

Juliana Camargo¹, Lycia Lima², Flavio Riva³ and André Portela Souza^{4*}

Technical Education, Non-cognitive Skills and Labor Market Outcomes: Experimental Evidence from Brazil

Abstract

This paper describes the results from an evaluation of a public policy that offers scholarships to current and former public high school students, so that they can attend technical and vocational education courses free of charge. We use a waiting list randomized controlled trial in four municipalities in a southern Brazilian State (Santa Catarina) to quantify the effects of the program on school progression, labor market outcomes and non-cognitive skills. Our intention-to-treat estimates reveal substantial gender heterogeneity two years after program completion. Women experienced large gains in labor market outcomes and non-cognitive skills. Employment rose by 21 percentage points (or approximately 33%) and the gains in earnings are of more than 50%. Also, women who received the offer scored 0.5σ higher on the synthetic index of non-cognitive skills and 0.69σ higher on an extraversion indicator. We find no effects on the male sub-sample. These findings corroborate the evidence on gender heterogeneity in the labor market effects of technical and vocational education programs. We also perform a series of exercises to explore potential channels through which these effects arise.

Current version: April 16, 2020
Keywords: Vocational Education, Human Capital, Non-cognitive Skills, Labor Market Outcomes
JEL-codes: C21, I38, J24
Corresponding author: André Portela Souza
andre.portela.souza@fgv.br

¹ Centers for Learning on Evaluation and Results, Brazil and Lusophone Africa, Fundação Getúlio Vargas (FGV EESP Clear), 286, Itapeva Street, São Paulo, SP, 01332-000, Brazil. E-mail: xcamargo@gmail.com.

² School of Business Administration, Fundação Getúlio Vargas (FGV EAESP) and Centers for Learning on Evaluation and Results, Brazil and Lusophone Africa, Fundação Getúlio Vargas (FGV EESP Clear), 286, Itapeva Street, São Paulo, SP, 01332-000, Brazil. E-mail: lycia.lima@gmail.com.

³ School of Business Administration, Fundação Getúlio Vargas (FGV EAESP), 286, Itapeva Street, São Paulo, SP, 01332-000, Brazil. E-mail: flaviorussoriva@gmail.com.

⁴ School of Economics, Fundação Getúlio Vargas (FGV EESP) and Centers for Learning on Evaluation and Results, Brazil and Lusophone Africa, Fundação Getúlio Vargas (FGV EESP Clear), 286, Itapeva Street, São Paulo, SP, 01332-000, Brazil. E-mail: andre.portela.souza@fgv.br.

1 Introduction

The fact that people might exit the educational system with low skill levels is a central policy concern because skills are seen as important determinants of labor market attachment, wages, and, ultimately, crime [Bertrand et al., 2017]. There is widespread agreement that the relevant set of skills is broad, and encompasses cognitive abilities, occupational-specific knowledge *and* non-cognitive or “soft” skills. In an extensive review of the topic, [Kautz et al., 2014] suggest that this distinction is also used by potential employers to evaluate workers, experiencing less difficulty “finding workers with adequate reading, writing, and math skills than with appropriate occupational, problem solving, teamwork, communication, and adaptability skills as well as positive work habits and a willingness to accept supervision” (p. 30). This places the question of how malleable the latter are, and what is their importance to the ease with which vulnerable youth can make the transition from school to work, at the forefront in the public debate on active labor market policies. These questions are of even greater importance in low or middle-income countries, where those unemployed out-of-school represent a large share of the society (approximately one in every five individuals from 15 to 23 years old in Latin America, cf., [Bertrand et al., 2017]).

