IDB WORKING PAPER SERIES Nº IDB-WP-1156

The Long-Term Effects of Job Training on Labor Market and Skills Outcomes in Chile

Annabelle Doerr
Rafael Novella

Inter-American Development Bank
Labor Markets Division

September 2020

The Long-Term Effects of Job Training on Labor Market and Skills Outcomes in Chile

Annabelle Doerr
Rafael Novella

Inter-American Development Bank
Labor Markets Division

September 2020

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Doerr, Annabelle.
The long-term effects of job training on labor market and skills outcomes in Chile / Annabelle Doerr, Rafael Novella.
p. cm. — (IDB Working Paper Series ; 1156)
Includes bibliographic references.
1. Employees-Training of-Chile. 2. Labor market-Chile. 3. Income-Chile. 4. Skilled labor-Chile. I. Novella, Rafael. II. Inter-American Development Bank. Labor Markets Division. III. Title. IV. Series.
IDB-WP-1156

JEL Codes: J08, J24, H43, O15
Keywords: Job training, skills, RCT, administrative data

http://www.iadb.org

The Long-Term Effects of Job Training on Labor Market and Skills Outcomes in Chile*

Annabelle Doerr[†‡] Rafael Novella[§¶‖]

Abstract

Job training programs can be an effective policy for improving productivity and labor market outcomes in low and middle income countries. We report medium and long-term impacts of a job training program for vulnerable workers in Chile on labor market and skill outcomes using experimental and administrative data. We find that the program fails on improving workers' skills and most labor outcomes but some evidence of a effect on labor income. We also find evidence of heterogeneous effects by course-type, training provider quality, and gender. This evidence aims at contributing to a better design of training programs and to a better use of public resources.

Keywords: Job training, Skills, RCT, Administrative data
JEL Codes: J08, J24, H43, O15

*The study uses the Chilean Unemployment Insurance database. We thank the Department of Employment of the Chilean Ministry of Labor and Pensions for dataset access. We also thank Macarena Alvarado, David Kaplan, and Graciana Rucci for their key contribution to the design and implementation of the program evaluation. Belén Conde provided excellent research assistance. The authors are responsible for all results and views, which do not represent the Ministry of Labor and Pensions or the Inter-American Development Bank. All the information used in this paper was kept anonymous. We do not use any data with individual indicators. The data were stored and managed on a secure server.

[†]UC Berkeley
[‡]University of Basel
[§]Inter-American Development Bank
[¶]University of Oxford (Oxford Department of International Development, ODID & the Centre on Skills, Knowledge and Organisational Performance, SKOPE)
[‖]University College London (UCL)

1 Introduction

Although education coverage in Latin America and the Caribbean (LAC) has substantially expanded in the last thirty years, the region still face the challenge of improving the quality and relevance of education. These deficiencies are reflected in the poor performance in standardized international tests, skills gaps reported by employers, high levels of informality, unemployment and low productivity of countries in the region. In this context, job training has received a lot of attention in LAC, as an effective policy aiming at redressing, in a short time, the lack of skills obtained during formal education and improving adults' labor outcomes (Escudero et al., 2019). Evaluations of such training programs in Argentina, the Dominican Republic, and Colombia show, in general, positive effects on different labor outcomes in the short-term (Attanasio et al., 2011; Card et al., 2011; Ibarrarán and Rosas Shady, 2009) that, according to the few long-term evaluations, not improve over time (Attanasio et al., 2017; Kugler et al., 2015; Alzúa et al., 2016; Ibarrarán et al., 2018). In this paper we evaluate the medium- and long-term effectiveness of a large training program in Chile on skills development and labor outcomes.

Chile is a pioneer country in LAC in implementing job training programs. However, there is only scarce evidence about their effectiveness to improve labor market outcomes. Available evidence from non-experimental evaluations in Chile show positive results on overall earnings and employment probabilities (Aedo and Pizarro, 2004; Centro Microdatos, 2006, 2008) and larger effects for vulnerable groups, as youths (Aedo and Pizarro, 2004) and less-educated workers (Novella et al., 2017). A randomized experiment gives us the opportunity to present evidence on the effectiveness of vocational training in one of the most dynamic economies in Latin-America. The *Formacion para el Trabajo* (FOTRAB) program was introduced in 2012 and persisted until 2018. It mainly provided four to six months of classroom and on-the-job or technical assistance training to women and men between the ages of 18 and 65 living in vulnerable conditions.

This study is based on a sample of more than 5,000 applicants to FOTRAB who were randomly assigned to receive treatment or being part of the control group in one cohort of applicants in 2013 using an over-subscription design. In the evaluation, we focus on 123 courses that were offered in three regions in Chile. The applicants were surveyed before training started, and 18 and 35 months after application. The surveys include tests of cognitive and non-cognitive skills that allows us to study the programs impact on skill development, in addition to labor market outcomes, which is important to gain insights about the mechanisms of the program. Moreover, we enrich our analysis with administrative data to evaluate the long-term effects of training (more than 4 years after application). This also serves as a robustness check for the validity of the self-reported labor market outcomes of survey participants.

We find that being assigned to and participate in the program has no effect on paid- or self-employment. Labor income increases in the short-run, but vanishes over time. When we distinguish the overall effect into the three different course types, we find that the most basic classroom training has a positive long-term effect on labor income. The course that combines classroom training with a technical assistance component increases the short-term probability to be self-employed and the probability of formal employment in the long-run. For the course that combines classroom training and on the job-training, the findings are more negative: the probability to be in paid employment, being formally employed, and also labor income significantly decrease. Interestingly, we find that the training provider quality is positively associated with a larger probability of being self-employed and labor income. The results from a subgroup analysis show that FOTRAB is more effective for males than for females. Moreover, the limited effects of the program could be related with its poor impact on skill-development.

This study contribute to the literature evaluating long-term effects of training programs in developing countries. We provide the first experimental evidence on a large-scale training program targeted to individuals living in vulnerable conditions without restricting the eligibility to a specific age-group or unemployment status in Chile. Instead, it covers unemployed and employed, women and men as well as young and older individuals. Furthermore by combining survey data and administrative records, we can observe individuals short, medium, and long-term labor market outcomes up to four years after training. Finally, by using data from training providers, we are able to explore whether labor market outcomes vary by providers' quality.

This article is organized as follows: Section 2 describes the program and the design of the experimental evaluation. Section 3 presents the data used and analyses sample attrition. Section 4 analyses sample balance and non-compliance. Section 5 discusses the empirical strategy. Section 6 presents the results on labor outcomes and Section 7 presents the effects on skills and discusses heterogeneous effects. Finally, Section 8 concludes.

2 Program description and experimental design

In the last three decades, Chile has been one of LAC's fastest-growing economies. It has reduced the poverty rate from 36% in 2000 to 8.6% in 2017, and is classified as a high-income economy, with a GDP per capita of US$ 15,923. However, the country still faces important challenges in term of reducing high inequality (the GINI index was 46.6 in 2017) and improving human capital. According to the World Banks' Human Capital Index, the country performs better than the average for LAC but worse than the average of the high-income countries group in helping new generations achieving their

human capital potential.[1] Compared to the other OECD countries, Chile shows a lower proportion of adults who have completed secondary education and a lower average score achieved by students in reading, literacy, maths and sciences in the PISA test.[2] In 2012, the overall employment rate in Chile was 55.7%, with much lower rates for females (43.4%) and youth (31.1%). Although unemployment was relatively low in the country (6.7% in 2012), informal employment, measured as the percentage of workers no contributing to social security, has been around 35% in the last decade.[3]

2.1 Formación para el Trabajo (FOTRAB)

The origins of FOTRAB date back to 1997, when it was implemented as the *Programa Especial de Jóvenes* (PEJ), a subsidized training program aimed at improving job skills, labor market integration and employability of young people (18 to 29 years) living in vulnerable conditions. Motivated by the positive results of non-experimental evaluations of PEJ (Jara, 2001; Centro Microdatos, 2006, 2008; Ministerio de Trabajo y Previsión Social, 2011), the program was expanded over time.[4] In 2011, the PEJ was extended to unemployed women aged 25-49 years and men and women older than 50 years. Over a short period of one year, the program operated under the title *Formación en Oficios*. In 2012, it was renamed to *Formación para el Trabajo* (FOTRAB) and covered among its beneficiaries men and women aged 18 to 65 and living in vulnerable conditions. Vulnerability was defined as belonging to the first two quintiles of the *Ficha de Protección Social* (FPS) which was until 2016 the country-wide tool for targeting public social programs.[5] FOTRAB offered job training to all individuals living in vulnerable conditions independent of their employment status. In addition, the program covered young parents between 16 and 17 years old. However, individuals with the following characteristics, independent of their FPS score, were excluded: individuals with completed tertiary education, current students in tertiary education, and those who participated in the program in the year before application or were participating at the moment of application.

The program was financed by the Chilean government and administered by the National Training and Employment Service (SENCE) all over Chile. SENCE regional offices identified the courses that needed to be offered under consideration of regional labor demand, designed the curriculum, and opened a call to private training providers *Organismos*

[1] The World Bank https://data.worldbank.org/country/chile (accessed 17 December 2019)
[2] OECD https://data.oecd.org (accessed 17 December 2019)
[3] Inter-American Development Bank https://www.iadb.org/en/sector/social-investment/sims/home (accessed 17 December 2019)
[4] Using non-experimental methods (before-after comparison, matching and difference-in-difference), these evaluations find that PEJ is associated with a reduction in the probability of being unemployed and with an increase in the probability of being employed, being formally employed and labor income.
[5] Eligible individuals must had a FPS score of 11,734 points or lower.

