted service providers to provide a flexible combination of services that could include intermediation, job placement, case management, training services, transportation allowances and made payments to service providers conditional on outcomes in formal employment. Specifically, providers received a payment if the beneficiary was placed in a formal job right after the intervention. Additional premia were paid if the participant remained on the job after three months and after five months, respectively. There were also additional premia paid 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.

- Twelve service providers were contracted to implement the program. Some were employment and training agencies, and others were large private firms with structured training programs. Due to operational difficulties and data availability, nine of the twelve service providers are included in this study.

- 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.

- However, as is often the case of voluntary programs, there was less than full compliance. A very 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 delayed provision of services, intense screening from program providers, and difficulties participating because of mandatory lockdowns due to the pandemic. However, non-compliance was mostly confined to the treatment group. Contamination among the control group was quite low: very few workers assigned to the control group found ways to access the services.

- The evaluation relies on baseline administrative data as well as administrative social security records to be able to follow the formal labor market experience of those who participated in the program. 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 that were 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, those who were assigned to the program were 9% more likely to have a formal job five to eight months after the treatment. This impact is higher for men, for whom the estimated effect is 17%, and among people without college, for whom the impact is of 10%. 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 on the monthly average wage in the short run (two to five months after program application) nor it has on monthly average wage in the medium run (five to eight months after the program application). However, people that participated with service providers that were firms, increased their monthly average wage in the medium run (five to eight months after program application) by COP 20,394 (US$ 5)[1] or 14% compared to those randomly assigned to the control group. This was the only group of interest that showed statistically significant 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) in 3%. These effects were larger for men (8%), people without college (5%) and individuals who are at least 30 years old (4%). When comparing impacts between different types of service providers, larger effects were observed among the subsample of beneficiaries that received the treatment from private firms.

- ITT estimates show no effect of the program on having a formal job, monthly average wage or working in a big firm for the three groups with additional premia paid for placement: women aged 40 or more, disadvantaged youth who participated in the program *Jóvenes en Acción* and people with disabilities

- The evaluation was complemented with a 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 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

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

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 the provision of services. Providers that had identified job opportunities were not able to start giving services on a timely manner and had difficulties reaching potential beneficiaries. The implementation schedule was also affected by administrative delays, which limited the ability to provide a wider set of 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 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 that benefitted from these 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 transfer 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.

1. Introduction

This paper presents the results of the experimental evaluation of the *Empléate* program, which was implemented in Colombia on September of 2020. The program was introduced in the middle of one of the worse recessions faced in 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, farther 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 unemployment reached 21% and was even higher among women and youth. Inactivity 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 introduce *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, resumepreparation assistance, help with wardrobe for interviews, soft skills training, technical training, transportation, post-placement assistance, etc. 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 questionnaire. 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, they were randomly assigned to the treatment. 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, most individuals assigned to the treatment did not participate in the program. On the contrary, there were no compliance issues in the control group: in general, those assigned to the control group did not participate in the program.

This paper shows estimated effects of *Empléate* on the labor market outcomes of individuals randomly assigned to the program from September 2020 to April 2021. The results 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 men (17%) and among people without college (10%). Also, there is a positive impact of 14% on monthly average wage five to eight months after program application, but only for people participating with private firms as service providers. 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%), individuals older than 30 (5%), and workers with no college (4%). There is no evidence that *Empléate* improved labor market outcomes of women and individuals younger than 30.

The rest of the report proceeds as follows. Section 2 provides a literature review of pay-for-performance programs. Section 3 describes the *Empléate* program. Section 4 describes the

randomization and the methodology used to estimate intention-to-treat effects as well as the administrative and survey data we use in the analysis. Section 5 presents the results of the overall impacts on employment, monthly wage, and firm size, as well as heterogeneous effects. In addition, this section explores the potential channels through which the impacts are working. Section 6 presents thecost-benefit analysis and Section 7 conclude.