Some technical and vocational education and training (TVET) programs are based on the idea that remediation is possible for children who grow up in disadvantaged environments and some were proven effective to promote vulnerable youth employment [Attanasio et al., 2011, Card et al., 2011, Kugler et al., 2015, Attanasio et al., 2017]. This issue is particularly important when considering the case of poor adolescents in developing countries, who only have access to low quality education and thus develop insufficient skills, which translates into access to low quality jobs and further contributes to the perpetuation of inequalities across time. One potential mechanism through which these programs may have affected beneficiaries is by building occupational-specific skills that are not taught in regular school and are demanded in the labor market. It is also conceivable that training impacts non-cognitive skills “workplace-based programs that teach non-cognitive skills appear to be effective remedial interventions for adolescents. They motivate acquisitions of work-relevant skills and provide for disadvantaged youth the discipline and guidance which is often missing in their homes and high schools”.[Kautz et al., 2014]. – such as job readiness, teamwork, discipline and responsibility, leadership, flexibility, communication capacity, perseverance and self-control –, either by explicitly teaching and exercising these skills or by providing an environment where they can be developed through the interaction with peers and teachers [e.g., Manual Operativo 2017; Strategy for technical and vocational education and training (TVET)(2016-2021) 2016], manualop]. One of the reasons why this is an important empirical question is that occupational-specific skills might become obsolete at a faster rate than non-cognitive skills, specially in the presence of rapid technological change. Although specific skills may facilitate an early transition into the labor market, [Hanushek et al., 2017] shows evidence that general human capital skills are more likely to bring benefits at later stages in life. The empirical evidence on whether training programs targeting youth can affect non-cognitive and general life “soft” skill formation, either by explicitly teaching and exercising these skills or by providing an environment where they can be developed, is relatively scant. It is thus somewhat surprising that the literature on TVET has focused almost exclusively on labor market outcomes. In sum, intuition, anecdotal evidence

and a growing empirical literature suggest that non-cognitive skills are important - maybe as important as cognitive skills - in succeeding in different domains of life [for instance, Mueller and Plug, 2006]. Nonetheless, the policy effects of bundling non-cognitive contents in vocational and skills training programs for youth are an understudied empirical question which could benefit from high-quality and large-scale empirical evidence.

This paper addresses some of these questions and enlarges the dimensions considered in the previous literature using a TVET program targeted at current and former high school students of the public educational system in Brazil, which is attended mostly by low income students. We describe the results from the evaluation of the Student Training Scholarship, a policy under a large public program — the National Program for Access to Technical Education and Employment (PRONATEC Program) — that offers scholarships so that eligible youth can attend TVET courses free of charge. We base our analysis of the causal effects of the program on a waiting list randomized controlled trial in four mid-sized municipalities in Santa Catarina State. Classes were offered by two of the main TVET providers in Brazil which, faced with excess demand, chose to make offers to their candidates through randomization in 2012, 2013 and 2014. These scholarships allowed interested eligible individuals enrolled in high school or high school graduates to attend to two years or 1,200 hours of in-classroom occupational training in mechanics, workplace safety, computer networks, electro-technology, food technology, and informatics, among others. Developing non-cognitive skills was among the programs' goals and there is direct evidence that employers consider them under-developed among youth in the setting of our study [FGV/CLEAR, 2017].

The analysis of the design generated by the waiting list shows that the sample is balanced across treatment offer arms after controlling for strata (course and time period in a given school) dummies. Also, despite the high rate of attrition (50%), the likelihood of being interviewed at endline is not significantly different in offer and no offer groups. We use survey and administrative data to provide intention-to-treat estimates of the scholarships on posterior investments in general education, labor market outcomes and non-cognitive skills. The offer to attend a technical course significantly increased the probability of graduating in the program and receiving a diploma in 45 percentage points (p.p.s), and increased the probability of graduating in any technical education program in 23 p.p.s. We consider the causal impacts of the offer as the relevant policy parameter in this case, but these can be easily inflated to local average treatment effects using the inverse of these differential take-up rates.

The main goal of the program is to generate better opportunities for beneficiaries in the labor market, so our analysis starts by describing the causal effect of offer assignment on employment, labor market participation, formality and earnings and find that men and women were affected very differently by the offer. The impacts on women are both large and strongly significant. They are also robust to nonparametric inference based on randomization inference tests. Employment rose by 21 percentage points (or approximately 33%), labor market participation rose by 26.2 p.p.s, formal employment rose by more than 35 percentage points (74%) and wages rose by roughly 23%. Considering the unconditional distribution of work earnings, we estimate an impact of approximately 513 Brazilian Reais or more than 58% of the no offer group mean. Point estimates for men are small in magnitude and statistically indistinguishable from zero.

The difference in effects between men and women we observe on labor market outcomes is also found in most of the standardized indicators of non-cognitive skills. This is reflected on the strong effects found on the synthetic index of non-cognitive skills for the subsample of women. For the different dimensions of non-cognitive, estimates for women tend to be large, and effects on men are not only small but frequently wrong-sided if we hypothesize that the program had positive effects on non-cognitive skills. This suggests that non-cognitive skills in women were malleable to the investments channeled through the program in dimensions that could play an important role in explaining the results found in the labor market.