Técnicos de Capacitación (OTEC) that could apply for the practical implementation of the courses. In the call, SENCE specified the subject of training and the topics that must be included. Since the main objective of the program is to improve the skills and labor market outcomes of participants the curriculum is build to improve participants human capital by providing them with cognitive and non-cognitive skills in addition to technical skills specific to the occupation. The program should increase the chances of unemployed individuals to find employment, support individuals to transit from informality to formal employment and to increase the wages of those currently employed, establish more stable employment relationships and provide the necessary skills to individuals who plan to become self-employed. Interested OTECs applied by presenting a proposal of how to practically implement the program and achieve the programs objectives. Finally, the proposals were evaluated by SENCE and ones with top scores were hired and paid for each trained participant. The first cohort entered the program in 2012 and the program was implemented until 2018.

Table 1: Training content and features by course type

I. Classroom training (CT)	
(1) Classroom training (min. 250 hours) (2) Accident insurance (3) Food and transportation subsidy (US$ 6 daily) (4) Certificates and licenses	
II. CT plus job training (CJT)	III. CT plus technical assistance (CTA)
(5) On-the-job training (360 hours) (6) Job-finding support	(5) Technical assistance (min. 40 hours) (6) Subsidy for tools (US$ 404)

Note: There are three different course types. Contents (1)-(4) are common to all three course types. CJT and CTA offers additional two components, respectively.

Eligible individuals could apply to a freely chosen subsidized job training course from a set of available courses in each region. The average duration of the offered courses ranged between four to six months. The courses can be classified in three types (see Table 1). The first type is classroom training (CT). Courses of this type mainly took place in the classroom and consists of a minimum of 250 hours of training. On average, classroom training courses lasted 500 hours and individuals spent four hours per day in the classroom. At least 60% of the total duration was devoted to teaching technical skills specific to the trained occupation, while the remaining time was spent to teach transferable skills (e.g., non-cognitive skills, digital and financial knowledge). The training providers were allowed to vary the distribution of hours between these modules and teaching methodologies conditional on offering training on specific technical and transferable skills. Participants

were provided an accident insurance, a daily subsidy of US$6 for transport and food[6] and certificates or licenses in case the occupation require them. These benefits were also offered in the two other course types.

The second course type consists of classroom training plus job training (CJT). In addition to the classroom component these courses offer 360 hours of on-the-job training and job-finding support. The focus of CT and CJT courses is on finding paid employment. The third course type consists of classroom training plus technical assistance (CTA). In contrast to the other two types, these courses are designed for individuals interested in being self-employed. In addition to the content taught in the classroom, the curriculum covers at least 40 hours of technical assistance (covering, for instance, topics like self-marketing, branding, and accounting) to improve self-employability and a subsidy of up to US$404 for buying tools for those who finalized the technical assistance phase. In 2013, the average cost per trainee to FOTRAB was about US$3,200.

2.2 Experimental design

The experimental evaluation of the program was designed in 2012, under the framework of a loan between SENCE and the Inter-American Development Bank (IDB), and motivated by the evolution and importance of the program and the absence of experimental evaluations of training programs in the country. The experiment took place in one of the three cohorts of applicants to FOTRAB in 2013 (*Llamado 22*). The experimental design involved several steps. First, a sample of 123 courses offered in three regions (Biobío, Metropolitana and Valparaíso) and organized by 12 training providers were selected. SENCE decided to implement the evaluation in the three regions with larger demand. The selection of courses was guided by two criteria: selected courses needed to represent the diversity of courses offered by FOTRAB and they had to show enough demand in previous *Llamados* to ensure that randomization by excess of applications could be implemented. The selected courses represent 46% of courses in the three regions (and 31% of the total courses implemented in the country in that *Llamado*). The type of offered courses varied by region and provider. A large majority of 67% were CTJ courses, 20% were CTA courses, and only 13% of pure CT courses. A distribution of course types by region and OTEC is presented in Table A.1 in Appendix A.

Second, training providers were asked to prepare a list of eligible applicants equivalent to twice the capacity of each training course. Third, SENCE/IDB received the list of eligible candidates and randomly assigned them to the treatment or the control group. Randomization was realized within each course and stratified by gender. Overall, 5,193 applicants were randomly assigned to the control and treatment group (columns 1 and 2

[6] The 2013 average exchange rate of US$1 = 495 Chilean Pesos (CLP) is used in the paper.

Table 2: Original randomization, realized assignment, and participation

	Original randomization		Realized assignment			Realized participation		
	Treatment Z* = 1	Control Z* = 0	Assigned Z = 1	Re-assigned Z = 1	Not-assigned Z = 0	Participant D = 1	No-show D = 0	Control D = 0
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All	2,631	2,562	2,631	1,020	1,542	2,464	1,187	1,542
CT	330	321	330	106	215	311	125	215
CJT	1,701	1,675	1,701	760	915	1,561	900	915
CTA	600	566	600	154	412	592	162	412

Note: Z* indicates original randomization. Training providers had the opportunity to fill empty slots of assigned applicants that did not show-up or dropped-out early. Z indicates if an applicant was randomly assigned to participate in a course either by original assignment or by re-assignment. D indicates participation in a course. The evaluation is based on the realized assignment. CT is the abbreviation for classroom training, CJT is classroom plus job training, and CTA is classroom plus technical assistance.

of Table 2). Fourth, the list of individuals assigned to the treatment group was sent to the training providers to proceed with their registration and to start the course. From applicants who were originally assigned to participate in the program about 36% never started or dropped-out early. Every time when an applicant originally assigned to the treatment group was not longer available or dropped-out within the initial 20% of the training, the provider should contact SENCE/IDB who would randomly select a replacement out of the list of controls.

The implementation of the evaluation followed the experimental design with the exception of the fourth step. When receiving the lists of applicants assigned to the treatment groups OTECs knew who were not in these lists. This allowed them to skip the step where they had to contact SENCE/IDB for getting a randomized replacement and, contrarily to the design, directly contact applicants originally assigned to the control group. Unfortunately, we do not know how often this happened and how the OTECs decided which individuals they offered a slot for being reassigned. We carefully investigate whether original and realized assignment is random based on the applicants information in our data (see Section 4). The advantage of our study is that we observe the baseline characteristics and outcomes for nearly 90% of applicants. This allows us to carefully investigate potential selection in the re-assignment procedure.

We focus on the impact of being offered to participate in FOTRAB either during the original randomization or by being re-assigned. The numbers of applicants assigned or re-assigned to the treatment and the control group are shown in columns (3)-(4) of Table 2. Caused by the re-assignment procedure, the treatment group is larger than the control group. We define an assigned person as participant if she or he passes the initial 20% of the course. The numbers of participants, no-shows (the ones who never started or dropped out earlier) and non-participants are reported in columns (6)-(8). The compliance rate

defined as the share of participants among those assigned to a course is 67%.

3 Data collection and sample attrition

Applicants' information was collected in four points in time in face-to-face interviews: (i) basic data was collected at the time of application; (ii) a baseline questionnaire (including, for instance, socio-demographic and labor characteristics and skills) in 2013; (iii) a medium-term follow-up survey was conducted at about 1.5 years after applying to FOTRAB; and (iv) a long-term follow-up survey was collected in 2016, about 3 years after applying to FOTRAB. In addition, we use data from administrative records that track individuals until August 2017, i.e., 51 months after application to the program. This allows us to estimate short, medium- and long-term impacts of the FOTRAB program.

3.1 Basic data and baseline survey

Basic data was collected among all individuals who applied to the program. It consists of basic demographic characteristics at the time of applying and information for confirming eligibility to FOTRAB. The baseline survey was conducted shortly before training started and includes a large set of information of applicants such as many socio-demographic variables as well as information about household characteristics, education and self-reported labor market status. An enormous advantage of the evaluation of the FOTRAB program is the extensive measurement of cognitive and non-cognitive skills.

Cognitive skills at baseline were measured with tests of literacy, numeracy, spatial ability, and fluid intelligence. Literacy was measured by a "banked gap-filling" test. Individuals were asked to complete 12 gaps in a text by selecting words from a "bank" of words randomly ordered outside the text. The test for numeracy skills included a basic arithmetic problem. Spatial ability was measured by a paper folding task (Ekstrom et al., 1976). The image of an squared piece of paper later folded in two was shown to individuals. Then, two holes were punched in two places through the folded paper. Individuals were asked to choose the picture that correctly represent how the paper would look like after being unfolded. Spatial ability is the capacity to understand and remember the spatial relations among objects and has been found to affect labor outcomes (Aldén et al., 2017). To measure fluid intelligence,[7] a set of non-verbal multiple choice questions were used. For each question, individuals were asked to select, from a set of four to eight choices, the picture that completes a specific visual geometric design.

Non-cognitive skills were measured with three concepts. First, the Ten Items Personal-

[7] Fluid intelligence (skills for solving new problems) and crystallized intelligence (skills for using knowledge and experience) are two domains of general intelligence (Carroll, 1993).

ity Inventory Test (TIPI) (Gosling et al., 2003) to measure the Big-Five personality traits (openness to experiences, conscientiousness, extraversion, agreeableness and neuroticism) was implemented. Second, the 8-item scale to assess applicants' locus of control was used. This test reflects how much individuals believe that they have control over the outcome of events in their lives, as opposed to external forces beyond their control (Rotter, 1966). Third, a 10-item Rosenberg self-esteem scale was included in the baseline survey (Rosenberg, 1965). This scale measures individuals overall evaluation of their own worth. An overview of the collected information in the baseline survey is provided in Table 3.

