Who benefits from job training programs? Evidence from a high-dosage program in Brazil*

Daniel Da Mata † Rodrigo Oliveira ‡ Diana Silva §

July 15, 2021

Preliminary version—Latest version available here.

Abstract

Using admission lotteries and registry data linking labor market outcomes, we study the effect of a vocational training program focused on disadvantaged individuals in Brazil. The intensive program is an 18-month classroom training coupled with a 6-month on-the-job training provided by government-sponsored training centers. When assessing the impacts on 15,000 winners and 200,000 nonwinners who graduated in different business cycle moments, we show that female students fare better than their male counterparts. Results are driven by courses in services and those located in faster-growing local labor markets. Investigating outcomes beyond employment and earnings, we do not find an impact on entrepreneurship or university admission.

Keywords: Job Training Programs, Employment, Entrepreneurship.

JEL Codes: I26, I28, J24, L26

*We thank Orazio Attanasio, Nathan Canen, Tiago Cavalcanti, Tulio Cravo, Bruno Ferman, Luiz Fontes, Daniel Haanwinckel, Bruno Ottoni, Vladimir Ponczek, Leonardo Rosa, Cezar Santos, Edson Severnini and Michèle Tertilt their valuable comments. Bruna Mirelle and Lucas Tumkus provided excellent research assistance. We are grateful to the Educational Department of Bahia State for making available the data for this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This study was approved by FGV's Research Ethics Board (IRB Approval Number 11/2020). Any findings or conclusions expressed are those of the authors.

†Sao Paulo School of Economics - FGV. Email: daniel.damata@fgv.br

‡Federal University of Bahia (UFBA) - Department of Economics. Email: rodrigo.coliveira13@gmail.com

§Federal University of Bahia (UFBA) - Department of Economics. Email: dlgsilva@ufba.br

1 Introduction

Governments worldwide have been devising policies to tackle persistent unemployment and improve the skills of more vulnerable workers. One key initiative is promoting vocational training programs, which aim to enhance worker-firm matching and ease the transition from education to work. Since public funds present high opportunity costs and returns to policies vary considerably (Hendren and Sprung-Keyser, 2020), a growing literature has been assessing the effects of government-sponsored active labor market policies (Card, Kluve, and Weber, 2010, 2018). An important aspect is that the returns to a specific policy are subject to aggregate shocks and may depend on the economic cycle (Rosenzweig and Udry, 2020). Lower-skilled and more vulnerable workers—the target population of most labor market policies—are usually hit hardest during recessions. Evaluating the returns to job training programs during economic downturns is then crucial to those workers.

In this paper, we study the effects of a "high-dosage" vocational training program in one of Brazil's poorest states, where the labor market is characterized by high informality, low educational attainment, and high unemployment. The program, called PROSUB ("Programa Subsequente" in Portuguese), offers training for disadvantaged individuals with a secondary degree from state-sponsored schools. The main goal is to help students overcome barriers to enter the formal labor market.

Four features of the program stand out: (i) target population, (ii) length and content of the program, (iii) government participation, and (iv) size of the program. Unlike many vocational training programs worldwide, all high-school graduates from public schools can apply for PROSUB, regardless of age or job market experience. Importantly, the program offers a mix of classroom training and internship for 24 months. More specifically, the intensive program offers an 18-month classroom training (6 months for general subjects such as math and Portuguese plus a 12-month period for occupation-specific subjects) coupled with a 6-month on-the-job training.[1] Additionally, while government-run training centers are integral to PROSUB, many vocational programs in developing countries evaluated by the literature were provided (entirely or partly) by private institutions.[2] Lastly, the program offers approximately 8,000 slots annually in a large variety of courses at dozens of government-run training institutes spatially distributed throughout the state territory.

The program is oversubscribed and selects students via admission lotteries. Because of the program's large size, our sample of lottery winners and nonwinners is larger than previously found in the literature: approximately 200,000 applicants and more than 15,000 lottery winners. We leverage randomized admission lotteries to assess the effects of the

[1] A high-dosage course comes at a cost (e.g., lower compliance with lottery results and higher dropout rates), as the opportunity cost for completing the degree is higher than that of short-duration vocational training programs typically studied in the literature.

[2] PROSUB is tuition free, but there is no financial assistance to cover living costs.

vocational training program. We start by estimating the impact of the program on its primary goal: formal employment. We also assess the effects of the program beyond labor market returns by checking whether the post-secondary training program is used as a stepping-stone for other objectives. As explained in detail in the institutional background (see Section 2), PROSUB students faced a severe recession at different moments after their graduation. We analyze a cohort that has graduated one year before the start of the recession, and another cohort whose students graduated at a recession.[3] As a result, our cohorts of recently trained workers carried out job search efforts during different points of the business cycle.

Our analysis innovates in at least three aspects by showing when, for whom, and how the large program works. First, we assess short- and medium-run effects of different PROSUB cohorts using nationwide (matched employee-employer) administrative data with pre- and post-program labor market outcomes. We estimate a Local Average Treatment Effect (LATE) using the lottery as an instrumental variable for program enrollment.[4] We show that the high-dosage training program generates a positive effect on labor market participation for female students who graduated before the recession. For those workers, we document a persistent and positive effect of 4–5 percentage points (on a baseline employment of 31 percent) during four to six years after the lottery. We also find an increase of 13% in the intensive margin as measured by the number of months working in formal employment. The effects concentrate on employment outcomes, with no distinguishable impact on earnings. When assessing the impacts on cohorts that have graduated during different points of the business cycle, we show that female students fare better than their male counterparts—regardless of when they have graduated. In particular, male workers are especially negatively affected when graduating at a recession.

Second, the program's characteristics and the large sample enable us to perform additional heterogeneity analyses to understand why female students perform better. We check how returns vary by students' experience, type of course, and characteristics of the local labor markets. We compare students with and without previous labor market experience within the same program—most of the literature we relate can only perform such comparison between policies. Previous labor market experience does not seem to interfere with female students, but experience matters to explain the results of male students (as less experience male students fare worse). Moreover, we verify if returns to different fields vary by separating the courses into two areas: (i) agriculture and manufacturing and (ii) services. Consistently with the role of services influencing female labor force participation (Ngai and Petrongolo, 2017; Buera, Kaboski, and Zhao, 2019), courses in services drive

[3] Brazil experienced a large recession between April 2014 and December 2016 when *per capita* Gross Domestic Product fell by 9 percent. Bahia state experienced also a similar severe recession.

[4] Unlike other settings studied in the literature, compliance with PROSUB lotteries is low: approximately 40% of lottery winners enroll in a course. See Section 2.

the positive results of female students who graduated before the recession. Female students who graduated at the recession in service-related courses presented a negative effect, but were relatively less affected than their male counterparts.[5] To shed light on the role of local economic growth, we compare the impacts of courses located in municipalities with above- and below-median GDP growth. With this exercise, we analyze the role of economic growth and depressed local labor markets within jobseekers of the same cohort. Our results suggest that female workers enrolled in courses in faster-growing local labor markets (municipalities with above-median GDP growth) are those who have benefited the most. Male workers in courses located in municipalities with below-median GDP growth are those more negatively affected.

Finally, we measure the program's impacts on entrepreneurship and university admission. This is important because assessing the impacts only on employment may understate training programs' benefits (Kugler, Kugler, Saavedra, and Herrera-Prada, 2020). The effects on entrepreneurship and higher education outcomes are conceptually ambiguous. PROSUB students acquire basic skills (see details in Section 2) that might be useful to run a small business or to enter a university. By contrast, the program's positive impact on student's labor outcomes may refrain them from engaging in entrepreneurship activities or attempts to enter a university. Results do not point out any impact on either entrepreneurship or admission at a selective university. Taken together, the results suggest that the program achieves its goal (for a subgroup of students) by promoting access to the formal labor markets and does not interfere with other outcomes.

This paper relates to the growing number of studies that generate experimental and lottery-based evidence of vocational training programs. Most of the papers find positive returns of vocational training for women in developing countries, such as Attanasio, Kugler, and Meghir (2011) and Attanasio, Guarín, Medina, and Meghir (2017) for Colombia; Acevedo, Cruces, Gertler, and Martinez (2017) for the Dominican Republic; Maitra and Mani (2017) for India; Field, Linden, Malamud, Rubenson, and Wang (2019) for Mongolia; and Camargo, Lima, Riva, and Souza (2020) for another program in Brazil. In a recent study comparing standard training courses and on-the-job training, Alfonsi, Bandiera, Bassi, Burgess, Rasul, Sulaiman, and Vitali (2020) find an improvement in employment rates of youth in Uganda but with different dynamics over time—returns to training courses emerge slowly but were long-lasting. A few papers found no impact on labor market outcomes (e.g., Hirshleifer, McKenzie, Almeida, and Ridao-Cano, 2016 in Turkey).[6]

[5] Descriptive evidence indicates that female students who graduated before the recession sorted into service-related courses and courses associated with a female gender bias in labor demand.

[6] Card, Ibarrarán, Regalia, Rosas-Shady, and Soares (2011) find no impact in the Dominican Republic, but Acevedo et al. (2017) find an effect studying the same program in different years. Attanasio et al. (2017) find that long-term returns of the program in Colombia are similar for males and females. Barrera-Osorio, Kugler, and Silliman (2020) find positive impacts for both men and women also in Colombia. In Argentina, Alzúa, Cruces, and Lopez (2016) find stronger results for men, but only in the short run.

Our contribution to this literature is threefold. First, we study long-term effects of an intensive program. The other papers that study long-term impacts focused on relatively shorter-duration programs (e.g., Attanasio et al., 2017 and Barrera-Osorio et al., 2020), while papers studying longer-duration programs (e.g., Field et al., 2019) do not study long-term effects. Second, this branch has largely overlooked whether and to what extent training is used as a stepping-stone for entrepreneurship or university admission.[7] We also add by exploiting our larger sample size to perform heterogeneity analyses (not yet done) to understand why female and male students perform differently.[8]

We also connect to the literature on the returns of education and training during different points in the business cycle (Oreopoulos, Von Wachter, and Heisz, 2012; Card et al., 2018; von Wachter, 2020). We contribute by examining the relative efficacy of vocational training for cohorts facing distinct initial labor market conditions, so our work is closely related to Field et al. (2019). Besides, we add by showing that students who present a greater "lock-in" effect—i.e., a reduction in employment during the courses—are those who also show persistent adverse effects on employment when graduating at times of lower economic growth. More specifically, in our setting, male students present strong lock-in effects, which translates into a persistent negative employment effect for those male students who graduate at a recession (*between* cohort analysis) or below-median growth localities (*within* cohort analysis).

Our work also relates to the strand studying the role of for-profit and nonprofit post-secondary educational institutions. Cellini and Turner (2019) point out that students of for-profit colleges do worse in the labor market. In the experimental and lottery-based literature we have reviewed, non-governmental institutions (either private or nonprofit entities) were training providers, and in some cases (e.g., India), only private companies participate. Therefore, we contribute by studying a setting with only government-sponsored educational institutions. One concern would be that the public sector could be "gaming" by hiring recent graduates from the program but we show that hiring comes from private sector firms.

This article proceeds as follows. Section 2 provides background on the vocational training program. We describe the data in Section 3. Section 4 details the empirical strategy. Section 5 presents the results. Section 6 concludes.

[7] A notable exception is Kugler et al. (2020), which study the impact on higher education enrollments.

[8] We also connect to the vast non-experimental literature. Chakravarty, Lundberg, Nikolov, and Zenker (2019) find positive impacts of vocational training in Nepal on non-farm employment. Studying other programs in Brazil, O'Connell and Mation (2019) find that an employer-informed vocational training program has a larger short-term effect on employment than a traditional program.

2 Background

Bahia—a large-sized state with about 15 million inhabitants and whose territory is about the size of France—is one of the poorest states in Brazil.[9] Bahia's labor market has a large share of informal jobs, less-educated workers, and high unemployment rates. According to the 2010 Population Census, informal jobs represented half of the total employment, and half of the workers had at most eight years of educational attainment.[10] In 2019, Bahia had the second-highest unemployment rate in Brazil—17% against the national rate of 11%.

To tackle undesirable labor market outcomes, Bahia's state government has been sponsoring post-secondary training programs for disadvantaged individuals. One chief initiative is PROSUB ("Programa Subsequente" in Portuguese), a vocational training program that public school students are eligible for after finishing the high school degree (i.e., completing 11 years of schooling).[11] Recent high school graduates and older students can apply for PROSUB, regardless of employment history. This is strikingly different from many vocational training programs that usually focus on youth unemployment participants and do not offer slots for displaced adults to find new careers. Vocational training for younger workers is extensively available worldwide—for instance, according to OECD (2020), 22% of 15-19 year-olds are enrolled in vocational education on average across OECD countries.

PROSUB offers tuition-free vocational training to help students from secondary public schools overcome barriers to enter the formal labor market. The program is tuition free, but students do not receive any scholarship to fund living expenses. PROSUB teachers have college degree (97%) and prior experience working outside the program (87%)—see Table A.1 in online Appendix A for more details on the teachers' characteristics. The program relies on a decentralized structure with 81 government-run centers (vocational training institutes) spatially distributed throughout 50 municipalities across the Bahia state territory.[12] Figure 1(a) shows the location of the state of Bahia. Figure 1(b) displays the location of PROSUB's training centers and the size of the informal sector in the municipalities where those centers are located. Informality is high but heterogeneous across Bahia territory, so there are centers in places with higher and relatively lower labor informality.