2. Literature Review

2.1. Literature about Pay-for-Performance Programs

Economists have long studied Pay-for-Performance (P4P) programs as a way of attracting higher-skilled workers, incentivizing greater effort, and improving productivity (Prendergast, 1999; Booth and Frank, 1999). Much of the earlier research on P4P focused on the use of these programs by firms or government agencies to incentivize their employees (Shearer, 2004; Lazear, 2000; Heckman et al., 1997; Lazear, 1986). While P4P programs are, for the most part, well researched, many factors could alter their effectiveness.

Monetary incentives can be effective, but their effectiveness can vary among more pro-social or civically minded professions (Ashraf et al., 2014a; Ashraf et al., 2014b). More pro-social professions may be more receptive to non-pecuniary benefits such as social rewards or certificates of achievement (Ashraf et al., 2014b). Similarly, people living in less individualistic countries (Hofstede and Minkov, 2013; Hofstede, 2006) may be less responsive to monetary incentives (Bandiera and Fisher, 2013). Individual monetary incentives versus piece-rate wages as a mechanism for increased performance can also result in varied effectiveness (Bandiera et al., 2005; Lazear, 1986). For example, individual monetary incentives can be less effective than piece rates if workers feel they impose a negative externality on their peers by obtaining the reward (Bandiera et al., 2005).

The size of the incentive is also essential in determining the effectiveness of the P4P program. Monetary incentives must be large enough to elicit the desired response (Luo et al., 2020; Miller and Babiarz, 2013; Gneezy and Rustichini, 2000). There is also evidence that a more visible non-pecuniary rewards could be a stronger than a monetary incentive not recognized by peers (Khan et al., 2019; Ashraf et al., 2014a; Ashraf et al., 2014b). The perception of the size of the incentive or their expectationof receiving the incentive could influence performance (Mbiti et al., 2019; Mas., 2006; Gneezy and Rustichini, 2000). Additionally, program structure could have an impact on outcomes. The simplicity or ease with which program participants understand the incentive structure could influence the program's effectiveness (Mbiti et al., 2019). Employment programs that require an opt-in can also suffer from self-selection, with individuals who are already more competitive in the labor force self-selecting into a training or job-placement program (Chen, 2009).

There is a substantive literature on program structure and design explicitly on incentive programs, contests, and award allocation (Benkert and Letina, 2020; Kireyev, 2020; Olszewski and Siegel, 2020; Bimpikis et al., 2019; Halac et al., 2017; Balafoutas et al., 2016; Strack, 2016; Xiao, 2016; Goltsman and Mukherjee, 2011; Sisak, 2009; Cohen et al., 2008; Che and Gale, 2003; Moldovanu and Sela, 2001; Fullerton and McAfee, 1999; Taylor, 1995; Lazear and Rosen, 1981). For example, Lemus and Marshall (2021) find that real-time public leaderboards, on average, increase outcomes in dynamic competitions. In the form of submissions, efforts increased by about 21%, and on average, the maximum score increased by 1.7%.

2.2. Literature about P4P in Contracted- Out Employment Programs

Another, smaller, branch of literature analyzes employment programs in which the government offers employment services such as job brokerage, counseling, training, and case-management through contracted service providers and gives financial incentives to these providers

for achieving specific employment outcomes. This type of contracting has been implemented in the US and the UK since the late 1980's. Since then, some OECD countries like Australia, the UK, Netherlands, Sweden and Italy have adopted models of large-scale outcomes-based contracting (OECD,2022).

There is some evidence of the effectiveness of this type of interventions and possible unintended consequences, 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) 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.