Table 3: Collected information and attrition rates

	Baseline	Follow-up 1 (after 18 months)	Follow-up 2 (after 35 months)
Socio-demographics	x	x	x
Household characteristics	x	x	x
Education and labor market experience	x	x	x
Labor market situation			
Employment	x	x	x
Contributions to social security	x	x	x
Labor income	x	x	x
Hours worked	x	x	x
Self-employment	x	x	x
Cognitive skills			
Literacy	x	x	
Numeracy	x	x	
Spatial orientation	x	x	
Fluid intelligence	x	x	
Reflection (CRT)			x
Non-cognitive skills			
BIG Five	x		
Locus of control	x		
Self-esteem (Rosenberg Test)	x	x	
CPS		x	
GRIT		x	
Self-efficacy			x

Notes: The first and second follow-up surveys were collected, on average, 18 and 35 months after individuals applied to FOTRAB, respectively. The first follow-up was collected between 16 and 22 months after application, and the second follow-up between 32 and 39 months.

3.2 Follow-up surveys

The first follow-up survey was conducted between the end of 2014 and beginning of 2015 which was on average 18 months after applying to FOTRAB. The second follow-up took place in 2016, on average 35 months after applying to FOTRAB. As in the baseline survey, personal and household characteristics as well as self-reported labor market status was collected in both points in time. The labor market information included the employment

or self-employment status, whether contributions to the social security systems are paid, monthly labor income and the number of hours worked per month.

Again, tests for the measurement of cognitive and non-cognitive skills were implemented in both follow-up surveys. With respect to cognitive skills, the first follow-up included similar test as in the baseline survey. These test were not repeated in the second follow-up but replace by a Cognitive Reflection Test (CRT) (Frederick, 2005). The CRT consists of three-items in order to measure the cognitive ability to reflect on a question (i.e., using effort, motivation, concentration, maths skills) and resist the impulse of reporting the first intuitive response that comes to mind. CRT has been found to be positively associated with risk-taking and patience (Frederick, 2005).

For non-cognitive skills, the Rosenberg test was repeated in the first follow-up and complemented by a 14-item scale GRIT measure (Duckworth and Quinn, 2009) and a Personal Competencies Scale (CPS) test (Brea, 2010). The GRIT test is designed to reveal an individuals' consistency, perseverance and passion for achieving long-term results. The CPS tests seven personal and social skills which are leadership, conflict resolution, self-esteem, relationship with others, organization, empathy and communication. As for the cognitive skills, the second follow-up included only one test for non-cognitive skills which is a 10-item scale of general self-efficacy to assess individuals' belief of their own ability to achieve goals (Jerusalem and Schwarzer, 1992).

3.3 Attrition at baseline and follow-up surveys

The overall attrition rates are relatively low in this evaluation. First, we observe some attrition between application and the baseline survey. 89% of applicants were interviewed at the baseline. In Table 4, we perform regressions to investigate if the attrition rates differ between applicants who were originally assigned to participate in the course and those assigned to the control group. We regress an indicator of participation in the baseline survey on an indicator for original assignment to the treatment group, course-by-gender fixed effects and basic characteristics collected at the time of application (see column 1 and 2). The coefficients are close to zero and not significant, which suggests that attrition between application and the baseline survey seemed to be random.

Next, we focus on attrition in the two follow-up surveys (see columns 3-6). Because re-assignment took place after the baseline survey but before the follow-ups, we regress an indicator of participation in the first or second follow-up survey, respectively, on an indicator for being originally or re-assigned to the treatment group, course-by-gender fixed effects and characteristics provided at application and at the baseline. We find a positive coefficient of the likelihood to participate in the first follow-up of about 3 percentage points (pp). Although the effects are not large they could lead to a bias in the empirical

Table 4: Impact of random assignment on probability to participate in the surveys

	Baseline survey		Follow-up 1		Follow-up 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Response rate	0.887		0.868		0.863	
Original assignment	0.010	0.010	0.028	0.022	0.019	0.011
	(0.008)	(0.009)	(0.009)	(0.011)	(0.010)	(0.011)
Basic characteristics	No	Yes	No	Yes	No	Yes
Number of observations	5,193	5,193	5,193	5,193	5,193	5,193

Notes: The table reports the coefficient of an indicator of whether the person participate in the baseline or the follow-up surveys, respectively, on the originals or realized assignment dummy. Robust standard errors are reported in parentheses. All regressions control for gender and course fixed effects. In columns (2), (4), and (6), we control for the characteristics provided at the time of application which are age, poverty score (FPS) and labor market indicators in the year previous to application.

results. We find no relation between being offered to participate in the program and in the second follow-up survey.

4 Experimental balance and non-compliance

4.1 Balancing of applicants characteristics

Table 5: Experimental balancing using applicants' characteristics

	Original randomization		Realized assignment		Realized participation	
	Control group	Diff Treatment	Not assigned group	Diff Assigned	Realized treated	Diff No-shows
	(1)	(2)	(3)	(4)	(5)	(6)
Age in years	32.45	-0.050	32.82	-0.075	31.95	0.129
		(0.106)		(0.112)		(0.140)
Poverty index (FPS)	5,900	-120	5,845	-27	5,884	-221
		(80)		(92)		(109)
Formal empl. 2011/12	0.581	-0.012	0.601	-0.046	0.539	0.068
		(0.014)		(0.015)		(0.020)
Formal empl. 2012	0.465	-0.012	0.475	-0.034	0.431	0.056
		(0.013)		(0.016)		(0.019)
Months formal 2012	3.645	-0.159	3.783	-0.368	3.235	0.641
		(0.130)		(0.151)		(0.194)
Labor income in 2012	128,282	-2,986	134,112	-11,585	115,800	20,944
		(4,722)		(5,454)		(6,522)
Num. of observations	5,193		5,193		3,651	
Test joint significance	p-value = 0.696		p-value = 0.115		p-value = 0.002	

Note: We control for course-by-gender fixed effects in all regressions. Robust standard errors are reported in parentheses.

If the randomization of applicants was successfully implemented, the characteristics of those who are assigned to the control group and those who are offered to participate

in training would be not statistically different. In Table 5, we focus on the balancing of characteristics at application on which randomization was based on. Since randomization took place within training provider and course and was stratified by gender, we control for course-by-gender fixed effects in all regressions. In columns (1) and (2), we show the mean characteristics of applicants randomly assigned to the control group and their difference to those assigned to the treatment group based on original assignment. In columns (3) and (4), we focus on the balancing of the realized assignment groups. If the original assignment as well as the re-assignment was random, the characteristics of the original assigned groups and the realized assigned groups should be similar, respectively. Additionally, we investigate the non-compliance behavior of the assigned applicants in columns (5) and (6). If it was random, the characteristics of those who comply with the assignment and those who did no show up or dropped out early should be similar as well.

Applicants were 32 years old on average and had an average poverty score of 5,838 which indicates that they belong to the poorest quintile of the population. 57% of the men and women that applied to FOTRAB were formally employed at some point of time between 2011 and 2012, but only 46% in the year before application. They were employed for 3.6 months on average in 2012 and had a monthly labor income of 126,998 Chilean Pesos. A test for the joint significance of applicants characteristics to predict treatment assignment is not significant. Thus, we conclude that the assignment procedure was randomly implemented.

For realized assignment, we find some significant differences of the labor market status directly before application. Those assigned to participate were less likely to be employed and had lower income. All information available at application are jointly not significant and can not predict the assignment status. This indicates that applicants were randomly assigned to participate in FOTRAB. However, we observe non-randomness in the no-show behavior. Participants were negatively selected with respect to their labor market history. Those who did not show-up were significantly more often formally employed and had higher labor income in the past. The characteristics jointly predict the compliance behavior of those assigned to the treatment group.

4.2 Balancing of characteristics at the baseline

The baseline survey collects a variety of applicants characteristics including household information, educational attainment, and the results from cognitive and non-cognitive tests. In Table 6, we show the balancing and non-compliance behavior based on the baseline characteristics. Because of attrition between application and the baseline survey, the statistics are based on a slightly smaller sample (89% of original applicants) of 4,608 applicants.

Table 6: Experimental balancing using baseline characteristics

	Original randomization		Realized assignment		Realized participation	
	Control group	Diff Treatment	Not assigned group	Diff Assigned	Realized treated	Diff No-shows
	(1)	(2)	(3)	(4)	(5)	(6)
Household characteristics						
Married/partnership	0.415	0.003	0.443	-0.040	0.390	0.042
		(0.015)		(0.017)		(0.021)
Household size	4.190	0.049	4.171	0.081	4.254	-0.036
		(0.049)		(0.054)		(0.076)
Number of children	1.885	-0.011	1.890	-0.009	1.892	-0.048
		(0.032)		(0.032)		(0.045)
Head of household	0.454	-0.030	0.467	-0.027	0.421	0.035
		(0.014)		(0.016)		(0.022)
HH members 0-5 yrs	0.408	0.014	0.420	-0.000	0.402	0.038
		(0.013)		(0.016)		(0.020)
HH members > 65 yrs	0.136	0.011	0.136	0.007	0.140	0.011
		(0.010)		(0.011)		(0.015)
Joint significance	p-value = 0.746		p-value = 0.162		p-value = 0.121	
Educational attainment and cognitive skills						
Years of education	11.47	0.104	11.43	0.076	11.54	-0.026
		(0.049)		(0.061)		(0.081)
Literacy test	0.876	0.004	0.874	0.008	0.881	0.006
		(0.006)		(0.007)		(0.007)
Numeracy test	0.749	0.011	0.747	0.006	0.753	0.012
		(0.010)		(0.011)		(0.012)
Spatial orientation	0.755	0.022	0.756	0.016	0.769	-0.006
		(0.013)		(0.016)		(0.018)
Fluid intelligence	53.73	0.682	54.06	0.161	54.46	0.853
		(0.649)		(0.614)		(0.987)
Joint significance	p-value = 0.489		p-value = 0.596		p-value = 0.219	
Non-cognitive skills						
Locus of control	33.75	0.149	33.74	0.196	33.79	0.351
		(0.140)		(0.163)		(0.188)
Rosenberg score	33.92	0.146	33.93	0.177	34.00	0.205
		(0.153)		(0.176)		(0.205)
BIG 5: Extraversion	4.600	0.033	4.600	0.035	4.621	0.008
		(0.044)		(0.049)		(0.066)
BIG 5: Agreeableness	5.259	-0.060	5.242	-0.018	5.241	-0.069
		(0.040)		(0.051)		(0.054)
BIG 5: Conscientious	5.860	-0.075	5.888	-0.061	5.794	0.029
		(0.037)		(0.041)		(0.051)
BIG 5: Neuroticism	5.097	-0.033	5.051	0.052	5.089	0.023
		(0.042)		(0.046)		(0.059)
BIG 5: Openness	5.737	-0.019	5.747	-0.021	5.708	0.050
		(0.043)		(0.051)		(0.067)
Joint significance	p-value = 0.308		p-value = 0.298		p-value = 0.408	
Joint significance (all)	p-value = 0.515		p-value = 0.231		p-value = 0.167	
Num. of observations	4,806		4,806		3,288	