The program offers approximately 8,000 annual slots in 50 types of post-secondary courses in areas ranging from health (e.g., nursing, nutrition) and non-health services (tourism, accounting, and management) to agriculture and manufacturing. Table A.2 in

[9]Brazil is a three-tiered federation with 26 states, a federal district, and 5,571 municipalities. Bahia state has 417 municipalities.

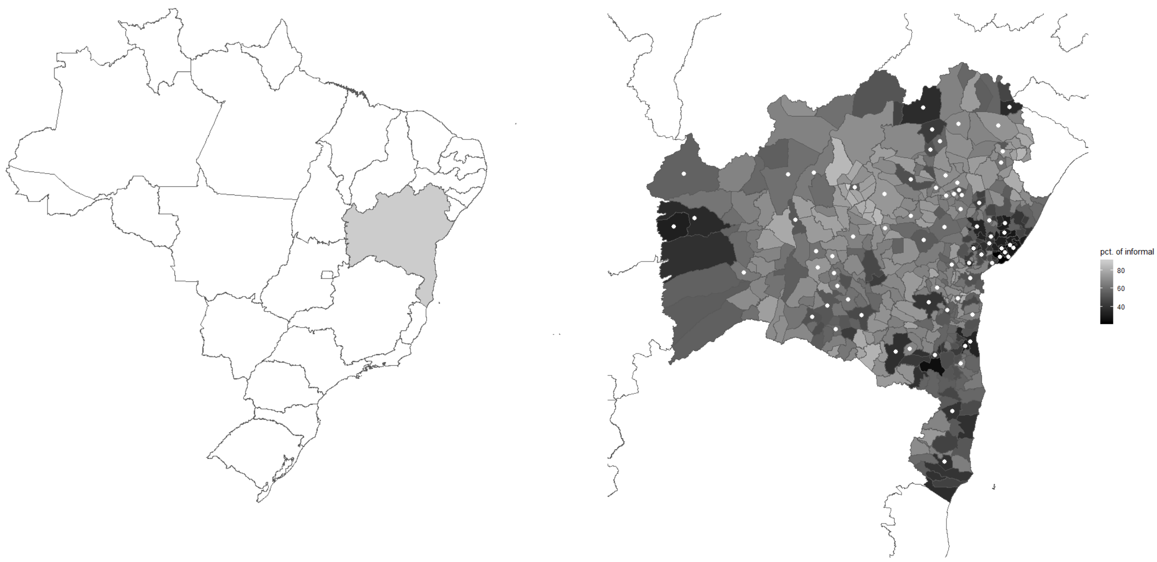
[10]In 2010, the share of informal workers—defined as those not contributing to social security—was 35.4% in Brazil and 49.4% in Bahia.

[11]In Portuguese, PROSUB is also known as "Educação Profissional Subsequente ao Ensino Médio".

[12]Bahia has the second largest number of training centers in Brazil (Bahia (2020)).

online Appendix A shows the list of courses offered. As discussed in more detail shortly, the program is oversubscribed and some courses have very high demand (e.g., more than 3,000 applicants for 15 seats). In 2011, the program was reformulated to be a high-dosage program. Since then, each course offers an intensive program of an 18-month classroom training coupled with a 6-month on-the-job training, totaling 1,600 hours over four semesters. In-class training consists of both (i) 6 months studying general subjects (e.g., math, Portuguese, management, and ethics) and (ii) 12 months dedicated to occupation-specific skills training (subjects focused on the specific chosen field). The on-the-job training component is an internship period at a formal firm located in Bahia. After completing the 24-month training period, students receive a certificate as technicians in the chosen field of study.[13]

Fig. 1: Bahia state and training centers



(a) Bahia state (gray) in Brazil

(b) Distribution of PROSUB centers across Bahia's territory

Notes. Figure 1(a) shows the location of Bahia in Brazil. Figure 1(b) shows the distribution of PROSUB's training centers across Bahia's territory, where each dot represents a training center. In Figure (b), the borders represent the municipalities of Bahia state. Labor informality share—defined as workers not contributing to social security divided by total employment—is in gray scale, according to 2010 Population Census.

The overall goal of PROSUB points us to investigate the effects on labor market outcomes. However, the course provides nonspecific training (one semester of "general" subjects), which may be important for encouraging entrepreneurship and are core subjects in university exam admissions, so we also study the impacts on those additional outcomes.

[13]Notice that PROSUB's course length is closer to a career technical degree in the United States than the typical vocational training program found in the literature.

Admission process. Application is free—likewise the program's tuition—and can be submitted via the education department's website. Each candidate can apply only for one course in one center per calendar year—this is enforced by having applicants declare their unique taxpayer identification number (CPF, "Cadastro de Pessoas Físicas") at the time of application. The program thus excludes the possibility of multiple applications by one person in a given calendar year.

The program selects students via admission lotteries when courses are oversubscribed. Ordinance n. 9,383/2011 established the rules governing the lotteries. During registration, applicants know they are applying for slots to be distributed by lotteries. The lotteries are widely publicized through social media, radio, and television. See Figure A.1 in online Appendix A for examples of marketing of the lotteries. As potential students apply for one course at one center, the education department runs one separate lottery for each course–center. As a result, a set of students applying for course "A" at center "B" would participate in one lottery. The education department's website states when, where, and how the lotteries happen.[14]

The program's selection creates randomized waitlists due to capacity constraints. Applicants are ordered randomly, and slot offers are made following that order until all seats are filled. Conceptually, applicants can be partitioned into three groups: (i) "initial offer" group (those who were ranked up to the number of slots), (ii) "replacement" group (individuals randomly ranked above the number of seats but who eventually get a slot offer), and (iii) "never offer" group (randomly ranked at lower positions and never offered a seat).

Table 1: Sample description: Applicants of the PROSUB lotteries

	Cohort 1: 190 waitlists			Cohort 2: 283 waitlists		
	Total	Male	Female	Total	Male	Female
# of Applicants	106,495	36,489	70,006	97,017	34,975	62,042
% IO (initial-offer) lottery winners	5.98	5.43	6.26	8.79	8.30	9.06
% IO winners who enroll in a course (compliance rate)	42.72	40.65	43.66	44.83	40.37	47.13
% students who finish the course (degree completion)	29.72	24.52	31.81	35.18	27.63	38.92

Notes. This table describes the sample of applicants of PROSUB lotteries. Lotteries of "Cohort 1" took place in Dec-2011 (and courses started Mar-2012) and those of "Cohort 2" took place in Feb-2013 (courses started Apr-2013). Initial-offer (IO) winners are those individuals randomly ranked up to the number of slots. The third row of the table shows the percentage of IO winners who complied with the result of the lottery and enrolled at one PROSUB course. The fourth row shows (conditional on enrollment) the percentage of the PROSUB students who finished the course.

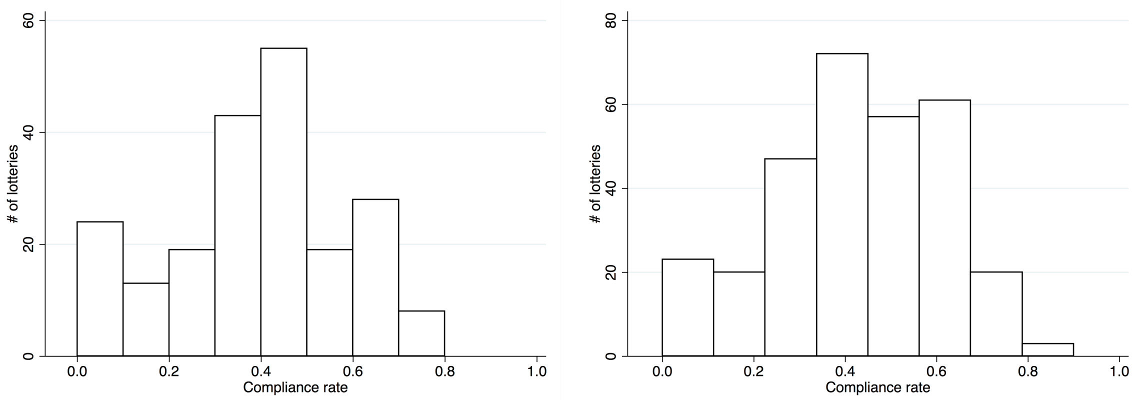
We work with the lotteries that happened on December 13, 2011 ("cohort 1") and on February 05, 2013 ("cohort 2")—the first and second rounds of lotteries after the

[14]The entire process is audited by members of the different supervising institutions and followed by applicants, their family members, teachers, and the local community.

program's reformulation to be considered high dosage.[15] The lotteries of cohort 1 consist of 190 randomized waitlists, which correspond to the number of oversubscribed courses at different centers.[16] These waitlists include 106,495 individuals, 6,367 of whom are in the initial offer group. For the lotteries of cohort 2, there are 283 waitlist, 97,017 applicants and 8,524 initial-offer winners. Table 1 reports that around 65% of applicants are women.

Table 1 also shows that between 43% and 45% of initial-offer winners complied with the result of the lottery by enrolling at a PROSUB course. The compliance with lottery results was higher among female lottery winners. The average compliance rate among initial-offer individuals is relatively low compared to that in other papers focusing on low- and middle-income countries (e.g., Attanasio et al., 2011). Figure 2 points out that the compliance rate varied considerably among the lotteries in our sample. Table 1 shows that degree completion for the students who have enrolled at some course is around 30–35%. Part of the low degree completion rate may be explained by the high opportunity cost associated with a 2-year training course. Degree completion (conditional on enrollment) is also lower than that found in other papers focusing on low- and middle-income countries.

Fig. 2: Compliance rate among lotteries



(a) Cohort 1: 190 waitlists

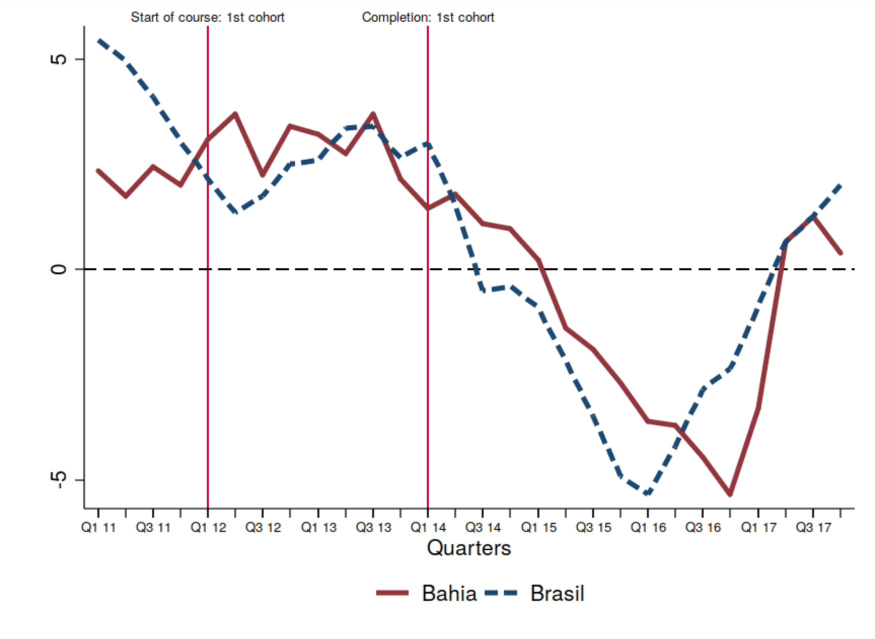
(b) Cohort 2: 283 waitlists

Notes. Figure 2(a) shows the compliance rate among initial-offer (IO) winners in each of the 190 lotteries of cohort 1's sample. Figure 2(b) shows the compliance rate for the cohort 2's 283 lotteries. The compliance rate is defined as the proportion of IO winners who complied with the lottery result by enrolling at a PROSUB course.

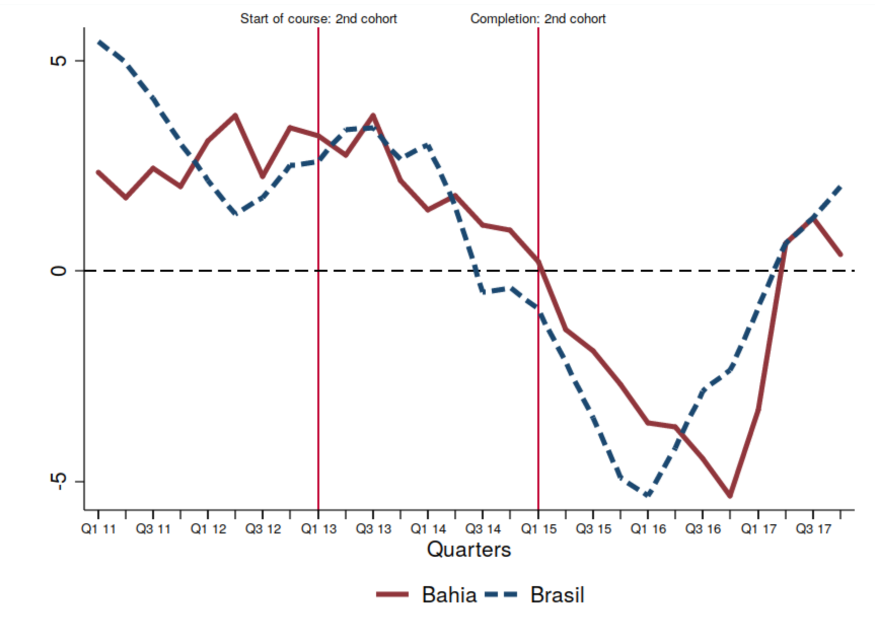
[15]From 2014, PROSUB cohorts were exposed to an additional program: an employment service program. See Da Mata, Oliveira, and Silva (2020) for an assessment of the employment service program.