The literature on the effects of TVET programs shows results are rather different if we consider the developing and developed countries contexts. In developed countries, evidence is mixed, registering both positive impacts of VET programs on labor market outcomes - such as wages ([Dearden et al., 2002], [Bishop and Mane, 2005], [Neuman and Ziderman, 1999]), and employment probability ([Bishop and Mane, 2005]; [Hanushek et al., 2017]) - as well as low or nonexistent economic returns found in other circumstances ([Woessmann, 2008], [Machin and Vignoles, 2005], [Jenkins et al., 2007], [Dearden et al., 2002], [Kane and Rouse, 1995], [Card, 1999]). The evidence also suggests that VET returns are higher in developing countries when compared to developed countries ([Attanasio et al., 2011]), specially on earnings ([Malamud and Pop-Eleches, 2010], [Tansel, 1998]). For Brazil in particular, some studies have previously found positive impacts of TVET on labor market variables ([Almeida et al., 2014], [Vasconcellos et al., 2010], [Assunção and Gonzaga, 2010], [Oliva, 2014], [Neri, 2010]). The gender heterogeneity we find on labor market outcomes has been pointed out in systematic reviews about technical and vocational education in developing countries [Blattman and Ralston, 2015] and other active labor market policies in general [Bergemann and Van den Berg, 2008, Card et al., 2010, Card et al., 2015, Caliendo and Kunn, 2015]. To our knowledge, [Alzúa et al., 2013] describe the only intervention where gender differences point to another direction, documenting that the Argentinian program “entra21” in Córdoba had positive and significant impacts on formal employment and earnings of men approximately 2 years after course completion, but not for women. We find strong evidence of heterogeneous effects in other dimensions, especially on non-cognitive skills, which have received less attention by the literature on TVET programs. Even though we are not able to gauge the specific mechanisms that underlie our results, we hypothesize that important traits related to sociability, organization and responsibility — as captured by the non-cognitive constructs of extraversion and conscientiousness — could have played a big role in explaining the results observed in the labor market dimension.

Our analysis comes with some caveats. First, our experimental data comes from four mid-sized municipalities in Brazil, and the public policy we evaluate is a national one. The setting we analyze could be very specific and differ from other locations in ways that limit our ability to generalize the results. In particular, it has been argued by previous non-experimental evaluations [Assunção and Gonzaga, 2015] that the impacts of the program and technical education in general vary according to region. Arguing against this, one could note that, although training courses are administered locally, quality is regulated by a set of federal guidelines. Also, the courses offered through the program are always provided by the same institutions in Brazil, which are part of a nationwide network and follow similar standards and overall curricula for provision across the country. Second, our analysis is limited to various short- or mid-term outcomes, averaging two years after program completion. Few studies provide impact estimates

for the period beyond two or three years after completion of training, but this shortcoming has been recently addressed in the case of the Colombian program “Jóvenes en Acción” [Atanasio et al., 2017]. Their results using administrative data suggest that the effects of the program are persistent. Following samples for a longer time to track whether the gains observed in non-cognitive skills can be persistent provides an interesting avenue for future research. Finally, our sample sizes are small, especially when compared to the most recent literature on TVET [see, for instance, Orazio et al, 2017]. The fact that randomization inference p -values are not different from the ones that rely on usual inference procedures nevertheless indicates that large-sample approximations are not relevant for the main qualitative results.

The remainder of the paper is divided as follows. Section 2 provides background information on the public school system in Brazil and on the design and implementation of the program. Section 3 describes the data (lottery, survey and administrative), the experimental design and the empirical strategy. Section 4 presents the intention-to-treat estimates of the offer on educational investments and labor market outcomes. Section 5 investigates possible channels through which PRONATEC could affect labor market outcomes. Section 6 concludes with a discussion of our findings.

2 Background and Program

Basic education in Brazil’s formal educational system after preschool is divided in preprimary (4-5 years old), primary (6-10 years old), lower secondary (11-14 old) and upper secondary or high school (15-17 years old). All educational levels can be completed either in private or in public institutions and one of the main features of the Brazilian school market is its segmentation. Students with lower family income are predominantly in public schools, which accounted for 86-90% of the enrollments in secondary school from 2000 to 2015. For instance, in 2014, only 2%, 3% and 4% of high school students in families in the first, second and third income quintile were in private schools, respectively [Almeida et al., 2017]. Quality indicators, like repetition and evasion rates are substantially higher in public schools [Costa, 2013] and students therein perform considerably worse than private high school students in tests that are used for admission at the tertiary level [26% worse in the National Secondary Education Exam, “Exame Nacional do Ensino Médio”, in 2005, cf. Braz et al, 2017].