Note: We control for course-by-gender fixed effects in all regressions. Robust standard errors are reported in parentheses.

42% of the applicants are married or live in a partnership. The average household size amounts to 4 persons with on average two children living the household. About 44% are household heads, 42% live with children under the age of six, and 14% with elderly family members. The household characteristics are well-balanced between the original assigned and realized assigned treatment and control group. For realized assignment, we find that those assigned to FOTRAB are significant less often married and head of the household. With respect to the no-show behavior, we find that assigned individuals who did not show up or dropped out early are more often married or live in a partnership and are more likely to live with young children compared to participants. Joint tests of these variables to predict assignment and compliance behavior are not significant.

Applicants to the program accomplished on average 11.5 years of education, which is just below the 12 years needed to complete secondary education. The test scores for cognitive skills support the view of a middle- to high-skilled sample. The achievement on the cognitive skill tests for literacy is close to 90%, the one on the tests for numeracy and spatial orientation amounts to 75%. Fluid intelligence is tested on a scale from 0 to 100 with an average score of 54. The educational level and the cognitive skills are well-balanced for all groups.

On average applicants show a high internal locus of control and a high self-esteem score (34 points out of 40). The BIG 5 personality traits are measured on a scale from zero to seven. The scores range between 4.6 and 5.9 points on average. The scores on conscientiousness, openness and agreeableness are higher than the ones on neuroticism and extraversion. We find a significant difference in conscientiousness between the randomized groups with a lower score for those assigned to receive treatment. The tests of joint significance of non-cognitive skills as predictors for assignment or no-show behavior are again not significant.

Overall, the baseline characteristics are well-balanced. If we test all characteristics together none of tests is significant. Although the tests indicate that assignment is random, we admit that the differences between assigned individuals who participate and those who do not show up or dropped-out early indicates more household-related responsibilities of the latter group. We take this and other imbalances we find carefully into account in the empirical analysis by controlling for the labor market history, household characteristics, and non-cognitive skills in all regressions.

5 Empirical strategy

5.1 Parameters of interest and estimation

In this study, we estimate two parameters. First, the Intention-to-treat effect (ITT) identifies the effect of being randomly assigned to a slot in the FOTRAB program. This a highly relevant parameter from policy perspective because, as for most active labor market policy programs, individuals can be offered a participation in FOTRAB but can not be forced to participate. Thus, the ITT corresponds to the effect of making the program available to eligible individuals. Especially in cases when it is expected that non-compliance occurs, the ITT shows whether a program is effective although compliance can not be enforced or fully controlled. We estimate the ITT using a specification of the form

$$Y_{ij} = \beta Z_i + X_i\lambda + \tau_j + \epsilon_{ij}, \tag{1}$$

where Y_{ij} is the outcome of applicant i to course j, Z_i is an indicator for being assigned to participate in the program either during original randomization or during the re-assignment process, X_i is a vector of individual controls and τ_j are course-fixed effects. ϵ_{ij} is a random error term. Course by gender fixed-effects account for the fact that randomization was implemented at course-level and by gender. We use the information submitted at application, household characteristics, and non-cognitive skill variables as controls, because we document some imbalances in these sets of covariates. The coefficient β gives a weighted average of gender-specific effects across the 123 different courses that were selected for the evaluation. The coefficient and significance of the ITT are presented in the first row in all panels of the result tables.

Second, we estimate the local average treatment effects (LATE) for compliers, i.e., for the subset of participants that comply with the assignment. In our application, members of the control group had no access to the participate in the program unless there were re-assigned to empty slots. This provides us with a situation of one-sided non-compliance (e.g., Angrist and Pischke, 2009) which means that the effect for compliers is equal to the treatment effect on the treated (TOT) which is another policy-relevant parameter because it identifies the effect for those who participated in the program and for whom public resources were actually used. We estimate the LATE in the following specification by using the random assignment to the courses as instrument for participation,

$$Y_{ij} = \beta D_i + X_i\lambda + \tau_j + \epsilon_{ij}, \tag{2}$$

where the participation dummy D_i is instrumented with the indicator for assignment to

the FOTRAB program Z_i in the first-stage equation

$$D_i = \pi Z_i + X_i\delta + \tau_j + \eta_{ij}. \tag{3}$$

We estimate these equations by two-stage least squares using course-by-gender fixed effects in all equations. The coefficient β is a weighted average of course-by-gender specific LATE effects defined as the impact of participation in the FOTRAB program on individuals who are induced to participate by being offered a slot. This is equal to an up-scaling of the ITT by the first-stage effect of the instrument on the probability of participation, see equation (3). The coefficient and significance of the first-stage are presented in the second row, and the LATE estimates in third row in all panels of the result tables. We use robust standard errors in all specifications.

For the identification of the LATE, we impose the assumptions that are standard in the relevant literature (e.g., Imbens and Angrist, 1994; Angrist et al., 1996; Angrist and Pischke, 2009). We assume that the instrument is randomly assigned and has a significant effect on the probability to participate in the program. Both assumptions are satisfied by our experimental design. Furthermore, we assume individual-level monotonicity, which rules out the existence of defiers. Defiers are applicants that behave always different to what they are assigned to, i.e., they would participate in the program when assigned to the control group and would deny participation when assigned to the treatment group. It is very unlikely that we observe such rebellious behavior especially because individuals voluntarily applied to the FOTRAB program. Finally, we assume that the original assignment has no direct effect on the outcome of interest. This assumption is known as exclusion restriction in the literature. It implies that those who do not participate in training when assigned to it, e.g., the no-shows, would have the same outcome when assigned to the control group. This assumption likely holds in the FOTRAB setting, because not showing up means to not participate in the program, thus the no-shows are treated the same was as those assigned to the control group. However, one could still think about other violations of this assumption, for example, if those assigned to the control group are so heavily discouraged by this outcome that their skills or labor market outcomes are affected. It is unlikely that such effects exist in the case of FOTRAB because the economic constraints that eligible individuals face make the chance of discouragement or mood alteration affecting their labor behavior very unlikely. If such effects would exist they should have only a short-term impact and can be negligible in the longer run.

5.2 Outcomes

Given the programs objectives, the main outcomes of interest in the evaluation are the labor market outcomes of applicants. As discussed above, in both follow-up surveys, individ-

uals were asked to report their employment status (i.e., paid employee or self-employed), their income and whether social security contributions are paid. Paid-employment includes salaried workers in the public or private sector, and self-employment includes employers and workers on their own account. In Chile, paid workers and employers must contribute to the social security system, however, at the time of the evaluation, contributions to the social security was optional for own-account workers. Thus, we cannot interpret contribution to social security as a direct measure of job formality. We complement the analysis with employment records from administrative data. We use data from the unemployment insurance (UI) system, which contains employment and earning histories of formal workers in the country.[8] This data is administered by the Unemployment Fund Administrator and contains records from all formal dependent workers until August 2017 which allows us to follow the applicants for the program over a period of 51 months after application.

Moreover, as shown in Table 3.1 several tests for cognitive and non-cognitive skills in the first follow-up and one test each in the second follow-up survey were collected. For the first follow-up, we organize the test scores into summary indices of cognitive and non-cognitive skills. Therefore, we standardize each individual skill outcome with the control group mean and the standard deviation. Then, we summarize the standardized test scores to one cognitive skill index and one non-cognitive skill index in the first follow up (similar indices are used in the early childhood education literature, e.g., Walters, 2015; Deming, 2009). The test scores in the second follow-up are standardized in the same way, however, since only one cognitive and non-cognitive test was implemented the skill outcomes in the second and first follow-up survey can not be compared over time.

6 The impact of FOTRAB on labor outcomes

We present the program impact on the probability of having a paid-employment and being self-employed, and labor income in the first and second follow-ups in Table 7. Being assigned to and participate in FOTRAB has no effect on the likelihood to work in paid-employment or to work as self-employed in both follow-ups. We find a significant increase of labor income, at least in the first follow-up, that vanishes over time. The raise of monthly income amounts to 9,000 Chilean Pesos for those assigned and to 12,000 Chilean Pesos for those who participate.

To complement the survey results, we show the effect of the program based on ad-

[8] UI is an individual savings account for each formal worker. Both the worker and the employer contribute to this fund, although UI is supplemented by the Solidarity Fund, which is financed by public and private (employer) contributions. The Unemployment Fund Administrator of Chile is the private manager of the mandatory UI.