[16]The waitlists are mutually exclusive since one person cannot perform multiple applications.

Fig. 3: GDP growth and PROSUB cohorts



(a) Cohort 1: Start in Mar-2012 and Completion in Dec-2013



(b) Cohort 2: Start in Mar-2013 and Completion in Dec-2014

Notes. The figures show the evolution of quarterly Gross Domestic Product (GDP) for Brazil (solid line) and for Bahia state (dashed line) between 2011 and 2017. Figure 3(a) shows the window between cohort 1's course start (March 2012) and course completion (December 2013). Figure 3(b) displays cohort's 2 course start (March 2013) and course completion (December 2014).

The cohorts of PROSUB students started their job search efforts amidst different points of the business cycle. Given the 24-month course, cohort 1 completed their degrees in December 2013, while cohort 2 finished in December 2014.[17] Brazil faced a severe recession between 2014 and 2016, when *per capita* Gross Domestic Product fell approximately 9 percent and unemployment increased to 12% (up from 7.2%) according to Brazil's Bureau of Statistics (see CODACE (2020)). Figure 3 shows that Bahia state also experienced a severe recession, but Bahia has reached a trough around three quarter after Brazil's GDP trough. Panel (a) of Figure 3 indicates that cohort 1's graduation took place one year before recession officially started in Bahia. Panel (b) of Figure 3 shows that cohort 2 graduated at the start of the recession period. We use the fact that cohorts graduate before and during a recession when interpreting the impacts of PROSUB.

Table 2 presents selected characteristics of our sample of waitlists. The median number of seats is 30. The median number of applicants per seat is approximately 8 (cohort 1) and 5 (cohort 2). There are some lotteries with very high demand: more than 100 applicants per seat. IO winners' non-compliance rate is approximately 56%. According to the program's implementers, initial-offer results were widely publicized and people knew whether they were selected.

Table 2: Sample description: Waitlist characteristics

	Median	Mean	St. Dev.	Min.	Max.
Panel (a): Cohort 1, 190 waitlists					
# of Applicants	246	577.8	919.4	12	6310
# of Seats	30	34.7	16.9	2	120
Applicants/Seats	8.37	16.2	20.2	1	118.1
% declining offers in initial offers	56.7	57.9	17.0	20	100
Panel (b): Cohort 2, 283 waitlists					
# of Applicants	190	443.9	789.7	9	5759
# of Seats	30	37.8	15.3	9	120
Applicants/Seats	5.37	11.6	19.1	1	155.9
% declining offers in initial offers	55.6	54.6	16.3	10	90

Notes. This table describes the characteristics of PROSUB's waitlists. Panel (a): descriptive statistics for the waitlists of PROSUB's cohort 1. Panel (b): descriptive statistics for the waitlists of PROSUB's cohort 2. The "% declining offers in initial offers" is defined as the proportion of initial-offer winners who declined (did not complied) with the lottery result by not enrolling at a PROSUB course.

[17] The lotteries in December 2011 allocated 95.4% of the students to start their courses early in 2012 (while 4.6% began their studies during the second semester of 2012).

3 Data and Summary Statistics

3.1 Data

In this paper, we use four administrative registries to create the relevant outcome variables. The four registries were linked by using a unique identifier.[18] The registries allow us to follow the individuals of our large sample over time, thus minimizing concerns about attrition.

RAIS. The labor market outcomes stem from RAIS ("Relação Anual de Informações Sociais"), a matched employee-employer dataset from Brazil's Ministry of Economy. The RAIS dataset has information on each formal worker at each plant in Brazil, as all establishments in Brazil are legally required to submit information to RAIS. We use yearly information for the period 2008–2018. We construct variables related to the extensive and intensive margins of formal employment. For the extensive margin, we create a employment dummy which equals one if the individual is formally employed at at least one month each year and zero otherwise. For the intensive margin, we calculate the number of months that each applicant spent on formal jobs. We also collect information on average earnings for each year.[19] Notice that RAIS has information (including socioeconomic characteristics) only for workers in the formal labor sector.

PROSUB registry. We merge the labor market dataset with PROSUB's registry to built a panel database for all lottery applicants. The longitudinal data track the complete history of lottery applicants in the formal labor market and identify changes in employment status and earnings. The PROSUB dataset allows us to identify individuals who are applicants, lottery winners (initial-offer winners), those who have enrolled in training courses, and those who have completed the entire training course. Even though the registry provides data on enrollment, it does not show if a person in the replacement group has been offered a seat. Many individuals of the replacement group were never contacted because they provided the wrong contact details.[20] Therefore, we work in this paper with initial-offer winners, and do not work with ever-offer winners due to data constraints (de Chaisemartin and Behagel, 2020). The registry also gives us a set of baseline characteristics for each applicant. Most characteristics stem from the (short) application form used by the program

[18]The unique identifier is the CPF ("Cadastro de Pessoas Físicas"), a nine-digit individual taxpayer identification number. In our data, all PROSUB applicants have a valid CPF number. More precisely, three registries were matched using the nine-digit CPF. One registry (CNPJ public dataset) provides only six digits of each CPF, so we proceeded as follows. Approximately 90% of individuals were uniquely matched using the six-digit CPF. For those individuals in which the six digits generated multiple matching, we identify the PROSUB applicant by checking the first name, the last name, and then the full name, accounting for minor misspellings. The final dataset is de-identified.

[19]For few individuals who have two or more jobs, we considered only the job with higher earnings.

[20]Differently from initial-offer results, subsequent offers were not widely publicized. According to the program's implementers, it was logistically more challenging to publicize and contact people in the replacement group (as some people put wrong phone numbers).

to contact trace winners (email and telephone number, for instance). We show later how we use the application form's information for the lotteries' balancing exercise.

CNPJ public dataset. We obtain the entrepreneurship outcome from the publicly available CNPJ (National Register of Legal Entities) registry. The CNPJ dataset—administered by Brazil's Internal Revenue Service—has establishment-level information on the date of opening, sector of economic activity, type of tax regime, list of owners, among other characteristics. We selected all establishments categorized as micro, small, and medium enterprises and determined those owned by the applicants of the PROSUB lottery.[21] We then selected the subgroup of establishments whose date of opening is after the lottery date. We set the entrepreneurship dummy variable to equal one for those applicants who are owners of establishments opening from January 2012 on.

UFBA university admission. We use the Federal University of Bahia (UFBA) dataset to check whether PROSUB applicants were eventually admitted into that university. UFBA is one of Brazil's largest universities and the most highly demanded one in the state of Bahia. Like all public universities in Brazil, UFBA is tuition free. University admission depends on the applicant's performance on an entrance exam.[22] We track each student admitted into UFBA and set a university admission dummy to equal one for those admitted who were applicants of the PROSUB lottery.

3.2 Summary Statistics and Balancing

Our sample is composed of the 106,495 applicants in cohort 1 and 97,017 applicants in cohort 2. Table 3 reports information for PROSUB's cohorts 1 and 2. Column (i) of Table 3 shows the baseline characteristics of the applicants. One year before the lottery, approximately 38% of applicants had a formal job in RAIS.

Balancing. We estimate the following specification:

$$y_i = \alpha + \rho Z_{is} + \gamma_s + \varepsilon_i \quad , \tag{1}$$

where y_i is a baseline characteristic of individual i, and Z_{is} is the initial-offer dummy, which equals one for initial-offer lottery winners of center-course s and zero otherwise (i.e., zero for the replacement group and the never offer group). We include the lottery fixed effects γ_s (center-by-course fixed effects), and ε_i is the idiosyncratic error term.

[21] More precisely, we selected establishments that belong to the SIMPLES ("Simples Nacional") and MEI ("Microempreendedor Individual") special tax regimes, whose goal is to encourage enterprises to operate in the formal economy by reducing formality costs (e.g., redtape and taxes).

[22] The admission exam has questions on math, Portuguese, physics, chemistry, geography, history, among other subjects. In 2016, the exam had roughly 200,000 participants for 8,875 slots.

Table 3: Lottery balancing

	All Applicants	Nonwinners	Difference IO Winners and Nonwinners	p-value
	(I)	(II)	(III)	(IV)
Panel (a): Cohort 1, 190 waitlists				
% Formal Employment in 2010	0.378	0.383	0.007	0.255
	(0.001)	(0.002)	(0.006)	
Earnings in 2010	328.060	333.500	6.099	0.429
	(1.780)	(1.847)	(7.710)	
# months in formal employment in 2010	3.581	3.626	0.096	0.160
	(0.016)	(0.016)	(0.069)	
Female dummy	0.657	0.655	-0.002	0.781
	(0.001)	(0.002)	(0.006)	
Labor experience dummy: > 1 month	0.531	0.537	-0.001	0.921
	(0.002)	(0.002)	(0.006)	
Labor experience dummy: at least 36 months	0.099	0.100	0.001	0.893
	(0.001)	(0.001)	(0.004)	
Address dummy: street	0.751	0.751	0.001	0.896
	(0.001)	(0.001)	(0.006)	
Address dummy: road	0.084	0.083	0.000	0.932
	(0.001)	(0.001)	(0.004)	
Email dummy: hotmail.com	0.729	0.728	0.003	0.624
	(0.001)	(0.001)	(0.006)	
Number of observations	106,494	100,127	106,494	.
Panel (b): Cohort 2, 283 waitlists				
% Formal Employment in 2011	0.384	0.392	0.002	0.772
	(0.002)	(0.002)	(0.006)	
Earnings in 2011	361.171	369.440	8.708	0.315
	(2.168)	(2.285)	(8.661)	
# months in formal employment in 2011	3.617	3.685	0.041	0.531
	(0.016)	(0.017)	(0.066)	
Female dummy	0.639	0.638	-0.004	0.483
	(0.002)	(0.002)	(0.006)	
Labor experience dummy: > 1 month	0.491	0.501	0.001	0.912
	(0.002)	(0.002)	(0.006)	
Labor experience dummy: at least 36 months	0.097	0.098	-0.001	0.722
	(0.001)	(0.001)	(0.004)	
Address dummy: street	0.753	0.752	-0.000	0.962
	(0.001)	(0.001)	(0.006)	
Address dummy: road	0.087	0.087	-0.001	0.849
	(0.001)	(0.001)	(0.004)	
Email dummy: hotmail.com	0.706	0.706	0.005	0.441
	(0.001)	(0.002)	(0.006)	
Number of observations	97,017	88,493	97,017	.

Notes. Panel (a): Balancing tests for 106,495 applicants, 6,367 of whom are in the initial-offer (IO) lottery winner group. Panel (b): Balancing tests for 97,017 applicants, 8,524 of whom are initial-offer (IO) lottery winners. Data are provided at the individual level. Column (I) shows the mean and standard deviation (in parentheses) of baseline characteristics for all the applicants. Column (II) presents the mean and standard deviation (in parentheses) of baseline characteristics of nonwinners applicants (replacement group plus never-offer group). Column (III) reports the coefficients estimated from Equation (1) and the last column the corresponding p-value. *** p<0.01, ** p<0.05, * p<0.1

Table 3 shows the balancing between initial-offer winners and other participants. Data for balancing come from RAIS and PROSUB registries. From RAIS, we use average monthly earnings, the dummy for the extensive margin of formal employment, and the measure of the intensive margin (number of months) in the pre-lottery year. In addition, we create two "previous labor market experience" dummies: the first equals one for those who have worked at least one month before the PROSUB lottery, and the second equals one for those with at least 36 months of formal employment experience. PROSUB's application form provides information on sex (the only socioeconomic variable) and contact information (email and address). The contact information data are used as "placebo" checks as follows. We create an email dummy that equals one if the applicant uses a hotmail account (the most used by applicants) and zero otherwise. We also generate two "address" dummies: the first equals one if the person's address suffix is "street"; the second dummy equals one if the address suffix is "road." Table 3 shows that lotteries created comparable groups. To expand on socioeconomic characteristics, we use RAIS data on age, educational attainment, race, and gender for the subgroup of applicants who had a formal job in the pre-lottery year. Consistent with the randomized design, Table A.3 in online Appendix A displays that the winners and nonwinners of this subgroup are comparable in socioeconomic characteristics.

4 Empirical Strategy

We focus our analysis on investigating the effects of enrolling at PROSUB. Recall that Z_{is} is the initial-offer dummy, which equals one for applicant i getting an initial offer at the time of the lottery for center-course s. Let D_{is} be the enrollment dummy ("immediate enrollment"), which equals one if individual i enrolled at center-course s. We assume the following cross-sectional specification:

$$Y_i = \beta D_{is} + X_i'\psi + \gamma_s + \upsilon_i \quad , \tag{2}$$

where Y_i is the outcome variable for individual i and γ_s represents lottery fixed effect or strata fixed effect, a set of dummy variables indicating our strata s (center-by-course fixed effects). We use γ_s for comparison of winners and nonwinners within the same lottery. X_i is the vector of baseline characteristics (included to improve the precision of the estimates) and υ_i stands for the error term. To obtain year-by-year effects, we estimate Equation (2) separately for each of the following years: the two years corresponding to the course period and the four years after finishing the course. If the program has positive impacts, the post-course coefficients should be positive and statistically significant.