Stapleton et al. (2008) analyzed the Ticket to Work and Self-Sufficiency (TTWSS) program. The

TTWSS was enacted in 1999 and provided a ticket (voucher) to eligible Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) beneficiaries to obtain employment and other employment support services (Morton, 2013). Overall, the TTWSS program saw an increase in the beneficiary enrollment rate of about 0.7 percentage points in its second year and no evidence of increased earnings for participants. Lu (2015) looks at the effect of Indiana's PBCs granted to non-profits for the job placement of their vocational rehabilitation (VR) program. Overall, Indiana PBC participants had an increase in the log-odds of employment by 0.76. After introducing the PBCs, individual employees in Indiana achieved employment outcomes in about 72 fewer days. Heinrich and Choi (2007) look at the change in contracting by the Wisconsin Works (W-2) program with public and private organizations agreeing to offer Jobs Opportunities and Basic Skills (JOBS) services to help welfare recipients gain employment. Participating organizations focused their efforts on the performance measures allocated the most weight in their contracts while their performance in measures indicated as optional was "comparatively poor." Koning and Heinrich (2013) evaluate a similar program in the Netherlandswhere employment incentives were offered via PBCs to private social welfare providers to help the unemployed, and disabled workers gain employment. While there was evidence of creaming and other forms of gaming, there was a roughly a three percentage point increase in job placement among the unemployed group, but not among the disabled worker group.

Colombia has also implemented this type of P4P contracted-out employment programs previously through the *Jóvenes en Acción* (JeA) program starting in 2002. This program had a P4P component that paid part of the cost of offering training to the providers up-front but paid the rest once the trainees were placed in an internship. Another difference between the JeA program and the *Empléate* program is that the former one offered 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.

Like the *Empléate* program, though, the JeA program also offered a monetary stipend to cover for transportation and childcare costs. Attanasio, Kugler and Meghir (2011) and Kugler et al. (2022) conduct a random 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 intoagreements 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 verifiedby 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. One year and three months after, this effect was still present, albeit smaller; of 8 percentage points (Chaparro et al, 2020).

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

As discussed in the previous section, Colombia 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.

Empléate, Prosperidad Social established a financing arrangement in which twelve service providers received payments contingent on the verified achievement of results. 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 $149 and USD $68, respectively. There were extra payments for employment results among beneficiaries with disabilities, women aged 40 or more, and graduates from *Jovenes 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)	
	Thousand COP	USD
Beneficiary gets a formal job at the end ofthe intervention	1,500	$405
Beneficiary remains in a formal job 3months after the intervention	550	$149
Beneficiary remains in a formal job 5months after the intervention	250	$68
Premium if beneficiary has a disability	250	$68
Premium if beneficiary is a women aged 40 or more	250	$68
Premium if beneficiary participated in the *Jóvenes en Acción* program	250	$68

Notes: Payments in dollars are estimated using the official exchange rate in September 1st,2020 (i.e., 3745 COP 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.

Our survey data, which we collected based on a random sample of people assigned to the treatment and control groups, shows that 99% of beneficiaries received basic intermediation services and 14.8% received an additional service, defined as "specialized". One beneficiary could receive more than one basic or specialized service. The most provided basic services were registry in the Public Employment Service Information System (78.9%), psycho-technical testing (33.6%), and job-search tools workshops (32.1%). In terms of specialized services, 13.3% received labor profiling services (e.g.,reconstruction of CVs, document management support), 8.4% received training (e.g., tailored training, technical certifications), 5.7% received selection process support (e.g., home visits, specific labor skills tests), and 5.6% received complementary support (e.g., dressing,hairdressing, transportation, and childcare support) (see Appendix 1 for more details).

Empléate operated nationwide[2] from September 2020 to November 2021. Our survey data shows that more than 70% were in the departments of Atlántico, Córdoba and Bogotá, their

[2] Service providers were based in he 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.

average age was 30 years old, and 77% were women. Moreover, 91.4% belong to two lowest socioeconomic strata, 19.7% had higher or postgraduate 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 aspired to have a salary equivalent to one minimum wage and 30.4% a salary of between one and two minimum wages. The majority 58.2% aspired to have a stable work in a company.