Table 7: Effects on labor outcomes

	Paid-employment		Self-employment		Labor income	
	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2
(1) ITT	-0.001	-0.013	0.013	0.015	9.00	-0.42
	(0.016)	(0.017)	(0.014)	(0.017)	(5.30)	(5.51)
(2) First stage	0.743	0.745	0.743	0.745	0.742	0.744
	(0.017)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)
(3) LATE=TOT	-0.001	-0.018	0.018	0.020	12.13	-0.56
	(0.022)	(0.023)	(0.019)	(0.018)	(7.15)	(7.60)
(4) Control means	0.528	0.548	0.196	0.185	172.12	207.21
Num. of obs.	4,101	4,055	4,095	4,053	3,951	3,906

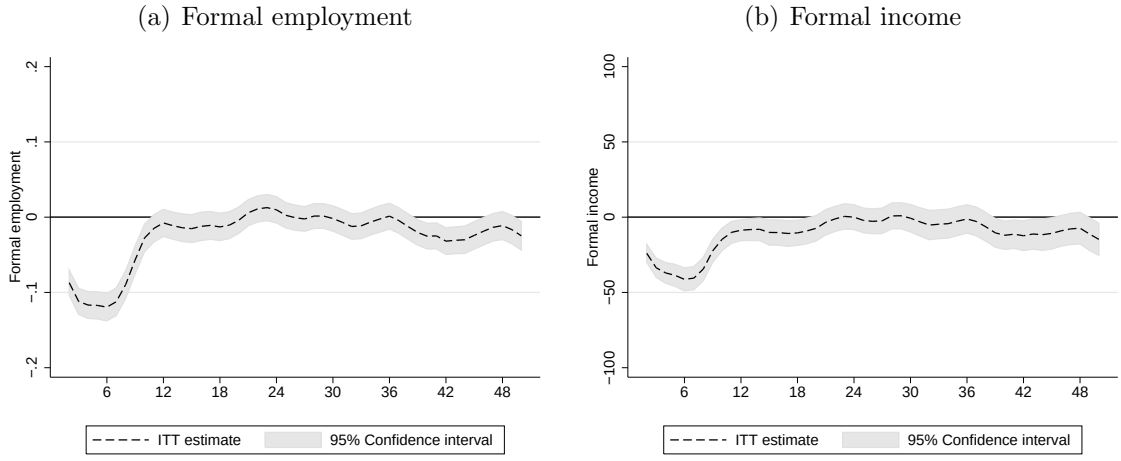
Note: This table reports the impact of being assigned to and participate in FOTRAB on different forms of employment and labor income measured in 1,000 CLP. FU 1 and FU 2 are the abbreviations for first and the second follow-up, respectively. The TOT estimates are obtained by using the random assignment as instrument for participation. Robust standard errors are reported in parentheses. All regressions control for course-by-gender fixed effects, basic characteristics provided at application as well as household characteristics and non-cognitive skills measured at the baseline. All missing covariates are set to zero, and dummies for missing values are included.

ministrative records of applicants in Figure 1. The results imply that being assigned to the program leads to significant negative lock-in effects for both outcomes, formal employment and income. These negative effects can be explained by a lower search intensity of participants during training. Around a year after applying to FOTRAB, the effects increases sharply but never turn positive over the entire observational period. Instead, it seems that they evolve to a negative effect on formal employment and formal income in the longer run. As robustness check, we implement a pretreatment outcome evaluation for the three years before training started. This allows us to investigate if there are any effects shortly before training, the so-called Ashenfelter's dip (e.g., Ashenfelter and Card, 1985) or other selectivity based on pre-treatment labor outcomes. As can be seen from the results in Figure B.1 in the Appendix, the effects are zero over almost the whole pre-training observation period.

Paying social security contributions is most comparable to the variable of having a formal employment from administrative data presented in Figure 1 which refer to those contributing to UI only. In Table C.1 in Appendix C, we provide results on the probability of paying social security contributions and other outcomes reported in the surveys (the hours worked, the probability to be unemployed, and the probability to be inactive, thus not participating in the labor market). The point estimates of the program impact on the probability of paying social security are similar to the ones from administrative data after 18 and 36 months. There are no effects on other outcomes.

Then we analyze whether the effect of FOTRAB varies across the three different course types presented in Section 2. We start with the most basic course type (CT courses) which is classroom training only. The impacts of this course type are presented in Panel

Figure 1: Impact of assignment to FOTRAB (using administrative records)



Note: These figures show the ITT effects of being assigned to the program on the probability to be formally employed and formal income on a monthly basis. The effects are presented as rolling averages using a three month window. The gray shaded area indicates the 95% confidence interval calculated based on robust standard errors.

A of Table 8. We find positive coefficients on the ITT and the TOT on the probability of having a paid-employment in both follow-up surveys that never become significantly different from zero, but a significantly positive ITT on labor income in the longer run. Monthly income increases by 26,690 Chilean Pesos. The effect for participants is positive as well, but estimated with much noise. The effects on the other outcomes reported in Table C.2 are not significant. We find the same pattern and a substantial variance due to the relatively low number of observations in the results obtained from administrative data in Figure 2. There is no lock-in period which can be explained by the fact that CT courses are designed for people who wanted to combine working and training, thus, many of them are employed during participation. The effects remain close to zero and insignificant over the whole period. Thus, CT courses neither increased nor decreases the employment outcomes of participants.

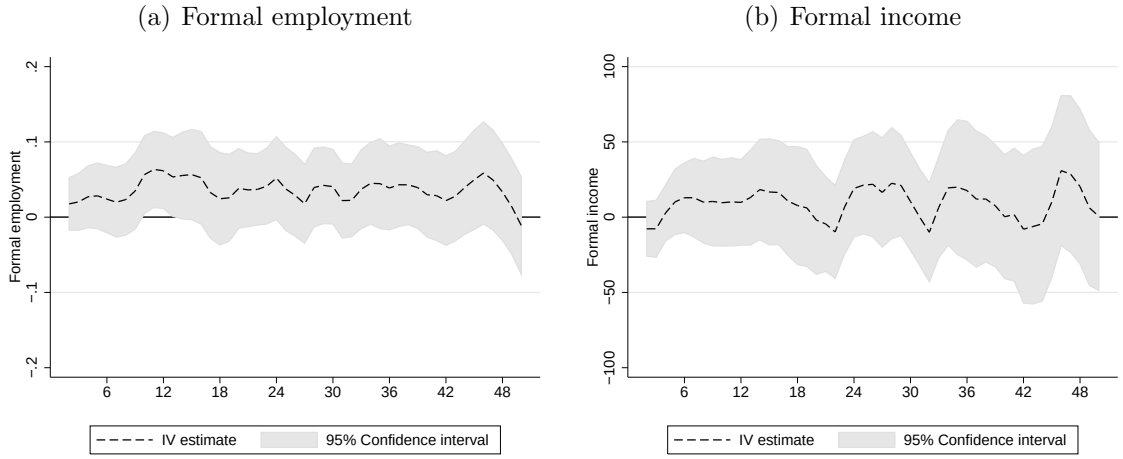
Next, we investigate the effects of being assigned and participate in CJT courses, which includes a job training period in addition to the classroom training (see Panel B of Table 8). We find a significantly negative ITT and TOT on paid-employment in the longer run about 36 month after application. Participation in CJT courses reduces the probability to be a paid-worker by 6 pp. We find no effect on the probability to be self-employed. The negative pattern in the long-run is also reflected in significant lower labor income in the second follow-up for those assigned to training which decreases significantly by 11,400 Chilean pesos. Also, the likelihood to pay social security contributions significantly decreases (see Table C.2 in the appendix). Moreover, we find the same pattern using the administrative data as can be seen from Figure 3. After a pronounced negative lock-in

Table 8: Effects on labor outcomes by course type

	Paid-employment		Self-employment		Labor income	
	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2
Panel A: CT courses						
(1) ITT	0.018	0.044	0.008	-0.004	11.40	26.69
	(0.062)	(0.057)	(0.041)	(0.051)	(13.41)	(11.88)
(2) First stage	0.798	0.812	0.798	0.812	0.801	0.811
	(0.025)	(0.025)	(0.025)	(0.025)	(0.026)	(0.026)
(3) LATE=TOT	0.023	0.055	0.010	-0.005	14.23	32.93
	(0.057)	(0.058)	(0.045)	(0.048)	(20.56)	(21.61)
(4) Control means	0.667	0.639	0.177	0.203	275.42	299.70
Num. of obs.	478	484	478	484	461	461
Panel B: CJT courses						
(1) ITT	0.014	-0.041	-0.007	0.014	7.44	-11.37
	(0.021)	(0.021)	(0.017)	(0.015)	(6.96)	(6.91)
(2) First stage	0.699	0.698	0.698	0.698	0.698	0.699
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
(3) LATE=TOT	0.021	-0.058	-0.011	0.020	10.66	-16.28
	(0.031)	(0.032)	(0.024)	(0.023)	(9.62)	(10.82)
(4) Control means	0.539	0.579	0.158	0.139	163.60	206.05
Num. of obs.	2,727	2,680	2,722	2,678	2,630	2,593
Panel C: CTA courses						
(1) ITT	-0.043	0.030	0.064	0.022	13.62	15.74
	(0.032)	(0.033)	(0.033)	(0.034)	(11.40)	(11.79)
(2) First stage	0.814	0.817	0.814	0.817	0.811	0.813
	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)
(3) LATE=TOT	-0.052	0.037	0.079	0.027	16.79	19.36
	(0.040)	(0.041)	(0.039)	(0.037)	(12.76)	(12.63)
(4) Control means	0.438	0.433	0.290	0.279	141.56	163.67
Num. of obs.	896	891	895	891	860	852

Note: This table reports the impact of being assigned to and participate in one of the course types on different forms of employment and labor income measured in 1,000 Chilean Pesos. FU 1 and FU 2 are the abbreviations for first and the second follow-up, respectively. The TOT estimates are obtained by using the random assignment as instrument for participation. Robust standard errors are reported in parentheses. All regressions control for course-by-gender fixed effects, basic characteristics provided at application as well as household characteristics and non-cognitive skills measured at the baseline. All missing covariates are set to zero, and dummies for missing values are included.