We instrument for immediate enrollment D_{is} by using the random variation results of the lottery. More precisely, the first-stage equation is as follows:

$$D_{is} = \eta_1 Z_{is} + X_i' \eta_2 + \mu_s + e_i \quad , \tag{3}$$

where the instrument is the initial-offer dummy Z_{is}. The balancing analysis provides support for the identifying assumption (exclusion restriction). The parameter η_1 in Equation (3) will show the compliance rate. Noncompliance occurs in the initial-offer and replacement groups (as some individuals do not enroll after being offered a seat), and also in the never-offer group. A subgroup of individuals of the never-offer group reapplied for PROSUB lotteries in later years (lotteries continue to take place from 2012 on), and some were lottery winners. Therefore, we aim to recover a Local Average Treatment Effect (LATE). We perform the analysis of the individuals in cohort 1 separately from those of cohort 2. For the estimation of the cohort 2, we use the first lottery results as the instrumental variable for course enrollment. Table A.4 in online Appendix A presents the first-stage results.

In the robustness exercise, we perform the analysis without the vector of baseline characteristics. In addition, we estimate an ANCOVA version of Equation (2) by adding pre-treatment values of the outcomes variable. We also estimate the model clustering standard errors at the lottery level—as aggregate shocks may affect the outcomes (Deeb and de Chaisemartin, 2020). To perform heterogeneity analysis, we explore the characteristics of the lotteries' stratification (course type and location of each center) as well as individual characteristics (sex and labor market experience).

To further understand the results of the policy, we estimate a variation of Equation (2) substituting the immediate enrollment dummy for a dummy for degree completion D'_{is}, which equals one if the individual finished the course after being selected by the PROSUB lotteries. For this latter exercise, the exclusion restriction weakens (compared to that of the immediate enrollment) as it requires that the lottery results do not affect outcomes directly, but only through degree completion.

Finally, to obtain an average effect of the program, we estimate the following pooled specification:

$$Y_{it} = \tau D_{is} + X_i' \theta + \delta_s + \lambda_t + \epsilon_{it} \quad , \tag{4}$$

where Y_{it} is the outcome variable for individual i, t years after the lottery date. λ_t is the time fixed effects, δ_s represents lottery fixed effect, and X_i is the vector of baseline characteristics. Standard errors are clustered at the individual level. We again instrument for immediate enrollment D_{is} by using the initial-offer dummy Z_{is}. In the pooled specification, we estimate the impact of the program for the four-year period after finishing the course.

5 Results

We divide the results into three parts. Firstly, we present the baseline results on labor market outcomes. Secondly, we perform heterogeneity analyses. Finally, we finish with further analyses, including the results on entrepreneurship and university admission.

5.1 Baseline results

We start by studying the effects of the program on labor market participation. Figure 4 and Table A.5 in online Appendix A present the baseline results. Panel (a) of Figure 4 shows the (yearly) instrumental variable estimates for cohort 1. During the 2-year course, there is no statistically significant difference in labor market participation. By contrast, after the course, we find an increasing effect on employment probability. Six years after the lottery, we find a statistically significant and positive effect of about 4 percentage points. When we pooled all years after the course (years 3 to 6 after the lottery), we find a positive effect of 2 percentage points (though not statistically significant)—see Column (I) of Table A.5.

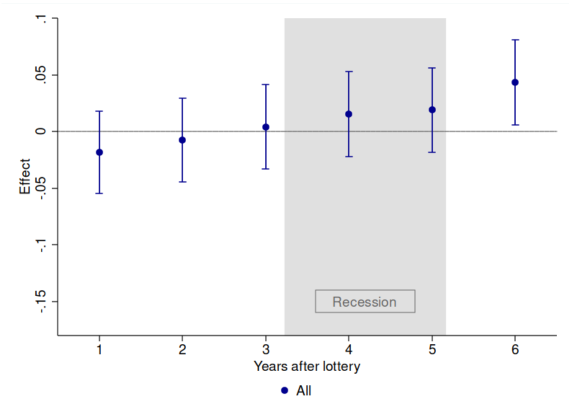
Panel (c) of Figure 4 reports the same set of regressions but splitting cohort 1's sample into female and male participants. For female winners, we find (i) no impact when still under PROSUB training and (ii) an increasing and persistent positive effect of 3–4 percentage points in employment during three to six years after the lottery (i.e., the positive impact starts two years after completing the course). While results for female participants appear immediately after finishing PROSUB, positive effects (though imprecisely estimated) for male participants only appear later on, about four years after finishing the course.

Acevedo et al. (2017) shows that women in vocational training acquire more skills, and men create higher expectations about future earnings. One implication is that men may turn down more easily job offers that they otherwise might have accepted. This is in line with the different results we find for female and male participants. In addition, male winners present a very strong negative effect on employment during the course years—a 6 percentage points decrease in employment probability during the first year after the lottery. This is consistent with a lock-in effect as students are likely to suspend job search efforts when still in the training program (Card et al. (2018)).

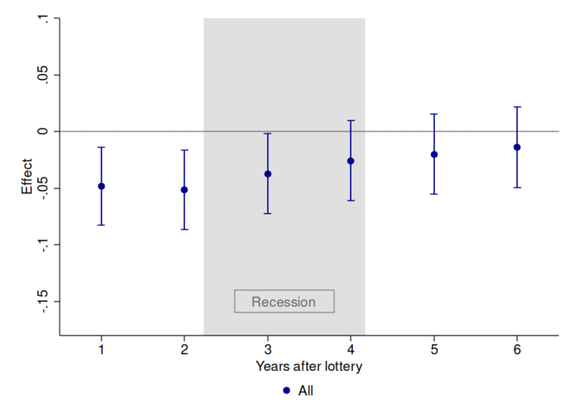
Results point out thus positive effects for female participants. Approximately 31% of the women in the control group were employed in 2010 (pre-lottery values). Our LATE estimates show an increase in female employment of about 4 percentage points, representing an impact of approximately 13% on employment likelihood.

The results are consistent with the literature providing experimental and lottery-based evidence on the effects of vocational training. Similarly to this literature branch, Table A.5 suggests that medium- and longer-term effects for male participants are zero. In addition, our results for women are comparable or stronger than the strand of the literature providing experimental and lottery-based evidence on the impacts of vocational training (e.g., Attanasio et al. (2011), Chakravarty et al. (2019), Alzúa et al. (2016), Field et al. (2019), Acevedo et al. (2017)).

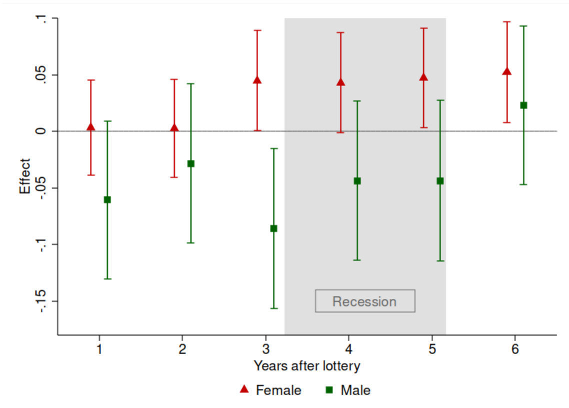
Fig. 4: Baseline estimates: Effects on Labor Market Participation



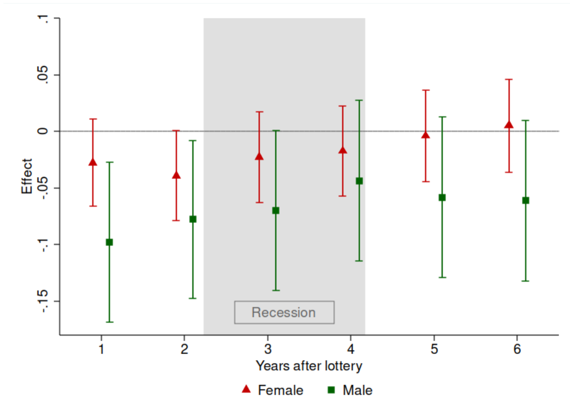
(a) Cohort 1: All applicants



(b) Cohort 2: All applicants



(c) Cohort 1: Female vs Male Applicants



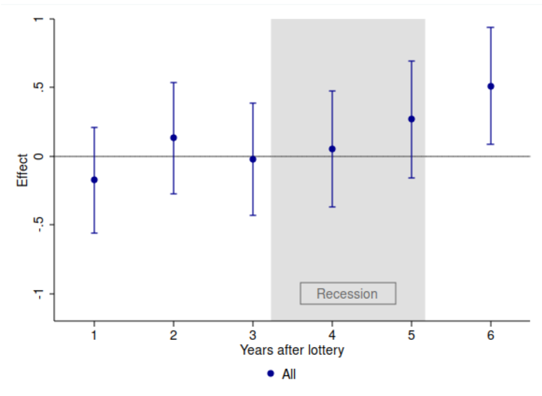
(d) Cohort 2: Female vs Male Applicants

Notes. The figure displays the baseline results for up to six years after the PROSUB lotteries. The horizontal axis shows the number of years after the PROSUB lotteries. PROSUB courses last two years. The dependent variable is an employment dummy. The gray area displays the recession period (see Section 2 for detail). Each estimate corresponds to a LATE effect—to the estimated coefficient of Equation (2)—where the enrollment dummy is instrumented by using the random variation results of the lottery. Panels (a) and (c) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Panels (b) and (d) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018.

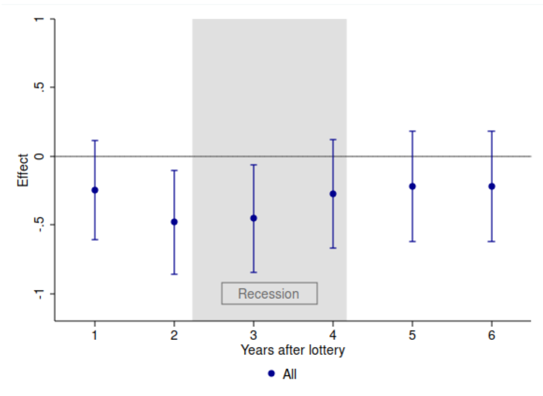
The results for cohort 2—reported in Panel (b) of Figure 4 and Column (V) of Table A.5—have a different pattern. Cohort 2 presents a stronger negative lock-in effect when it comes to employment during the course years. Besides, the negative effect persists until six years after the lottery. Panel (d) of Figure 4 points out that these negative effects are mostly driven by male winners. The negative effect on employment for female winners is transitory (reaching zero effect in the medium-run), while the negative effects for male winners are persistent (of about 6 percentage points on a baseline formal employment of 51 percent).

We show in Table A.6 in the online Appendix A that the results are highly robust when we re-run the regressions with an ANCOVA specification. Results are also highly stable to excluding controls (see Appendix Table A.7) and to clustering the standard errors at the lottery (strata) level (see Appendix Table A.8).

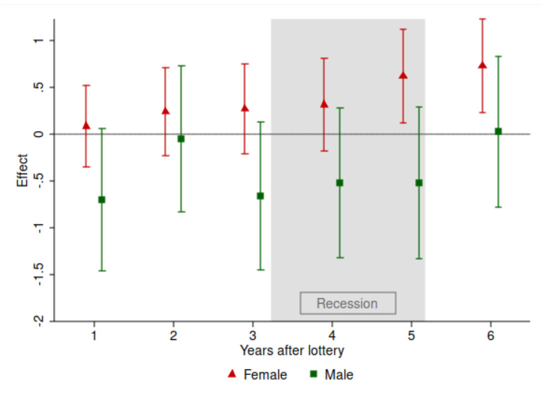
Fig. 5: Estimates: Intensive Margin of Formal Employment



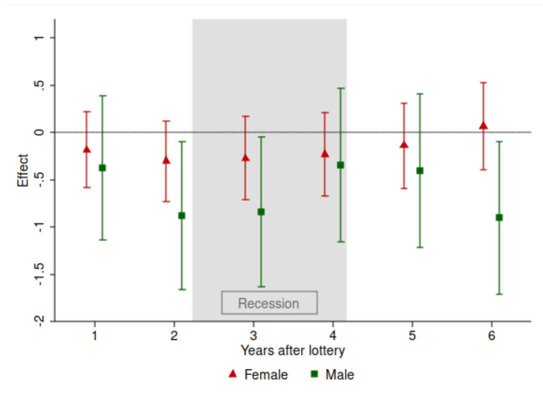
(a) Cohort 1: All applicants



(b) Cohort 2: All applicants



(c) Cohort 1: Female vs Male Applicants



(d) Cohort 2: Female vs Male Applicants

Notes. The figure displays the results of the employment's intensive margin for up to six years after the PROSUB lotteries. The horizontal axis shows the number of years after the PROSUB lotteries. PROSUB courses last two years. The dependent variable is the number of months in formal jobs. The gray area displays the recession period (see Section 2 for detail). Each estimate corresponds to a LATE effect—to the estimated coefficient of Equation (2)—where the enrollment dummy is instrumented by using the random variation results of the lottery. Panels (a) and (c) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Panels (b) and (d) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018.

To complement the analysis of labor market participation, we evaluate the effects on the intensive margin as measured by the number of months working in formal jobs. Figure 5 and Appendix Table A.9 display the results. Consistently, there is a positive effect for women in the first cohort of about 13%, and a negative effect on formal employment of about the same magnitude for males in the second cohort. We also observe lock-in effects for male participants.

In addition to looking directly at employment outcomes, we examine the impacts on earnings. Table A.10 shows that the program has no effects on earnings. Since the program does not seem to impact total earnings, we focus the remaining of the analysis on understanding the potential drivers of the results of labor market participation.