3. Methodology

3.1. 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 social or employment programs) 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, as it was mentioned before, 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 private and non-profit organizations as well as family compensation funds located in different parts of the country.[3] The randomization was done using a random number generator in Stata, which ranked individuals within each service provider. Then, the top 85th percentile of applicants according to ranking were chosen to be in the treatment group and the lowest 15th percentile were left in the control group.[4] For each randomization, we verified that the baseline characteristics, including gender, age, level of education, and housing conditions were not significantly different from each other in the treatment and control groups. The group that received the treatment was a non-randomly selected subset of those assigned to the treatment group. 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, eligible participants assigned to the treatment were recontacted too late, when they had already engaged in different activities. 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 and that administrative difficulties caused important delays, a vast number of those assigned to the treatment, did not participate in the program. Table 2 below shows how the number of people who were randomly assignment (Zi) and who participated in the program (Di). Between September 2020 and April 2021, 21.928

[3] 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.

[4] A majority was assigned to the treatment group as requested by the program administrators. Given thatthe 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 and contamination.

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.

3.2. Identification Strategy

Given random assignment, we estimate the causal effects of the *Empléate* program using the

following regression model:

$$Y_i = \alpha + \rho Assigned_i + \theta_i + \eta_i \tag{1}$$

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. In addition, we estimate the following specification in which we also include characteristics of the individuals to control for any remaining differences between the treatment and control group:

$$Y_i = \alpha + \rho Assigned_i + \theta_i + \beta X_i + \eta_i \tag{2}$$

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. We also include the baseline score in Colombia´s official proxy means test to target social spending (SISBEN), which is an indicator of poverty and socioeconomic status. We also include labor history indicators 3 years before program participation: number of months with formal employment and employment in a large firm. And we include an indicator that shows if the person had a formal job 6 months before the program. We also control for regional fixed effects.

We estimate fully saturated models as well for different groups of workers to examine potential heterogeneous effects. We estimate separate models for men and women, for those older

and younger than 30 years of age, for those with and without high school degree, and for those with and without college. We also derive separate models for individuals receiving services from two different types of service providers: private firms, and employment and training agencies.

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 eliminate potential bias introduced by creaming if one was to instead to compare participants and non-participants and estimate Average Treatment Effects (ATE). However, the ITT may be downwardly-biased as many of those assigned to the treatment group did not actually receive treatment.

4. Data Description

4.1. 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, and salary. 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 in 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 characteristics of and services provided by

the *Empléate* program. The survey includes questions about how the participants learned about the program; when and how they applied; what their expectations were; what services did they receive, 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.

4.2. 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: 85% 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 generally balanced, indicating that the randomization worked well. The only exceptions are the higher education dummies, with the control group having a slightly higher probability (0.017 percentage points) of having a higher education degree that is significant at 5%, and age, with the control group being on average 0.36 years younger than the treatment group, that is significant at 10%.

As shown below, 24% of those in the treatment and control groups are male. The average age is 36 years old and 29% are married. Among those in the control group, 45.7% have a high

school degree, while 44% have a high school degree among those in the treatment group. Overall, 37% have a maximum 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[5]. A small share of approximately 6% participated in the *Jóvenes en Acción* program. Also, around 28% of participants had a formal job six months before program participation; 86% worked with a big firm in the three years before the program and the number of registers in PILA in the three years before the start of program are around 7. There are slight differences in these job history variables but none of them are statistically significant.

We also use administrative data from social security records in PILA to build variables that describe formal labor history. As seen in Table 3, 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]		

[5] 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 the rural areas: 0 to 36.83.

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 (>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		

5. Impact of *Empléate* Program

Table 5 shows ITT estimates of the effects of the program on employment outcomes. 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, but five to eight months after the treatment. Column (3) show the impact on monthly average wage two to five months after program application. Column (4) shows the impact on monthly average wage five to 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 next panels show results 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 and beneficiaries that received services from private firms with training programs.