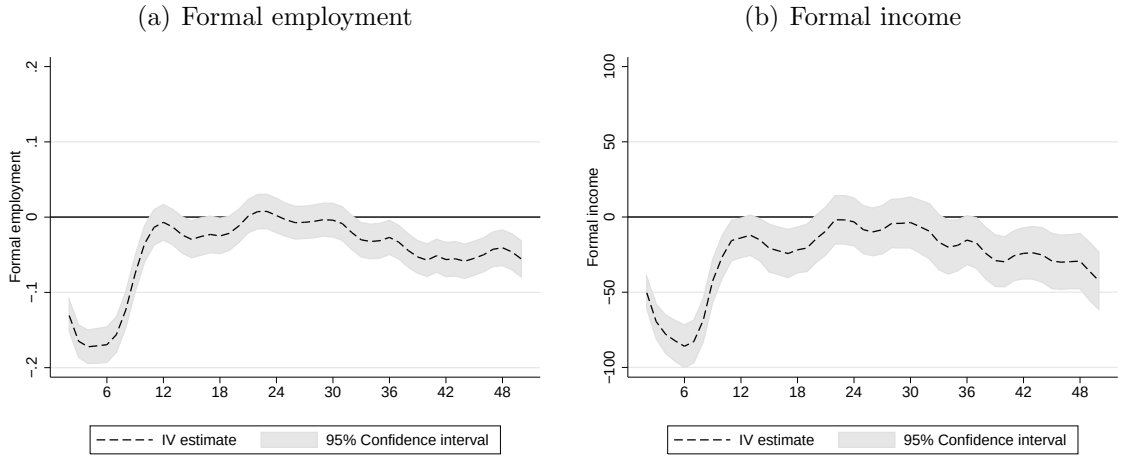
Figure 2: Impact of assignment to CT courses (using administrative records)



Note: These figures show the ITT effects of participation in CT courses on the probability to be formally employed and formal income on a monthly basis. The effects are presented as rolling averages using a three month window. The gray shaded area indicates the 95% confidence interval calculated based on robust standard errors.

period during which the probability to be formally employed decreases by about 20 pp and formal income by almost 100,000 Chilean Pesos compared to the control group, the effects increase sharply after participation. However, in the longer run about 36 months after application the effects start to decrease and become significantly negative.

Figure 3: Impact of assignment to CJT courses (using administrative records)

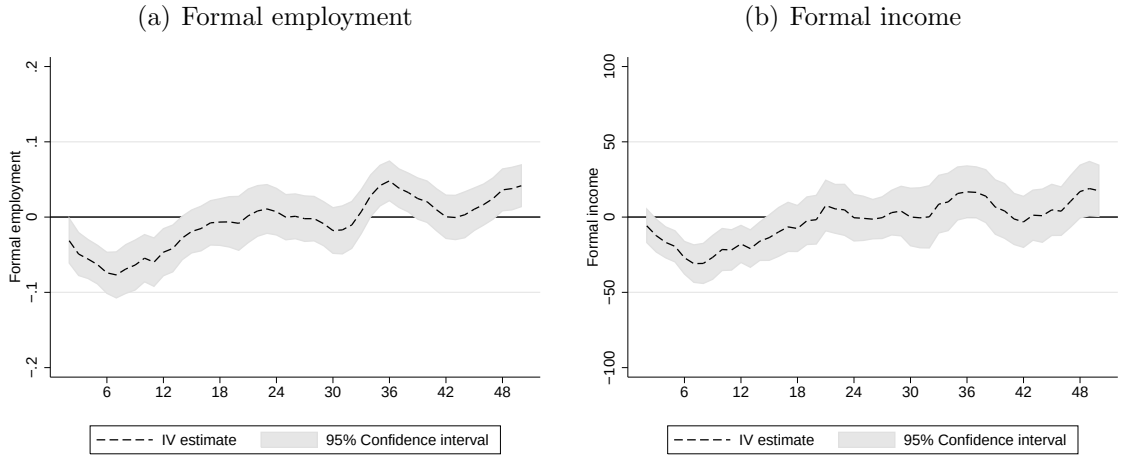


Note: These figures show the ITT effects of participation in CJT courses on the probability to be formally employed and formal income on a monthly basis. The effects are presented as rolling averages using a three month window. The gray shaded area indicates the 95% confidence interval calculated based on robust standard errors.

CTA courses consist of classroom training complemented with technical assistance. The target group of these courses are applicants that seek to be self-employed. The program effects for these courses are presented in Panel C of Table 8. We find negative but

not significant point estimates on the probability to be a paid-employee in the first follow-up. At the same time, the probability for self-employment increases significantly by almost 8 pp. However, this positive effect vanishes in the long-term. Overall, these results indicate that participants may have tried to substitute paid-employment by self-employment at least in the short-run. In the second follow-up, the effect on self-employment is still positive but no longer significant. Labor income is not affected in both follow-ups. From the results using administrative records on formal employment and income in Figure 4, we observe significant negative lock-in effects. The effects on both outcomes increase over time and become significantly positive after 48 months.

Figure 4: Impact of assignment to CTA courses (using administrative records)



Note: These figures show the ITT effects of participation in CTA courses on the probability to be formally employed and formal income on a monthly basis. The effects are presented as rolling averages using a three month window. The gray shaded area indicates the 95% confidence interval calculated based on robust standard errors.

7 Heterogeneous effects and potential mechanisms

7.1 Subgroup analysis by gender and age groups

We report heterogeneous effects by gender and age-group in Table 9. In Panel A, we estimate the effect of being offered training for the subgroup of women and the interaction indicating if training is more or less effective for males compared to women. We find negative effects on paid-employment for women in the long run. Men who were offered training are more likely to have a paid-employment and also earn significantly more than women in the long-term. These results contrast with the average results in LAC (Escudero et al., 2019) and worldwide (Card et al., 2018) but are aligned with the findings of Alzúa et al. (2016) in Argentina.

In Panel B, we investigate heterogeneity by age-group. Although not statistically significant, the results suggest that the program is less effective for youth (i.e., younger than 30) than for adults. The coefficients on the interaction term have a negative sign for almost all outcomes. This finding is in line with the evidence of training programs for OECD countries (Kluve et al., 2019). However, it contrasts with the evidence for LAC and with the one coming from non-experimental evaluations of the predecessor programs that originated FOTRAB (Jara, 2001; Centro Microdatos, 2006, 2008; Ministerio de Trabajo y Previsión Social, 2011).

Table 9: Effects on labor outcomes by gender and age-group

	Paid-employment		Self-employment		Labor income	
	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2
Panel A: Effects by gender						
ITT female	0.005	-0.036	0.013	0.027	4.30	-7.02
	(0.021)	(0.020)	(0.019)	(0.017)	(5.63)	(6.36)
ITT × male	-0.013	0.067	-0.001	-0.034	14.00	23.51
	(0.035)	(0.037)	(0.029)	(0.030)	(11.83)	(10.88)
Panel B: Effects by age-group						
ITT adult	0.014	-0.017	0.022	0.030	16.71	3.80
	(0.022)	(0.024)	(0.020)	(0.019)	(7.78)	(7.52)
ITT × young	-0.034	0.008	-0.019	-0.035	-17.35	-9.23
	(0.031)	(0.036)	(0.025)	(0.026)	(11.50)	(11.49)

Note: This table reports the effects of being assigned to FOTRAB for females and the difference between men and women on different forms of employment and labor income measured in 1,000 CLP. Additionally, we split the sample into the three course types. FU 1 and FU 2 are the abbreviations for first and the second follow-up, respectively. Robust standard errors are reported in parentheses. We control for course fixed effects, basic characteristics provided at application as well as household characteristics and non-cognitive skills measured at the baseline. All missing covariates are set to zero, and dummies for missing values are included.

7.2 Subgroup analysis by OTEC quality

OTECs are non-profitable organizations devoted to the training of vulnerable people. To provide publicly-sponsored courses, they have to meet requirements with respect to infrastructure and staff. However, we observe substantial variation of the implementation of these requirements which allos us to analyze whether the effectiveness of the program varies by the quality of the OTEC.

For the quality measure, we use course level information submitted by the OTECs to SENCE when applying for funding and detailed provider characteristics obtained from survey data at OTEC-level. In 2013, an OTEC questionnaire was administered to the 12 OTECs where the 123 courses that are part of the evaluation took place. We provide summary statistics of the variables we used to construct a quality index in Table 10. The

characteristics on the course level are shown in Panel A. We consider course capacity, teachers teaching and general work experience and the number of teaching hours devoted to technical skills and transversal ability. On average there were 21.42 persons attending a course. Teachers have an average teaching and working experience of 2.2 years and 2.6 years, respectively. The number of hours for teaching technical skills amounts to 387 hours, and for transversal abilities 95 hours.

The characteristics in Panel B are measured on the level of training providers. More than half of the provides report, that they invest in teachers training. The average share of employed teachers with university degree is 44%, and on average all OTEC's teachers have a teaching experience of 5 years. About 20% of providers apply psychological tests in the recruitment interviews, and almost half of them test students before course start.