Taken together, three patterns emerge from the baseline results. First, female winners fare better than their male counterparts after the course (regardless of the cohort we analyze). Second, male winners present a strong lock-in effect. Third, the results suggest persistent benefits for women in the first cohort, and persistent losses for men from the second cohort. The differential impact when one compares distinct PROSUB cohorts suggests a role for economic cycle—recall that only the second cohort graduated amidst a recession (see Section 2). To further understand the role of economic growth in driving the results, we perform below a within-cohort analysis by verifying the impacts of differential local labor market growth.

5.2 Heterogeneous effects

This subsection presents further heterogeneous analysis for labor market participation to understand what characteristics are associated with the positive effect on cohort 1's female workers and the negative effects on cohort 2's male workers. In the interest of space, we only report tables of the exercises for female workers and present the tables for male workers in the online Appendix A.

Experience. We initially ask whether results differ by workers' labor market experience. Since experience may be an important driver of labor market outcomes, we explore the fact that PROSUB offers seats for individuals regardless of age or experience. We compare then students who differ in their previous labor market experience within the same program (differently from other studies that perform such comparison between policies). We consider students with previous experience those who have worked at least 36 months before the PROSUB lottery.[23]

Columns (I) and (II) of Table 4 presents the results for cohorts 1's women (with and without previous experience, respectively). Female workers who have previous experience (i) start with a greater lock-in effect and (ii) are able to access the formal labor market more quickly after finishing the course, but the positive effects wane over time (are similar

[23]The same dummy variable was used in the balancing analysis of Section 3.

five years after the lottery and smaller in the sixth year). As a consequence, the findings suggest that displaced older female adults do not seem to fare worse.

Table 4: Effects on Employment: Experience (Women)

	Cohort 1		Cohort 2	
	(I) Experience	(II) No experience	(III) Experience	(IV) No experience
During course				
Year 1	-0.0532	-0.00220	-0.0690**	-0.0170
	(0.0362)	(0.0205)	(0.0343)	(0.0205)
Year 2	0.0167	-0.0463**	-0.0828**	-0.0220
	(0.0366)	(0.0234)	(0.0343)	(0.0225)
After course				
Year 3	0.0542	0.00488	-0.0403	-0.0205
	(0.0367)	(0.0254)	(0.0346)	(0.0233)
Year 4	0.0630*	0.000624	-0.0366	-0.0117
	(0.0370)	(0.0257)	(0.0355)	(0.0230)
Year 5	0.0360	0.0300	-0.0190	0.000662
	(0.0377)	(0.0257)	(0.0353)	(0.0240)
Year 6	0.0269	0.0438*	0.0145	-0.00189
	(0.0377)	(0.0265)	(0.0354)	(0.0245)
Pooled years				
Years 3 to 6	0.0450	0.0198	-0.0204	-0.00837
	(0.0297)	(0.0209)	(0.0285)	(0.0192)
Control: gender	No	No	No	No
# Applicants	31,576	38,428	25,446	36,591

Notes. The table presents the results from Year 1 to Year 6 after the PROSUB lotteries. PROSUB courses last two years. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(II) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (III)–(IV) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Table 4 also displays in Columns (III) and (IV) the results for women of cohort 2. Female students with experience also present a greater lock-in effect. Those students present a more negative impact during the first years after the course, but are able to revert the negative impacts by Year 6 after the lottery. By and large, Table 4 indicates that experience does not seem to be a key driver of employment for female workers when we compare different cohorts.

We inspect then if the previous experience is relevant for explaining the impacts for male students. Table A.11 in the online Appendix A documents that the group of less experienced workers of both cohorts has a relatively higher negative impact on employment. To summarize the implications from Tables 4 and A.11, while previous labor market experience does not seem to affect the results of female workers, experience matters to explain the results of male workers (as less experience male workers fare worse).

Table 5: Heterogeneous Effects: Type of Course (Women)

	Cohort 1		Cohort 2	
	(I) Agriculture & Manufactur.	(II) Services	(III) Agriculture & Manufactur.	(IV) Services
During course				
Year 1	-0.0421	0.00784	-0.0420	-0.0160
	(0.0650)	(0.0227)	(0.0287)	(0.0272)
Year 2	-0.0292	0.00612	-0.0545*	-0.0268
	(0.0690)	(0.0233)	(0.0297)	(0.0277)
After course				
Year 3	0.0636	0.0406*	-0.0281	-0.0187
	(0.0720)	(0.0237)	(0.0301)	(0.0280)
Year 4	-0.00445	0.0467**	-0.0193	-0.0159
	(0.0717)	(0.0238)	(0.0300)	(0.0277)
Year 5	0.0185	0.0510**	0.00676	-0.0130
	(0.0709)	(0.0236)	(0.0304)	(0.0282)
Year 6	0.0271	0.0546**	0.0134	-0.00188
	(0.0718)	(0.0239)	(0.0310)	(0.0283)
Pooled years				
Years 3 to 6	0.0262	0.0482**	-0.00681	-0.0124
	(0.0593)	(0.0196)	(0.0252)	(0.0232)
Control: gender	No	No	No	No
# Applicants	8,441	61,525	28,264	33,069

Notes. The table presents the results from Year 1 to Year 6 after the PROSUB lotteries. PROSUB courses last two years. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(II) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (III)–(IV) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Type of course. We now split the lotteries into two areas: (i) agriculture and manufacturing, and (ii) services. The related literature points out vocational training favors women in different countries, but less emphasis is given to the type of courses offered by the distinct programs. An advantage of our setting is the wide range of courses offered by PROSUB. Male and female workers tend to segregate into different industries and occupations, and gender bias in labor demand is one important factor (Olivetti and Petrongolo (2014)). As a result, gender bias in labor demand can disproportionally favor men or women and can play a role in explaining labor market performance.

Table 5 shows what type of courses seems to influence the results on female employment. Columns (I)–(II) present the results for cohort 1. Female students in services' courses have a large positive and persistent impact after the lotteries (see Column (II)). Column (I) of Table 5 shows that students in courses in agriculture and manufacturing have lower short and medium-run impacts in terms of employment (and statistically non-significant). Notice that in Bahia, services correspond to approximately two-thirds of the GDP. The results are consistent with Ngai and Petrongolo (2017), which highlights the role of services in explaining women's labor force participation.

In Columns (III)–(IV), we show the results for cohort 2's female winners. None of the course areas presents a statistically significant effect, consistent with the overall zero impact on employment for cohort 2's female winners. Even though services present a negative effect, it is not precisely estimated.

Table A.12 in the online Appendix A reports the results for male winners. For the cohort 1, Columns (I)-(II) show that no area presents a statistically significant result. The coefficients are somewhat large, but not precisely estimated. Columns (III)-(IV) show the results for cohort 2 and suggested that the negative effects are driven by services. For the cohort 2, the negative impact for male students in services (Column (IV) of Table A.12) is larger than that for female students in service-related courses (Column (IV) of Table 5). Coefficients for agriculture and manufacturing are large but imprecisely estimated.

Collectively, the analysis by course points out that services drive the positive results of female workers (in cohort 1). The adverse effects on male winners of cohort 2 seem to be accompanied by negative effects of all sectors (though only services are statistically significant).

Faster-growing local labor markets. To gain additional insights, we check whether results are guided by local labor markets' economic growth. The idea to explore heterogeneity in local growth is to understand further the role of economic growth (of regional labor markets) in explaining the results. Notice that labor markets may play an important role because the program is spread throughout a large territory and growth is spatially uneven. For each cohort, we separate lotteries in municipalities with above and below median GDP growth between 2012 and 2017. Students in locations with below-median local growth suffer from continuing exposure to adverse labor market conditions during and after the course.

Table 6: Heterogeneous Effects: Local economic growth (Women)

	Cohort 1		Cohort 2	
	(I) Above median: GDP growth	(II) Below median: GDP growth	(III) Above median: GDP growth	(IV) Below median: GDP growth
During course				
Year 1	0.00822	-0.00206	-0.0112	-0.0528
	(0.0277)	(0.0335)	(0.0241)	(0.0337)
Year 2	-0.00644	0.0133	-0.0303	-0.0529
	(0.0284)	(0.0345)	(0.0248)	(0.0347)
After course				
Year 3	0.0375	0.0530	-0.0181	-0.0303
	(0.0290)	(0.0353)	(0.0252)	(0.0347)
Year 4	0.0618**	0.0211	-0.0198	-0.0138
	(0.0292)	(0.0351)	(0.0248)	(0.0347)
Year 5	0.0451	0.0500	-0.00784	0.00164
	(0.0289)	(0.0349)	(0.0254)	(0.0350)
Year 6	0.0563*	0.0480	-0.000311	0.0131
	(0.0293)	(0.0354)	(0.0259)	(0.0352)
Pooled years				
Years 3 to 6	0.0502**	0.0430	-0.0115	-0.00734
	(0.0242)	(0.0290)	(0.0210)	(0.0290)
Control: gender	No	No	No	No
# Applicants	21,734	48,272	21,766	40,276

Notes. The table presents the results from Year 1 to Year 6 after the PROSUB lotteries. PROSUB courses last two years. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(II) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (III)–(IV) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Table 6 provides evidence that local labor market conditions are important for career development: courses in faster-growing localities are associated with large and positive employment effects. More specifically, Column (I) indicates that female students (of cohort

1) in courses at high-growth municipalities present large employment gains. We detect no statistically significant effects for courses in lower-growth municipalities but coefficients are large. In addition, we find no significant difference in employment for female winners of cohort 2 in either above median or below median localities (see Columns (III) and (IV)).

We also tested how local growth affects male winners. Table A.13 in online Appendix A indicates that the negative effects for male winners take place in localities with below-median GDP growth. The negative impacts are stronger for winners in cohort 2.

The effects of local labor market conditions differ then by gender. The analysis indicates that the positive results of female workers in cohort 1 are associated with courses in municipalities with higher GDP growth. Male winners of cohort 2 were negatively affected, especially in courses located in municipalities with lower GDP growth.

5.3 Further Analysis

Public employment. We now investigate whether results are driven by employment in the public sector. One concern would be that the public sector could be "gaming" by hiring recent graduates from the program. If this is the case, the observed differential increase in employment may be "artificially" created by government hiring. We would need then to assess the net-of-government labor market impact. We are able to identify whether a person is employed in the private sector or the public sector by using the Classification of Economic Activities (CNAE) code or the legal nature of companies in RAIS dataset.

Table A.14 in the online Appendix A performs a similar exercise as the baseline analysis, but with a different dependent variable (a public employment dummy, which equals one if working in a entity related to the public sector). Results indicate that the bulk of the effects comes from private firms hiring, not from the public sector entities. The results indicate then that, although the program is funded by the public sector, private firms are guiding the results.

Course completion. We aim to provide further evidence by substituting the immediate enrollment dummy for a dummy for degree completion in Equation (2). In Table A.15 in the online Appendix A, we re-run the regressions with the dummy for degree completion instrumented by using the initial-offer dummy. The results present similar patterns to the baseline analysis: positive effect on cohort 1's female workers and the negative effects on cohort 2's male workers. However, the magnitude is greater. For instance, completing the course increases the probability employment for women of cohort 1 by 13–15 percentage points, which means a 42–48% increase in employment.

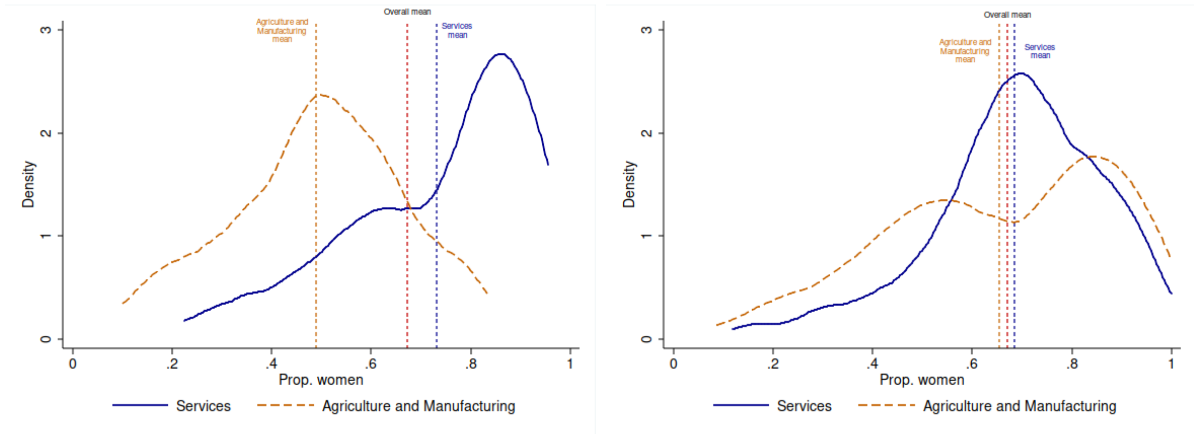
Channels. To provide suggestive evidence on the mechanisms underlying our findings, we perform three additional analyses using the dataset at the lottery level. We check whether preferences over specific characteristics of the courses (sector and "attractiveness") differ by gender. We assess if women sort into services-related courses, courses associated with a female gender bias in labor demand, and courses with higher demand per seat, which may be a proxy for potential employment opportunities.

First, we look at descriptive evidence on whether preferences for course areas differ by gender. Recall that around 65% of applicants are women. Figure 6 displays the distribution of the proportion of female applicants by lottery, dividing into two areas: lotteries of service-related courses and lotteries of agriculture- and manufacturing-related courses. For cohort 1 (Figure 6(a)), the proportion of female applicants is greater in service-related courses. On average, service courses have 76% of female applicants, while other courses have 50%. For cohort 2 (Figure 6(b)), the proportion of female applicants is similar among different areas. For cohort 1, the descriptive evidence thus suggests that women sort into services (the course which leads the positive results) since application. In cohort 2, men sort relatively more into services, the course in which they are made particularly worse off.