There are no short-run effects two to five months after the treatment, which could be explained by locking-in effects or delays in social security registration. The results are only observed on the probability of being employed five to eight months after the treatment. This suggests 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 men (17%), people with high school and no college (10%), and those receiving services from private firms (12%).

Also, the program shows no short-run nor long-run effects on monthly average wage. Only people that 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)[6] or 14% compared to those randomly assigned to the control group. This effect is 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 without college (4%) and people

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

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.

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					
Eligible	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					
Eligible	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					
Eligible	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					
Eligible	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					
Eligible	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					
Eligible	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					
Eligible	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					
Eligible	-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					
Eligible	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					
Eligible	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					
Eligible	-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					
Eligible	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					
Eligible	-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					
Eligible	-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

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.

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, this 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	1,824	1,924	2,142	2,333	2,271	4,562	9,698**	11,534	11,676	11,575	11,784
	(3,627)	(3,295)	(5,695)	(5,694)	(5,684)	(5,684)	(4,999)	(4,614)	(7,918)	(7,918)	(7,906)	(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* (0.014)	0.025* (0.014)	0.027* (0.015)	0.026* (0.015)	0.026* (0.015)	0.026* (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.

6. 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 these greater effects could, to some extent, 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[7]. (Appendix 3).

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

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

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.

7. 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 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, 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. There were also positive impacts on the probability of working on a large formal sector firm for men (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

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., 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.

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.

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)*.

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

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.

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.

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.

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.

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.

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 [0.003]	0.329 [0.012]	-0.090***	0.000***
Age	36.084 [0.082]	39.061 [0.323]	-2.977***	0.000***
Married	0.305 [0.003]	0.160 [0.010]	0.145***	0.000***
Primary education or less	0.043 [0.001]	0.045 [0.006]	-0.002	0.703
High school education	0.449 [0.003]	0.306 [0.012]	0.143***	0.000***
Technical Education	0.374 [0.003]	0.457 [0.013]	-0.083***	0.000***
College education or more	0.087 [0.002]	0.134 [0.009]	-0.048***	0.000***
SISBEN score	21.612 [0.088]	21.380 [0.363]	0.232	0.507
Former participant of Jóvenes enAcción program	0.056 [0.002]	0.083 [0.007]	-0.027***	0.000***
Employed 6 months before treatment	0.278 [0.004]	0.305 [0.015]	-0.028*	0.070*
Big formal employer (>20 employees) (3 years before treatment	0.879 [0.003]	0.899 [0.010]	-0.019*	0.077*
# of months contributing to social security before treatment (3 yearsbefore treatment)	6.511 [0.064]	9.143 [0.264]	-2.632***	0.000***
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 [0.004]	0.261 [0.004]	-0.036***	0.000***
Age	38.370 [0.109]	34.237 [0.094]	4.134***	0.000***
Married	0.292 [0.004]	0.298 [0.004]	-0.006	0.291
Primary education or less	0.047 [0.002]	0.039 [0.002]	0.008***	0.001***
High school education	0.496 [0.004]	0.397 [0.004]	0.099***	0.000***
Technical Education	0.326 [0.004]	0.423 [0.004]	-0.096***	0.000***
College education or more	0.099 [0.003]	0.081 [0.002]	0.018***	0.000***
SISBEN score	25.236 [0.106]	18.491 [0.107]	6.745***	0.000***
Former participant of Jóvenes enAcción program	0.064 [0.002]	0.053 [0.002]	0.011***	0.000***
Employed 6 months before treatment	0.290 [0.005]	0.271 [0.006]	0.018**	0.018**
Big formal employer (>20 employees) (3 years before treatment	0.883 [0.004]	0.877 [0.004]	0.006	0.298
# of months contributing to social security before treatment (3 yearsbefore treatment)	7.969 [0.089]	5.596 [0.073]	2.372***	0.000***
N	12,380	13,737		