We combine the information of these components to one quality indicator. The reason behind is that no component is a perfect proxy for quality but each single component provide an error-prone predictor of it. We use a first principal component analysis to obtain one quality index. We use this quality index and split the 123 courses in two quality groups. Courses with a quality index value below average within course type are defined as low quality courses. Courses with an index above average within course type are defined as high quality courses. In Table 10, we check the course level and provider characteristics by quality level and find significant differences between them. For example high quality courses have a lower capacity, teachers show longer teaching and work experience, they invest more often in teacher training and use more often testing procedures. The values for low and high quality courses are very plausible and support the view that the constructed index separate low and high quality courses from each other.

In Table 11, we investigate heterogeneous program effects by course quality. Since the applicants were not allocated randomly across courses with different quality the reported results presents mere correlations. We find significant higher effects for applicants assigned to a high quality course on the probability to be self-employed in the first follow-up and on labor income in the longer run. These results are aligned to the ones of Galdo and Chong (2012), which are the only other paper looking at the differential effect of training programs and provider quality.

7.3 Program impact on skill outcomes

In Table 12, we investigate if FOTRAB improves the skill development of applicants and participants. The program has a significant positive effect on the cognitive skill index in the first follow-up. This is an important finding because basic literacy and numeracy skills are highly demanded by employers and are a requirement for learning other skills and performing better in the workplace. We cannot investigate whether this is a long-term

Table 10: Variables used to construct quality measure

	All	Quality index		
		Low	High	Diff
Panel A: Course-level information				
Course capacity	21.42	22.00	20.97	-1.03
	(0.45)	(1.20)	(3.11)	(0.45)
Years of teaching experience of teachers	2.16	2.00	2.29	0.29
	(0.07)	(0.00)	(0.55)	(0.07)
Years of working experience of teachers	2.63	2.09	3.06	0.97
	(0.11)	(0.29)	(0.75)	(0.11)
Number of teaching hours in technical abilities	387.12	411.74	367.86	-43.89
	(25.51)	(202.91)	(54.66)	(25.51)
Number of teaching hours in transversal abilities	95.06	80.07	106.78	26.71
	(6.96)	(28.85)	(44.27)	(6.96)
Panel B: Provider-level information				
Investments in teacher training	0.54	0.15	0.84	0.69
	(0.07)	(0.36)	(0.37)	(0.07)
Share of teachers with university degree	0.44	0.06	0.74	0.68
	(0.07)	(0.23)	(0.44)	(0.07)
Years of experience of teachers	4.98	4.22	5.57	1.34
	(0.37)	(2.38)	(1.70)	(0.37)
Use of psychological tests in teacher recruitment	0.20	0.06	0.32	0.26
	(0.07)	(0.23)	(0.47)	(0.07)
Test students skills before course start	0.46	0.09	0.74	0.65
	(0.07)	(0.29)	(0.44)	(0.07)

Note: The standard deviation of the means are reported in parentheses. The difference between the characteristics of low and high quality training course are calculated from a regression of the characteristics on the indicator for high quality. The standard error of the difference is reported in parentheses.

Table 11: Effects on labor outcomes by provider quality

	Paid-employment		Self-employment		Labor income	
	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2
ITT lower quality	0.019	-0.030	-0.022	0.017	3.628	-11.827
	(0.023)	(0.023)	(0.019)	(0.017)	(7.117)	(8.005)
ITT × higher quality	-0.039	0.033	0.065	-0.005	8.519	22.379
	(0.032)	(0.033)	(0.027)	(0.026)	(10.541)	(11.749)

Note: This table reports the effects of being assigned to FOTRAB for applicants in lower quality OTECs and the difference between to higher quality OTECs on different forms of employment and labor income measured in 1,000 Chilean Pesos. FU 1 and FU 2 are the abbreviations for first and the second follow-up, respectively. Robust standard errors are reported in parentheses. We control for course fixed effects, basic characteristics provided at application as well as household characteristics and non-cognitive skills measured at the baseline. All missing covariates are set to zero, and dummies for missing values are included.

effect because another skill measure was used in the second follow-up.

For the CRT (cognitive reflection test) measured three years after application and the other non-cognitive skill measures in the follow-up surveys we find no effects. Although FOTRAB aimed at improving workers' skills, the program seems to be ineffective in helping applicant to the develop skills, at least the ones measured in the surveys. The importance of skills for achieving labor outcomes (Heckman et al., 2006), together with these results, can help in the interpretation of the low effectiveness of FOTRAB.

Table 12: Effects on skill outcomes

	Cognitive skills		Non-cognitive skills	
	Index FU 1	CRT FU 2	Index FU 1	Self-efficacy FU 2
(1) ITT	0.038	0.015	-0.031	-0.002
	(0.021)	(0.033)	(0.026)	(0.028)
(2) First stage	0.743	0.746	0.743	0.743
	(0.009)	(0.009)	(0.009)	(0.009)
(3) LATE=TOT	0.051	0.021	-0.042	-0.003
	(0.029)	(0.047)	(0.035)	(0.036)
(4) Control means	-0.012	-0.039	0.026	0.007
Num. of obs.	4,107	4,045	4,107	4,590

Note: This table reports the impact of being assigned to and participate in FOTRAB on different skill outcomes. The first follow-up indices correspond to the average of the z-scores (the score on each skill test was standardized taking into account the control group mean and the standard deviation) of the cognitive (literacy, numeracy, spatial orientation, and fluid intelligence) and non-cognitive (self-esteem, CPS, and grit) tests included. The test scores (CRT and self-efficacy) in the second follow-up are standardized in the same way. FU 1 and FU 2 are the abbreviations for first and the second follow-up, respectively. CRT stands for cognitive reflection test. The TOT estimates are obtained by using the random assignment as instrument for participation. Robust standard errors are reported in parentheses. All regressions control for course-by-gender fixed effects, basic characteristics provided at application as well as household characteristics and non-cognitive skills measured at the baseline. All missing covariates are set to zero, and dummies for missing values are included.

8 Discussion and conclusion

The experimental implementation of one cohort of the FOTRAB program offers the first opportunity to evaluate the causal effect of a large public-funded job training program in Chile. The experimental design, together with access to survey and administrative data allowed us to measure the effects of FOTRAB on labor market and skills outcomes more than 4 years after application.

FOTRAB aimed at improving the labor market outcomes and skills of vulnerable individuals in the country by offering a comprehensive job training program. Our results, however, show that the program has limited effectiveness. Overall, we find that FOTRAB

only improves labor income in the short-term. Also, we find evidence of heterogeneous effects. First, while course combining classroom training and technical assistance improve the probability of being self-employed, those combining classroom and on-the-job-training reduce the probability of employment and labor income. Second, we find that FOTRAB is more effective for male than females. Third, we find evidence of a positive association between labor outcomes and the quality of the trainer provider.

The programs reduced effectiveness could be related to many factors. It might be explained by the almost null effects of FOTRAB on skills development. Although the program devoted a considerable proportion of the training duration to skills development, we find almost no evidence that FOTRAB improve the cognitive and non-cognitive skills measured in the evaluation.

Although FOTRAB operated as such until 2018, SENCE continues offering programs with similar designs in Chile. Considering the recent evidence of what works for training programs and the design of FOTRAB we could also hypothesized about other features that could have improved its effectiveness. First, the program could have benefited from a better connection with the labor demand that would contribute to the design of the curricula and program (Alzúa et al., 2020). A better definition and supervision of the quality standards of training providers is also key (Galdo and Chong, 2012). In the case of FOTRAB we could not find, for instance, any evidence of mechanisms to secure that the non-cognitive skills component was taught by accredited professionals. Having expanded the targeted population of FOTRAB to individuals from different backgrounds could have also affected the effectiveness found by the non-experimental evaluations of the precedent programs (Duflo et al., 2011). Finally, as with other SENCE programs (Novella et al., 2017), the program might have benefit from offering vocational orientation and labor market information (e.g., returns to certain training) aiming at improving the match between applicants and the supply of training courses.

References

Aedo, C. and M. Pizarro (2004). Rentabilidad económica del programa de capacitación laboral de jóvenes chile joven. *INACAP and Mideplan. Santiago de Chile. Processed*.

Aldén, L., M. Hammarstedt, and E. Neuman (2017). All about balance? a test of the jack-of-all-trades theory using military enlistment data. *Labour Economics 49*, 1 – 13.

Alzúa, M. L., S. Batbekh, A. Batchuluun, B. Dalkhjav, and J. Galdo (2020, 06). Demand-Driven Youth Training Programs: Experimental Evidence from Mongolia. *The World Bank Economic Review*.

Alzúa, M. L., G. Cruces, and C. Lopez (2016). Long-run effects of youth training programs: Experimental evidence from argentina. *Economic Inquiry 54*(4), 18391859.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91(434)*, 444–472.

Angrist, J. D. and J. S. Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press.

Ashenfelter, O. and D. Card (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics 67*(4), 648660.

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*(2), 131143.

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*(3), 188220.

Brea, M. (2010). Interpretación de las escalas cps, rosenberg y grit y propuestas de revisión para mejorar la confiabilidad. Technical report.

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*(2), 267300.

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Centro Microdatos (2006). Evaluación de impacto del programa de formación en oficios para jóvenes de escasos recursos, informe final. Technical report, Centro Microdatos.

Centro Microdatos (2008). Evaluación de impacto del programa de formación en oficios para jóvenes de escasos recursos, ejecución 2005. Technical report, Centro Microdatos.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics 1*(3), 111–134.