Second, we check for another piece of evidence on preferences for different course areas: whether female students sort into courses associated with gender bias in labor demand. Males and females tend to segregate into different industries and occupations, and gender bias in labor demand is one important factor (Olivetti and Petrongolo, 2014). As a result, course choices coupled with gender bias in labor demand can disproportionally favor female workers and can play a role in explaining our results. We match courses to occupations to create the "share of female workers in each occupation" from Brazil's 2010 Population Census. To assess the correlation, we run a cross-section OLS with the measure of gender bias in labor demand (share of female workers in a given occupation) as the dependent variable, and the proportion of female applicants in courses associated with that occupation as the explanatory variable. Table A.16 in online Appendix A shows the results. For cohort 1 (Columns (I)–(II)), we find that there is a positive and statistically significant correlation between the measure of gender bias in labor demand and the proportion of women applicants. This evidence is consistent with (Olivetti and Petrongolo, 2014) that gender bias in labor demand is one important factor explaining differential labor market performance. For cohort 2 (Columns (III)–(IV)), we do not find any correlation.

Finally, we inspect which type of lottery is driving the positive impacts for females. There is a considerable variance when it comes to the number of applicants per seat among lotteries. The number of applicants per seat can be considered a measure of the "attractiveness" of a given course. One possible hypothesis is that applicants in lotteries with more demand might perceive, among other factors, the courses as providing better potential employment opportunities. We run a cross-section OLS with applicants per seat as the dependent variable, and the proportion of women applicants as the explanatory variable. We do not find an association that the proportion of women applicants is greater in more attractive courses.

Fig. 6: Distribution of the proportion of female applicants by the type of course



(a) Cohort 1 (b) Cohort 2

Notes. This figure presents the distribution of the proportion of female applicants by lottery, dividing the lotteries into two areas: lotteries of service-related courses and lotteries of agriculture- and manufacturing-related courses.

An additional channel that might be operating is related to job search efforts. Take the case of the negative lock-in effects we find for male students, which suggest that those students seem to be giving up significant labor market experience during the training. The negative lock-in translates into a persistent negative effect on employment for those students who face lower economic growth in both *between* and *within* cohort analyses. Comparing different cohorts, our results indicate that male students of the cohort directly affected by the recession present a negative effect on labor market participation (Column (VII) of Appendix Table A.5) and a reduction in the number of months in formal employment (Column (VII) of Appendix Table A.9). Comparing students of the same cohort, our findings show that the lock-in effect translates into a persistent negative employment effect for those living in below-median growth localities (Columns (II) and (IV) of Appendix Table A.13).

We hypothesize that workers search less intensely and receive fewer new job opportunities in situations of lower economic growth (as in Oreopoulos et al., 2012). This is also consistent with findings from Acevedo et al. (2017) that men create higher expectations about future earnings—and thus we hypothesize that they may be quitting jobs more often during training and searching longer after training. Therefore, changing job search efforts and expectations can reconcile the adverse effects we find for a subgroup of male students. Further research is need to assess the role of job search and expectations.

Additional outcomes. We turn to investigating whether the program is used as stepping stones for other objectives. More specifically, we check if the program generates measurable positive results on entrepreneurship and university admission. Assessing outcomes beyond employment and earning is relevant to check the program's overall returns. We estimate a cross-section version of Equation (2) using two dependent variables. The first dependent variable is a university dummy, which equals one for those entering the a more selective university during or after the vocational training course. The second dependent variable is an entrepreneurship dummy, which equals one if a PROSUB student became an owner of any establishment during or after the vocational training.

Table 7: Effects on Additional Outcomes

	University			Entrepreneurship		
	(I)	(II)	(III)	(IV)	(V)	(VI)
	All	Female	Male	All	Female	Male
Panel (A): Cohort 1						
Year 6	0.00161	0.00253	-0.000968	0.00511	0.0114	-0.00847
	(0.00171)	(0.00208)	(0.00292)	(0.00741)	(0.00782)	(0.0167)
Control: gender	Yes	No	No	Yes	No	No
# Applicants	106,490	70,003	36,487	106,490	70,003	36,487
Panel (B): Cohort 2						
Year 6	-0.000777	-0.000858	-0.000552	-0.00458	0.000917	-0.0163
	(0.000760)	(0.000826)	(0.00167)	(0.00666)	(0.00662)	(0.0158)
Control: gender	Yes	No	No	Yes	No	No
# Applicants	97,017	62,042	34,972	97,017	62,042	34,972

Notes. Each cell of this table presents the results for the coefficient β estimated from Equation (2) for the "Year 6" after the lottery for each cohort. PROSUB courses last two years. The dependent variable in Columns (I)–(III) is a university admission dummy, while in Columns (IV)–(VI) is an entrepreneurship dummy. Panel I is for the "Cohort 1" applicants, while Panel II is for the applicants of "Cohort 2". Robust standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Table 7 reports the results of cohort 1 (Panel (A)) and cohort 2 (Panel (B)). Columns (I)–(III) suggest that there was no effect on the probability on university admission for either female or male groups. Therefore, the program seems not to be associated with further educational investments. Besides, Columns (IV)–(VI) of Table 7 displays a very small and non-significant effect on the probability of becoming entrepreneurs. The results suggest then that the program has no impact on either entrepreneurship.

Cost-benefit analysis. We perform back-of-the-envelope calculations to inform on the cost-effectiveness of the program. Our estimates guide the program's benefits, and we compare them with the program's costs. To evaluate the costs, we calculate the total labor costs of PROSUB's teachers using the information from Appendix Table A.1, which stem from RAIS dataset in 2018—the only year in which we obtained teachers' information. Yearly labor costs are approximately R$ 103 million. Notice that these labor costs correspond to teachers who (i) allocate their time for two PROSUB cohorts simultaneously in a given year and (ii) share their teaching load with another program (according to the program's implementers), so we will assume that PROSUB's labor cost is R$ 25 million. To compare with benefits, we deflate labor costs to 2010 values (approximately R$ 15 million).

To obtain the program's fiscal externalities, we simulate the revenue stream of payroll taxes over 30 years for each worker who benefited from PROSUB. We suppose that the real wage is the same over time—baseline monthly earning in 2010 was R$ 328. Payroll taxes on top of earnings correspond to 68,17% or 96,75% of formal worker's earnings depending on the labor contract regime (monthly salary or hourly wage, respectively). To calculate the benefits, we will assume that the payroll tax corresponds to 96,75% of earnings for each worker.[24] Therefore, each worker generates yearly tax revenue of R$ 3,808 or a 30-year present-value revenue stream of R$ 86,993 (supposing a discount rate of 2%).

We ask how many formal jobs the program would need to generate such that fiscal revenues would break even the program's costs. Comparing the program's labor costs and the present-value stream of tax revenue, each cohort would need to generate additional employment of 170 students. When using the number of seats, compliance rate, and our coefficient estimates for the 2012 cohort, the program generated differential employment for 118 students—i.e., approximately two-thirds of the number to break-even benefits and costs. The second cohort did not present any positive impact on labor market participation in the time frame we analyze. Therefore, the simple calculations—along with the main results of this paper—suggest that the program did not seem to generate net benefits.

6 Concluding Remarks

This paper leverages admission lotteries to study the effect of a large-sized job training program in Brazil on individual labor market outcomes. The state-sponsored program is an intensive 24-month period combining in-class and on-the-job training attributes. We use a broad set of administrative registries to compare the employment of lottery winners and non-winners over time. When assessing the impacts on cohorts that have graduated during different points of the business cycle, we find that female students fare better than their male counterparts. We also find large and persistent positive effects of the program for female students who finished the course before a recession period. In addition, results suggest that female students in service-related courses and faster-growing

[24]The federal government collects payroll taxes in Brazil. States collect sales taxes, but we do not have data firm-level data on sales.

local labor markets benefited the most from the program. Male students of the cohort directly affected by the recession present a negative effect. Inexperienced male students in service sectors and areas with slower growth are particularly negatively affected. We do not find evidence of "gaming" by the public sector by hiring students: the positive effects we found are driven by hiring from private firms. Besides, we do not find evidence that the program is used as a stepping-stone for traditional academic programs or entrepreneurial activities.

Promoting access to quality education and decent jobs are two of the United Nations' sustainable development goals. High unemployment is a major social problem because—on top of its effects on poverty and inequality—it also impacts other important dimensions such as health and well-being. This paper aims to contribute by providing evidence on a type of program that countries worldwide promote to tackle persistent unemployment and improve the skills of more vulnerable workers. For instance, according to OECD (2020), 22% of 15-19 year-olds are enrolled in vocational education on average across OECD countries. Training can be particularly crucial during economic downturns as shocks may bring about shifts in skill needs and require training to combat unemployment.

References

Acevedo, P., G. Cruces, P. Gertler, and S. Martinez (2017): "Living up to expectations: How job training made women better off and men worse off," Tech. rep., National Bureau of Economic Research.

Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2020): "Tackling Youth Unemployment: Evidence from a Labour Market Experiment in Uganda," *Econometrica*, forthcoming.

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

Attanasio, O., A. Guarín, C. Medina, and C. Meghir (2017): "Vocational training for disadvantaged youth in Colombia: A long-term follow-up," *American Economic Journal: Applied Economics*, 9, 131–43.

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

Bahia (2020): "Apresentação: Educação Profissional," *Governo do Estado da Bahia.*, Acessed on August 24, 2020. http://escolas.educacao.ba.gov.br/apresentacaoep.

Barrera-Osorio, F., A. D. Kugler, and M. I. Silliman (2020): "Hard and Soft Skills in Vocational Training: Experimental Evidence from Colombia," Working Paper 27548, National Bureau of Economic Research.

Buera, F. J., J. P. Kaboski, and M. Q. Zhao (2019): "The Rise of Services: The Role of Skills, Scale, and Female Labor Supply," *Journal of Human Capital*, 13, 157–187.

Camargo, J., L. Lima, F. Riva, and A. Souza (2020): "Technical Education, Non-cognitive Skills and Labor Market Outcomes: Experimental Evidence from Brazil," *IZA Journal of Labor Economics*.

Card, D., P. Ibarrarán, F. Regalia, D. Rosas-Shady, and Y. Soares (2011): "The labor market impacts of youth training in the Dominican Republic," *Journal of Labor Economics*, 29, 267–300.

Card, D., J. Kluve, and A. Weber (2010): "Active labour market policy evaluations: A meta-analysis," *The economic journal*, 120, F452–F477.

——— (2018): "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, 16, 894–931.

Cellini, S. and N. Turner (2019): "Gainfully Employed?: Assessing the Employment and Earnings of For-Profit College Students Using Administrative Data," *Journal of Human Resources*, 54, 342–370.

Chakravarty, S., M. Lundberg, P. Nikolov, and J. Zenker (2019): "Vocational training programs and youth labor market outcomes: Evidence from Nepal," *Journal of Development Economics*, 136, 71–110.

CODACE (2020): "Comunicado 30-10-2017," *Brazilian Business Cycle Dating Committee (CODACE).*, Acessed on August 24, 2020. https://portalibre.fgv.br/sites/default/files/2020-03/comite-de-data_o-de-ciclos-econ_micos-comunicado-de-30_10_2017-_1_.pdf.

Da Mata, D., R. Oliveira, and D. Silva (2020): "Do Job Service Programs unleash the Benefits of Vocational Training?" Mimeo, Sao Paulo School of Economics.

de Chaisemartin, C. and L. Behagel (2020): "Estimating the Effect of Treatments Allocated by Randomized Waiting Lists," *Econometrica*, 88, 1453–1477.

Deeb, A. and C. de Chaisemartin (2020): "Clustering and External Validity in Randomized Controlled Trials," Mimeo.

Field, E. M., L. L. Linden, O. Malamud, D. Rubenson, and S.-Y. Wang (2019): "Does vocational education work? Evidence from a randomized experiment in Mongolia," Tech. rep., National Bureau of Economic Research.

Hendren, N. and B. Sprung-Keyser (2020): "A Unified Welfare Analysis of Government Policies," *The Quarterly Journal of Economics*, 135, 1209–1318.

Hirshleifer, S., D. McKenzie, R. Almeida, and C. Ridao-Cano (2016): "The impact of vocational training for the unemployed: experimental evidence from Turkey," *The Economic Journal*, 126, 2115–2146.

Kugler, A., M. Kugler, J. E. Saavedra, and L. O. Herrera-Prada (2020): "Long-Term Educational Consequences of Vocational Training in Colombia: Impacts on Young Trainees and their Relatives," *Journal of Human Resources*.

Maitra, P. and S. Mani (2017): "Learning and earning: Evidence from a randomized evaluation in India," *Labour Economics*, 45, 116–130.

Ngai, L. R. and B. Petrongolo (2017): "Gender Gaps and the Rise of the Service Economy," *American Economic Journal: Macroeconomics*, 9, 1–44.

O'Connell, S. and L. Mation (2019): "Making public job training work: Evidence from a policy experiment on decentralized targeting using firm input," *Working paper.*

OECD (2020): *Education at a Glance 2020*, OECD Publishing, Paris.

Olivetti, C. and B. Petrongolo (2014): "Gender gaps across countries and skills: Demand, supply and the industry structure," *Review of Economic Dynamics*, 17, 842 – 859.

Oreopoulos, P., T. Von Wachter, and A. Heisz (2012): "The short-and long-term career effects of graduating in a recession," *American Economic Journal: Applied Economics*, 4, 1–29.

Rosenzweig, M. R. and C. Udry (2020): "External Validity in a Stochastic World: Evidence from Low-Income Countries," *The Review of Economic Studies*, 87, 343–381.

von Wachter, T. (2020): "The Persistent Effects of Initial Labor Market Conditions for Young Adults and Their Sources," *Journal of Economic Perspectives*, 34, 168–94.

ONLINE APPENDIX - NOT FOR PUBLICATION

Online Appendix to "Who benefits from job training programs? Evidence from a high-dosage program in Brazil""

Daniel Da Mata, Rodrigo Oliveira and Diana Silva

July 15, 2021

A Appendix: Figures and Tables

Fig. A.1: PROSUB Lotteries

(a) (b)

Notes. The pictures show examples of advertisement of PROSUB lotteries.

Table A.1: PROSUB: Teachers' Characteristics

PROSUB Teachers in 2018	
# of teachers	2,461
% found in RAIS registry	99.87 %
Educational attainment	
High school + vocational training	2.59%
Tertiary degree	96.71 %
Master's and Ph.D.	0.70 %
Working Experience	
Experience working outside the Education Department	86.64 %
Experience in education	19.65 %
Experience in sectors compatible with PROSUB courses	29.37 %
Years of experience working in the Education Department	
0-2 years	41.01 %
2-5 years	15.45 %
5-10 years	3.7 %
>10 years	39.83 %
Earnings in BRL	
Mean	R$ 3847.861
Max.	R$ 15895.41
Min.	R$ 974.65

Notes. This table uses RAIS data to present selected characteristics of PROSUB's teachers. See the data section for details on RAIS.

Table A.2: PROSUB Courses

Course	2012 (%)	2013 (%)	Course	2012 (%)	2013 (%)
Management	7.69	9.70	Nursing	20.25	14.86
Social service	0.07	-	Health management	0.8	0.89
Agriculture	0.06	0.07	Tourism	1.11	0.76
Land surveyor	0.15	-	Hospitality	0.06	0.09
Agroecology	0.16	0.26	TI	4.05	2.98
Agro-industry	1.07	0.43	Music	1.30	0.43
Agro-business	0.27	0.18	Oil and Gas	-	0.73
Agriculture and livestock	0.67	0.81	Logistic	5.1	8.93
Clinical analysis	7.94	5.47	Maintenance and support	1.11	-
Chemical analysis	0.56	0.45	Mechanics	0.26	0.60
Leather articles	0.03	0.15	Automobiles Mechanics	1.68	0.38
Bio-fuel	0.21	0.15	Electron-Mechanics for autos	0.81	1.19
Biotechnology	0.35	-	Environment	1.26	1.31
Sales strategy	0.42	0.50	Mining	0.41	-
Visual communication	0.62	0.06	Nutrition	6.38	4.98
Conservation and environment	0.17	-	Petrochemical	1.36	2.88
Accounting	3.13	1.59	Fashion Design	0.67	0.03
Cooperatives	0.18	-	Gaming Program	0.13	-
Cooking	0.25	1.15	Chemistry	0.57	0.69
Design in construction	1.33	0.90	Computer networking	0.6	3.73
Furniture design	0.04	-	Dentistry	2.28	3.16
Musical documentation	0.08	0.06	School Secretariat	0.4	0.36
Building	5.83	6.53	Work Safety	11.17	11.05
Electronics	0.94	1	Sales Strategy	0.12	0.25
Electrical-Mechanics	5.49	5.15	Zootechnics	0.03	0.05
Electrical	0.51	1.86			

Notes. This table presents the list of PROSUB courses and the corresponding share of candidates in the lotteries.

Table A.3: Lottery balancing: RAIS data

	All Applicants	Nonwinners	Diff. between IO Winners and Nonwinners	p-value
	(I)	(II)	(III)	(IV)
Panel (a): Cohort 1, 190 waitlists				
Age	28.709	28.710	0.143	0.423
	(0.036)	(0.037)	(0.178)	
% with high school	0.784	0.785	-0.000	0.986
	(0.002)	(0.002)	(0.010)	
Black and Indigenous dummy	0.675	0.680	0.008	0.480
	(0.002)	(0.002)	(0.011)	
Female dummy	0.539	0.537	-0.011	0.295
	(0.002)	(0.003)	(0.011)	
Number of observations	40,288	38,392	40,288	.
Panel (b): Cohort 2, 283 waitlists				
Age	28.928	28.910	0.030	0.861
	(0.040)	(0.041)	(0.173)	
% with high school	0.784	0.785	-0.001	0.902
	(0.002)	(0.002)	(0.009)	
Black and Indigenous dummy	0.679	0.687	-0.007	0.499
	(0.002)	(0.002)	(0.010)	
Gender	0.523	0.523	0.004	0.654
	(0.003)	(0.003)	(0.010)	
Number of observations	37,303	34,716	37,303	.

Notes. This table uses RAIS data for the subgroup of PROSUB applicants who had a formal employment at any point in the pre-lottery year. See the data section for details on RAIS. Panel (a): Balancing tests for 40,288 applicants, 1,896 of whom are in the initial-offer (IO) lottery winner group. Panel (b): Balancing tests for 37,303 applicants, 2,587 of whom are initial-offer (IO) lottery winners. Data are provided at the individual level. Column (I) shows the mean and standard deviation (in parentheses) of baseline characteristics for all the applicants. Column (II) presents the mean and standard deviation (in parentheses) of baseline characteristics of nonwinners applicants (replacement group plus never-offer group). Column (III) reports the coefficients estimated from Equation (1) and the last column the corresponding p-value. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: First-Stage Results

	Cohort 1		Cohort 2	
	Enrollment dummy	Degree completion dummy	Enrollment dummy	Degree completion dummy
Initial-Offer dummy	0.3533***	0.1116***	0.3537***	0.1290***
	(0.0066)	(0.0044)	(0.0056)	(0.0040)
Observations	106,495	106,495	97,017	97,017
F-statistics	1,459	324	1,980	517
Strata fixed effects	Yes	Yes	Yes	Yes
Control: Gender	Yes	Yes	Yes	Yes

Notes. The table presents the first-stage results. The initial-offer dummy equals one for the applicant getting an initial offer at the time of the lottery. The enrollment dummy equals one if the individual enrolled at a PROSUB course (after being selected by the PROSUB lotteries). The degree completion equals one if the individual finished the course after being selected by the PROSUB lotteries. Robust standard errors (in parentheses). *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.5: Baseline estimates: Effects on Employment

	Cohort 1				Cohort 2			
	(I) All	(II) Female	(III) Male	(IV) p-value	(V) All	(VI) Female	(VII) Male	(VIII) p-value
During course								
Year 1	-0.0182	0.00344	-0.0606*	0.123	-0.0484***	-0.0277	-0.0981***	0.086
	(0.0185)	(0.0215)	(0.0355)		(0.0176)	(0.0198)	(0.0359)	
Year 2	-0.00770	0.00275	-0.0283	0.461	-0.0514***	-0.0392*	-0.0779**	0.347
	(0.0189)	(0.0221)	(0.0358)		(0.0178)	(0.0203)	(0.0357)	
After course								
Year 3	0.00422	0.0447**	-0.0856**	0.002	-0.0374**	-0.0229	-0.0712**	0.245
	(0.0192)	(0.0226)	(0.0361)		(0.0180)	(0.0205)	(0.0361)	
Year 4	0.0155	0.0429*	-0.0437	0.041	-0.0259	-0.0174	-0.0438	0.526
	(0.0192)	(0.0226)	(0.0360)		(0.0180)	(0.0203)	(0.0362)	
Year 5	0.0189	0.0474**	-0.0436	0.032	-0.0201	-0.00409	-0.0573	0.202
	(0.0191)	(0.0224)	(0.0362)		(0.0181)	(0.0207)	(0.0362)	
Year 6	0.0434**	0.0525**	0.0231	0.488	-0.0141	0.00499	-0.0613*	0.113
	(0.0192)	(0.0227)	(0.0358)		(0.0182)	(0.0209)	(0.0362)	
Pooled years								
Years 3 to 6	0.0205	0.0469**	-0.0375	0.014	-0.0244	-0.00987	-0.0584**	0.153
	(0.0157)	(0.0187)	(0.0288)		(0.0149)	(0.0171)	(0.0293)	
Control: gender	Yes	No	No	-	Yes	No	No	-
# Applicants	106,495	70,006	36,489	-	97,017	62,042	34,972	-

Notes. The table presents the results from Year 1 to Year 6 after the PROSUB lotteries. PROSUB courses last two years. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(IV) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Column (IV) shows the p-value of the difference in the coefficients of female and male regressions (in Columns (II) and (III)). Columns (V)–(VIII) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Column (VIII) shows the p-value of the difference in the coefficients between female and male regressions (in Columns (VI) and (VII)). Robust standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Effects on Employment: ANCOVA specification

	Cohort 1			Cohort 2		
	(I) All	(II) Female	(III) Male	(IV) All	(V) Female	(VI) Male
During course						
Year 1	-0.0179	-0.00269	-0.0495	-0.0472***	-0.0273	-0.0959***
	(0.0170)	(0.0197)	(0.0330)	(0.0164)	(0.0184)	(0.0337)
Year 2	-0.00746	-0.00240	-0.0193	-0.0503***	-0.0389**	-0.0760**
	(0.0177)	(0.0206)	(0.0336)	(0.0169)	(0.0193)	(0.0339)
After course						
Year 3	0.00445	0.0402*	-0.0771**	-0.0365**	-0.0226	-0.0696**
	(0.0182)	(0.0215)	(0.0344)	(0.0173)	(0.0197)	(0.0345)
Year 4	0.0158	0.0386*	-0.0355	-0.0250	-0.0171	-0.0421
	(0.0184)	(0.0216)	(0.0344)	(0.0173)	(0.0196)	(0.0349)
Year 5	0.0191	0.0433**	-0.0360	-0.0192	-0.00381	-0.0558
	(0.0184)	(0.0215)	(0.0350)	(0.0175)	(0.0200)	(0.0351)
Year 6	0.0436**	0.0487**	0.0306	-0.0133	0.00524	-0.0598*
	(0.0187)	(0.0221)	(0.0347)	(0.0178)	(0.0203)	(0.0353)
Pooled years						
Years 3 to 6	0.0207	0.0427**	-0.0295	-0.0235*	-0.00959	-0.0568**
	(0.0148)	(0.0176)	(0.0270)	(0.0141)	(0.0162)	(0.0278)
Control: gender	Yes	No	No	Yes	No	No
# Applicants	106,495	70,006	36,489	97,017	62,042	34,972

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. The regressions include pre-treatment values of labor market participation. Columns (I)–(III) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (IV)–(VI) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Effects on Employment: Robustness without controls

	Cohort 1	Cohort 2
	(I) All	(II) All
During course		
Year 1	-0.0165	-0.0501***
	(0.0186)	(0.0177)
Year 2	-0.00608	-0.0531***
	(0.0190)	(0.0180)
After course		
Year 3	0.00597	-0.0391**
	(0.0193)	(0.0182)
Year 4	0.0173	-0.0274
	(0.0193)	(0.0181)
Year 5	0.0206	-0.0217
	(0.0192)	(0.0183)
Year 6	0.0451**	-0.0157
	(0.0193)	(0.0184)
Pooled years		
Years 3 to 6	0.0222	-0.0260*
	(0.0158)	(0.0151)
Control: gender	No	No
# Applicants	106,495	97,017

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. The regressions do not include any control. Columns (I)–(III) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (IV)–(VI) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses). *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.8: Effects on Employment: Clustering at the Lottery Level

	Cohort 1			Cohort 2		
	(I) All	(II) Female	(III) Male	(IV) All	(V) Female	(VI) Male
During course						
Year 1	-0.0182	0.00344	-0.0606*	-0.0484***	-0.0277	-0.0981***
	(0.0176)	(0.0221)	(0.0368)	(0.0176)	(0.0213)	(0.0349)
Year 2	-0.00770	0.00275	-0.0283	-0.0514***	-0.0392*	-0.0779**
	(0.0184)	(0.0211)	(0.0386)	(0.0185)	(0.0205)	(0.0354)
After course						
Year 3	0.00422	0.0447**	-0.0856**	-0.0370*	-0.0230	-0.0701*
	(0.0189)	(0.0222)	(0.0377)	(0.0192)	(0.0211)	(0.0419)
Year 4	0.0155	0.0429**	-0.0437	-0.0259	-0.0174	-0.0437
	(0.0190)	(0.0214)	(0.0391)	(0.0202)	(0.0219)	(0.0406)
Year 5	0.0189	0.0474**	-0.0436	-0.0203	-0.00403	-0.0584
	(0.0201)	(0.0239)	(0.0392)	(0.0189)	(0.0221)	(0.0375)
Year 6	0.0434**	0.0525**	0.0231	-0.0141	0.00499	-0.0613
	(0.0201)	(0.0241)	(0.0355)	(0.0197)	(0.0220)	(0.0427)
Pooled years						
Years 3 to 6	0.0205	0.0469**	-0.0375	-0.0243	-0.00985	-0.0584*
	(0.0161)	(0.0190)	(0.0317)	(0.0164)	(0.0184)	(0.0338)
Control: gender	Yes	No	No	Yes	No	No
# Applicants	106,495	70,006	36,489	97,017	62,042	34,972

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(III) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (IV)–(VI) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses) are clustered at the lottery level. *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Effects on the Intensive Margin Number of Months in Formal Jobs