Duckworth, A. L. and P. D. Quinn (2009). Development and validation of the short grit scale (grit–s). *Journal of Personality Assessment 91*(2), 166–174.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review 101*(5), 1739–74.

Ekstrom, R. B., D. Dermen, and H. H. Harman (1976). *Manual for kit of factor-referenced cognitive tests*, Volume 102. Educational testing service Princeton, NJ.

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

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 25–42.

Galdo, J. and A. Chong (2012). Does the quality of public-sponsored training programs matter? evidence from bidding processes data. *Labour Economics 19*(6), 970–986.

Gosling, S. D., P. J. Rentfrow, and W. B. Swann Jr (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality 37*(6), 504–528.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

Ibarrarán, P., J. Kluve, L. Ripani, and D. Rosas Shady (2018). Experimental evidence on the long-term effects of a youth training program. *ILR Review*, 0019793918768260.

Ibarrarán, P. and D. Rosas Shady (2009). Evaluating the impact of job training programmes in latin america: evidence from idb funded operations. *Journal of Development Effectiveness 1*(2), 195–216.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62(2)*, 446–475.

Jara, O. (2001). Informe ejecutivo de estudio. seguimiento y evaluación programa de formación en oficios para jóvenes de escasos recursos. Technical report, Geo Consultores.

Jerusalem, M. and R. Schwarzer (1992). Self-efficacy as a resource factor in stress appraisal processes. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action*, pp. 195–213. New York: Taylor and Francis Group.

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

Kugler, A., M. Kugler, J. Saavedra, and L. O. H. Prada (2015). Long-term direct and spillover effects of job training: Experimental evidence from colombia. Technical report, National Bureau of Economic Research.

Ministerio de Trabajo y Previsión Social (2011). Informe final. comisión revisora del sistema de capacitación e intermediación laboral. Technical report, Ministerio de Trabajo y Previsión Social.

Novella, R., G. Rucci, C. Vazquez, and D. S. Kaplan (2017). Training vouchers and labour market outcomes in chile. *LABOUR*.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied 80*(1), 1.

Walters, C. R. (2015). Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start. *American Economic Journal: Applied Economics 7(4)*, 76–102.

A Summary of course characteristics

Table A.1: Distribution of courses by region and OTECs

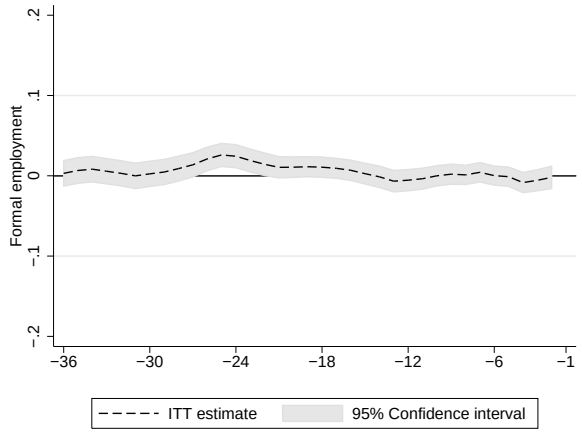
	All	CT	CJT	CTA
Region 1: Valparaíso	22	0	22	0
OTEC 1	13	0	13	0
OTEC 2	9	0	9	0
Region 2: Biobío	44	5	31	8
OTEC 1	20	4	16	0
OTEC 3	4	0	4	0
OTEC 4	3	0	3	0
OTEC 5	4	0	4	0
OTEC 6	9	1	9	8
OTEC 7	4	0	4	0
Region 3: Metropolitana	57	11	29	17
OTEC 1	8	3	5	0
OTEC 3	6	0	3	3
OTEC 6	11	0	0	11
OTEC 8	3	0	3	0
OTEC 9	16	8	8	0
OTEC 10	5	0	2	3
OTEC 11	4	0	4	0
OTEC 12	4	0	4	0
All	123	16	82	25

Note: CT is classroom training, CJT is classroom plus job training, and CTA is classroom plus technical assistance.

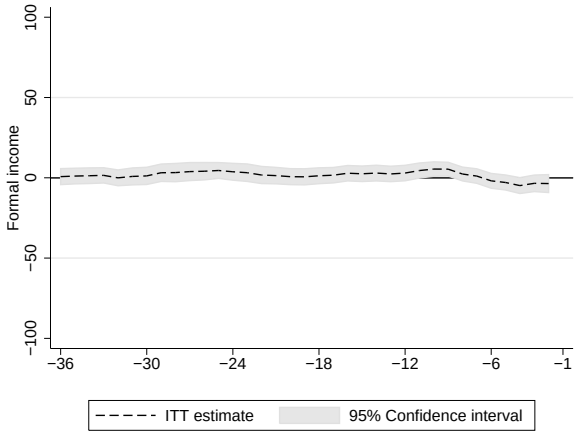
B Impacts on pre-treatment outcomes

Figure B.1: Pre-treatment analysis (using administrative records)

(a) Formal employment



(b) Formal income



Note: These figures show the effects of being assigned to the program on the probability to be formally employed and formal income on a monthly basis for the three years prior to application. The effects are presented as rolling averages using a three month window. The gray shaded area indicates the 95% confidence interval calculated based on robust standard errors.

C Impacts on more outcomes

Table C.1: Impacts on more labor market outcomes

	Pay social security		Hours worked		Unemployment		Inactive	
	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2
(1) ITT	-0.016	-0.015	-0.18	-0.06	-0.013	-0.003	0.000	0.001
	(0.018)	(0.018)	(0.75)	(0.73)	(0.009)	(0.010)	(0.015)	(0.013)
(2) First stage	0.739	0.745	0.743	0.744	0.743	0.745	0.743	0.745
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
(3) LATE = TOT	-0.022	-0.021	-0.24	-0.08	-0.018	-0.004	0.000	0.002
	(0.023)	(0.023)	(0.98)	(0.99)	(0.014)	(0.013)	(0.018)	(0.018)
(4) Control means	0.466	0.473	27.21	28.41	0.088	0.084	0.187	0.182
Num. of obs.	3,870	4,034	4,082	4,033	4,101	4,055	4,101	4,055

Note: This table reports the effect of being assigned to and participate in FOTRAB on the probability to pay social security, hours worked, the probability to be unemployed, and the probability to being inactive. Labor income is measured in 1,000 Chilean Pesos. FU 1 and FU 2 are the abbreviations for first and the second follow-up, respectively. The IV estimates are obtained by using the random assignment as instrument for participation in a course. Robust standard errors are reported in parentheses. All regressions control for course-by-gender fixed effects, basic characteristics provided at application as well as household characteristics and non-cognitive skills measured at the baseline. All missing covariates are set to zero, and dummies for missing values are included.

Table C.2: Impacts on more labor market outcomes by course type

	Pay social security		Hours worked		Unemployment		Inactive	
	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2	FU 1	FU 2
Panel A: CT courses								
(1) ITT	0.019	-0.012	0.40	0.77	-0.030	-0.033	0.004	-0.007
	(0.066)	(0.056)	(1.47)	(1.60)	(0.030)	(0.023)	(0.033)	(0.026)
(2) First stage	0.794	0.812	0.798	0.809	0.798	0.812	0.798	0.812
	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
(3) LATE = TOT	0.023	-0.014	0.50	0.95	-0.038	-0.041	0.005	-0.009
	(0.058)	(0.057)	(2.26)	(2.26)	(0.031)	(0.030)	(0.035)	(0.031)
(4) Control means	0.643	0.633	36.50	37.37	0.068	0.082	0.088	0.076
Num. of obs.	463	484	478	479	478	484	478	484
Panel B: CJT courses								
(1) ITT	-0.010	-0.038	0.93	-0.83	-0.018	0.013	0.009	0.013
	(0.023)	(0.024)	(1.00)	(0.93)	(0.013)	(0.013)	(0.020)	(0.017)
(2) First stage	0.694	0.699	0.698	0.696	0.699	0.698	0.699	0.698
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
(3) LATE = TOT	-0.015	-0.055	1.33	-1.19	-0.026	0.018	0.013	0.019
	(0.032)	(0.032)	(1.36)	(1.36)	(0.020)	(0.019)	(0.025)	(0.026)
(4) Control means	0.493	0.496	25.85	28.26	0.106	0.082	0.197	0.199
Num. of obs.	2,572	2,664	2,709	2,665	2,727	2,680	2,727	2,680
Panel C: CTA courses								
(1) ITT	-0.010	-0.038	0.93	-0.83	-0.018	0.013	0.009	0.013
	(0.023)	(0.024)	(1.00)	(0.93)	(0.013)	(0.013)	(0.020)	(0.017)
(2) First stage	0.694	0.699	0.698	0.696	0.699	0.698	0.699	0.698
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
(3) LATE = TOT	-0.015	-0.055	1.33	-1.19	-0.026	0.018	0.013	0.019
	(0.032)	(0.032)	(1.36)	(1.36)	(0.020)	(0.019)	(0.025)	(0.026)
(4) Control means	0.493	0.496	25.85	28.26	0.106	0.082	0.197	0.199
Num. of obs.	2,572	2,664	2,709	2,665	2,727	2,680	2,727	2,680

Note: This table reports the effect of being assigned to and participate in one of the course types on the probability to pay social security, hours worked, the probability to be unemployed, and the probability to being inactive. Labor income is measured in 1,000 Chilean Pesos. FU 1 and FU 2 are the abbreviations for first and the second follow-up, respectively. The IV estimates are obtained by using the random assignment as instrument for participation in a course. Robust standard errors are reported in parentheses. All regressions control for course fixed effects, basic characteristics provided at application as well as household characteristics and non-cognitive skills measured at the baseline. All missing covariates are set to zero, and dummies for missing values are included.