	Cohort 1				Cohort 2			
	(I) All	(II) Female	(III) Male	(IV) p-value	(V) All	(VI) Female	(VII) Male	(VIII) p-value
During course								
Year 1	-0.174	0.0882	-0.700*	0.0787	-0.247	-0.186	-0.378	0.662
	(0.195)	(0.224)	(0.389)		(0.185)	(0.204)	(0.389)	
Year 2	0.133	0.244	-0.0483	0.529	-0.479**	-0.304	-0.880**	0.203
	(0.207)	(0.241)	(0.398)		(0.193)	(0.216)	(0.398)	
After course								
Year 3	-0.0213	0.273	-0.659	0.0481	-0.453**	-0.275	-0.839**	0.223
	(0.209)	(0.243)	(0.404)		(0.199)	(0.224)	(0.405)	
Year 4	0.0519	0.318	-0.513	0.0830	-0.273	-0.232	-0.347	0.807
	(0.214)	(0.251)	(0.408)		(0.202)	(0.227)	(0.412)	
Year 5	0.268	0.622**	-0.513	0.0192	-0.221	-0.140	-0.402	0.579
	(0.217)	(0.254)	(0.413)		(0.204)	(0.230)	(0.414)	
Year 6	0.511**	0.734***	0.0294	0.145	-0.221	0.0634	-0.901**	0.0413
	(0.218)	(0.256)	(0.410)		(0.205)	(0.234)	(0.411)	
Pooled years								
Years 3 to 6	0.203	0.487**	-0.414	0.0226	-0.292*	-0.146	-0.622*	0.221
	(0.178)	(0.209)	(0.335)		(0.169)	(0.192)	(0.339)	
Control: gender	Yes	No	No	-	Yes	No	No	-
# Applicants	106,495	70,006	36,489	-	97,017	62,042	34,972	-

Notes. The table presents the results from Year 1 to Year 6 after the PROSUB lotteries. PROSUB courses last two years. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is the number of months in formal jobs. Columns (I)–(IV) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Column (IV) shows the p-value of the difference in the coefficients of female and male regressions (in Columns (II) and (III)). Columns (V)–(VIII) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Column (VIII) shows the p-value of the difference in the coefficients between female and male regressions (in Columns (VI) and (VII)). Robust standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1

Table A.10: Effects on Earnings

	Cohort 1		Cohort 2	
	(I) Total Earnings	(II) Positive Earnings	(III) Total Earnings	(IV) Positive Earnings
During course				
Year 1	-19.39	10.50	-62.53**	-26.70
	(25.22)	(36.21)	(25.21)	(35.76)
Year 2	-56.08**	-33.97	-64.66**	-7.395
	(26.91)	(35.51)	(28.97)	(41.13)
After course				
Year 3	-13.20	-14.61	-76.27**	-54.51
	(32.94)	(42.37)	(35.02)	(48.67)
Year 4	-27.98	-30.88	-9.039	68.32
	(34.57)	(42.64)	(38.31)	(53.36)
Year 5	34.72	-11.65	4.342	109.6*
	(37.98)	(46.48)	(40.43)	(56.39)
Year 6	17.26	-38.27	17.26	99.15*
	(42.92)	(55.08)	(41.75)	(55.30)
Pooled years				
Years 3 to 6	2.698	-23.58	-15.93	53.66
	(31.85)	(38.82)	(33.55)	(45.19)
Control: gender	Yes	Yes	Yes	Yes
# Applicants	106,495	64,699	97,017	56,596

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable in Columns (I) and (III) is total earnings, while in Columns (II) and (IV) is total earnings for those who are employed. Earnings are measured in BRL. Columns (I)–(II) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (III)–(IV) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.11: Effects on Employment: Experience (Male)

	Cohort 1		Cohort 2	
	(I) Experience	(II) No experience	(III) Experience	(IV) No experience
During course				
Year 1	-0.0571	-0.0497	-0.0500	-0.136***
	(0.0393)	(0.0548)	(0.0426)	(0.0523)
Year 2	-0.0168	-0.0605	0.0325	-0.205***
	(0.0393)	(0.0616)	(0.0414)	(0.0563)
After course				
Year 3	-0.0374	-0.164**	-0.0193	-0.143**
	(0.0397)	(0.0642)	(0.0432)	(0.0568)
Year 4	-0.0163	-0.0814	-0.0358	-0.0553
	(0.0403)	(0.0649)	(0.0447)	(0.0564)
Year 5	-0.0128	-0.0774	-0.0582	-0.0651
	(0.0412)	(0.0640)	(0.0439)	(0.0577)
Year 6	0.0162	0.0551	-0.0793*	-0.0313
	(0.0405)	(0.0654)	(0.0444)	(0.0581)
Pooled years				
Years 3 to 6	-0.0126	-0.0669	-0.0482	-0.0738
	(0.0308)	(0.0511)	(0.0351)	(0.0451)
Control: gender	No	No	No	No
# Applicants	24,941	11,544	22,135	12,818

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(II) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (III)–(IV) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Students with previous experience have worked at least three years before the PROSUB lottery. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.12: Heterogeneous Effects: Type of Course (Male)

	Cohort 1		Cohort 2	
	Agriculture & Manufactur.	Services	Agriculture & Manufactur.	Services
During course				
Year 1	-0.135**	-0.0187	-0.0976*	-0.0985*
	(0.0579)	(0.0449)	(0.0508)	(0.0509)
Year 2	-0.0706	-0.00401	-0.101**	-0.0562
	(0.0572)	(0.0457)	(0.0504)	(0.0504)
After course				
Year 3	-0.194***	-0.0223	-0.0257	-0.114**
	(0.0588)	(0.0456)	(0.0508)	(0.0513)
Year 4	-0.0830	-0.0195	0.0305	-0.113**
	(0.0578)	(0.0459)	(0.0508)	(0.0518)
Year 5	-0.0209	-0.0562	0.00749	-0.118**
	(0.0576)	(0.0464)	(0.0509)	(0.0515)
Year 6	0.0544	0.00725	-0.0167	-0.103**
	(0.0576)	(0.0456)	(0.0511)	(0.0513)
Pooled years				
Years 3 to 6	-0.0608	-0.0227	-0.00112	-0.112***
	(0.0466)	(0.0364)	(0.0409)	(0.0421)
Control: gender	No	No	No	No
# Applicants	16,075	20,368	11,564	22,032

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(III) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (IV)–(VI) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses) are clustered at the lottery level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.13: Heterogeneous Effects: Local economic growth (Male)

	Cohort 1		Cohort 2	
	(I) Above median: GDP growth	(II) Below median: GDP growth	(III) Above median: GDP growth	(IV) Below median: GDP growth
During course				
Year 1	-0.0818*	-0.0321	-0.120**	-0.0737
	(0.0448)	(0.0577)	(0.0481)	(0.0539)
Year 2	0.00754	-0.0761	-0.0612	-0.0960*
	(0.0445)	(0.0591)	(0.0478)	(0.0534)
After course				
Year 3	-0.0763*	-0.0981*	-0.0382	-0.107**
	(0.0449)	(0.0593)	(0.0482)	(0.0541)
Year 4	0.00743	-0.112*	0.0252	-0.119**
	(0.0442)	(0.0597)	(0.0483)	(0.0545)
Year 5	0.0114	-0.117*	0.00724	-0.128**
	(0.0446)	(0.0598)	(0.0483)	(0.0543)
Year 6	0.0489	-0.0114	0.0386	-0.170***
	(0.0443)	(0.0590)	(0.0481)	(0.0549)
Pooled years				
Years 3 to 6	-0.00214	-0.0847*	0.00823	-0.131***
	(0.0359)	(0.0470)	(0.0391)	(0.0441)
Control: gender	No	No	No	No
# Applicants	10,548	25,941	11,271	23,701

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(II) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (III)–(IV) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses) are clustered at the lottery level. *** p<0.01, ** p<0.05, * p<0.1

Table A.14: Employment in the Public Sector

	Cohort 1			Cohort 2		
	(I) All	(II) Female	(III) Male	(IV) All	(V) Female	(VI) Male
During course						
Year 1	-0.00237	-0.00219	0.000486	0.000982	0.0125	-0.0193
	(0.0103)	(0.0116)	(0.0208)	(0.00981)	(0.0113)	(0.0193)
Year 2	0.00806	0.00265	0.0233	-0.00154	0.0110	-0.0222
	(0.0110)	(0.0126)	(0.0216)	(0.0101)	(0.0115)	(0.0204)
After course						
Year 3	0.00903	0.00161	0.0290	-0.00228	0.0107	-0.0270
	(0.0113)	(0.0128)	(0.0228)	(0.0102)	(0.0116)	(0.0205)
Year 4	0.0162	0.00998	0.0308	0.00145	0.0103	-0.0131
	(0.0115)	(0.0131)	(0.0232)	(0.00980)	(0.0111)	(0.0198)
Year 5	0.0139	0.0114	0.0218	-0.000218	0.00745	-0.00776
	(0.0112)	(0.0128)	(0.0223)	(0.0109)	(0.0123)	(0.0218)
Year 6	0.0110	0.00377	0.0292	0.00415	0.00694	0.00591
	(0.0122)	(0.0139)	(0.0243)	(0.0113)	(0.0127)	(0.0227)
Pooled years						
Years 3 to 6	0.0125	0.00669	0.0277	0.000777	0.00886	-0.0105
	(0.0104)	(0.0118)	(0.0210)	(0.00952)	(0.0107)	(0.0192)
Control: gender	Yes	No	No	Yes	No	No
# Applicants	106,495	70,006	36,489	97,017	62,042	34,972

Notes. Each cell corresponds to a LATE effect where the enrollment dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy in any entity related to the public sector. Columns (I)–(III) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (IV)–(VI) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses) are clustered at the lottery level. *** p<0.01, ** p<0.05, * p<0.1

Table A.15: Effects on Employment: Course Completion

	Cohort 1			Cohort 2		
	(I) All	(II) Female	(III) Male	(IV) All	(V) Female	(VI) Male
During course						
Year 1	-0.0577	0.0103	-0.224*	-0.136***	-0.0703	-0.361***
	(0.0584)	(0.0641)	(0.132)	(0.0492)	(0.0501)	(0.134)
Year 2	-0.0244	0.00820	-0.105	-0.144***	-0.0997*	-0.286**
	(0.0597)	(0.0659)	(0.133)	(0.0499)	(0.0514)	(0.133)
After course						
Year 3	0.0134	0.133**	-0.317**	-0.105**	-0.0583	-0.262*
	(0.0606)	(0.0677)	(0.135)	(0.0505)	(0.0521)	(0.134)
Year 4	0.0492	0.128*	-0.162	-0.0724	-0.0443	-0.161
	(0.0607)	(0.0676)	(0.133)	(0.0503)	(0.0517)	(0.134)
Year 5	0.0600	0.141**	-0.162	-0.0562	-0.0104	-0.211
	(0.0605)	(0.0670)	(0.135)	(0.0509)	(0.0525)	(0.134)
Year 6	0.137**	0.156**	0.0855	-0.0394	0.0127	-0.225*
	(0.0609)	(0.0679)	(0.132)	(0.0511)	(0.0531)	(0.135)
Pooled years						
Years 3 to 6	0.0650	0.140**	-0.139	-0.0682	-0.0251	-0.215**
	(0.0497)	(0.0560)	(0.107)	(0.0418)	(0.0434)	(0.109)
Control: gender	Yes	No	No	Yes	No	No
# Applicants	106,495	70,006	36,489	97,017	62,042	34,972

Notes. Each cell corresponds to a LATE effect where the course completion dummy is instrumented by using the random variation results of the lottery. In rows from "Year 1" to "Year 6", each cell corresponds to the estimated coefficient of Equation (2). In row "Years 3 to 6" (Pooled Years), each cell corresponds to the estimated coefficient of Equation (4). The dependent variable is an employment dummy. Columns (I)–(III) are for the "Cohort 1" applicants: the years 2012 and 2013 are the course years, while the after courses years are from 2014 to 2017. Columns (IV)–(VI) are for the applicants of "Cohort 2": courses years are 2013 and 2014, while the after course years are from 2015 to 2018. Robust standard errors (in parentheses) are clustered at the lottery level. *** p<0.01, ** p<0.05, * p<0.1

Table A.16: Channels: Correlation with Gender Bias Labor Demand and Courses' Attractiveness

	Gender Bias Labor Demand analysis				Course Attractiveness analysis			
	Cohort 1		Cohort 2		Cohort 1		Cohort 2	
	Share of female workers (I)	Share of female workers (II)	Share of female workers (III)	Share of female workers (IV)	Applicants per Seat (V)	Applicants per Seat (VI)	Applicants per Seat (VII)	Applicants per Seat (VIII)
Share of female applicants	0.0043***	0.0051***	0.0022	0.0008	-14.0653	-0.3195	-2.8115	-1.7446
	(0.0011)	(0.0014)	(0.0018)	(0.0021)	(9.4548)	(9.7377)	(7.3167)	(8.3074)
Observations	190	190	283	283	190	190	283	283
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes. The table presents the results of a cross-section OLS with the share of females workers in a given occupation as the dependent variable (columns (I) to (IV)) and the proportion of female applicants in lotteries associated with that occupation as the explanatory variable. The dependent variable in columns (V) to (VIII) is the number of applicants per seat in each lottery. Regressions are at the lottery level. There are 190 lotteries for cohort 1 and 283 lotteries for cohort 2. The odd-numbered columns present results without controls, while the even-numbered columns present results controlling for the municipalities where each course is located. Robust standard errors (in parentheses). *** $p<0.01$, ** $p<0.05$, * $p<0.1$