Basic education is mandatory in Brazil so, in theory, all students must complete regular high school. Secondary technical education must be obtained in addition to regular high school. The share of students enrolled in such programs during high school is relatively low in Brazil, as compared to other countries [at 14%, compared to 33% in Chile and 60% in Italy, cf. Rita et al., 2015]. As a result of the federal government’s decision to massively scale up policies to increase the offer of professional training (including TVET), the Ministry of Education (MEC) created the National Program for Access to Technical Education and Employment (“Programa Nacional de Acesso ao Ensino Técnico e Emprego”, PRONATEC) in 2011. The program’s main goal is to expand the supply of training among current and former public school students — specifically to non-metropolitan areas of the country —, and it provided an umbrella for several overlapping policies that were already in place across the country [Almeida et al., 2015]. In 2014, year in which the program reached its maximum number of enrollments (roughly 8 million), PRONATEC offered 646 short qualification courses and 227 technical courses across

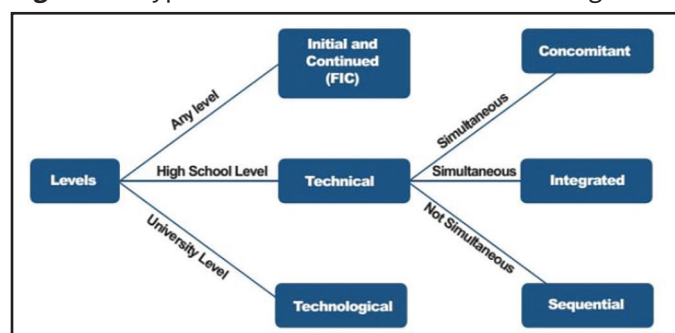
4,300 municipalities in Brazil [out of approximately 5,500, cf. Vivian et al., 2015]. In particular, between 2007 and 2013, investments in TVET programs increased from 0.04 to 0.2% of the GDP and enrollments have also increased sharply, by roughly 90%.

Through PRONATEC, public and private providers offer professional training course tracks for different age profiles and with different duration. Figure 1 illustrates the options available to potential beneficiaries. First, initial and continued training courses, which are short-term, with duration of less than 400 hours (minimum 160 hours). Available FIC courses are organized in conformity to the Brazilian Classification of Occupations (“Classificação Brasileira de Ocupações”, CBO). A full list of courses offered can be found in http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=41261-guia-pronatec-de-cursos-fic-2016-pdf&Itemid=30192 http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=41261-guia-pronatec-de-cursos-fic-2016-pdf&Itemid=30192. These courses do not count toward the completion of a formal general education level, and thus they do not qualify students to take courses at the next level. Second, TVET courses, which are offered in three modalities: (i) concomitant, by which regular education is taken independently of the technical education components, in an alternative shift, and often in a different institution; (ii) integrated, in which the student completes only one joint program that includes both general and vocational education; and (iii) sequential, which are attended by individuals who have already graduated secondary education, thus being only enrolled in the technical component. Finally, technological education courses, which can be attended only by students with a high school degree and are equivalent to a tertiary education diploma.

The focus of this paper is the Student Training Scholarship (“Bolsa Formação Estudante”) program, an initiative under PRONATEC that offers scholarships to students enrolled in the public high school system and to former students of public schools, so that they can attend initial and continued training courses or TVET courses (in the sequential and concomitant modalities) free of charge. The scholarships are not provided directly by the government to beneficiaries: through the program, MEC provides financial support to providers in exchange of vacancies to be filled by eligible individuals. This financial support cover all the expenses that would be paid by students to attend, including learning materials, food and transportation.

In our empirical setting, the scholarships allowed interested eligible individuals to attend to two years of a TVET course, which consists of in-classroom training at SENAI and SENAC, with a course load of 1,200 hours in a variety of subjects, such as mechanics, workplace safety, computer networks, electro-technology, food technology, and informatics, among others. Courses were offered at two of the most important private providers within the so-called S

Figure 1 Types of Courses Tracks Offered Through PRONATEC [Amorim et al., 2015].



System. The S System is a group of 10 “autonomous social services”, non-profit private entities that exert private activities of public interest within specific economic sectors (industry, commerce, agriculture), the National Service of Industrial Training (“Serviço Nacional de Aprendizagem Industrial”, SENAI) and the National Service of Commercial Training (“Serviço Nacional de Aprendizagem Comercial”, SENAC). As in most TVET programs, courses offered by these institutions are focused on providing trainees with occupational skills. Nevertheless, they explicitly aimed (as described in the curricula) to develop “soft” skills such as communication, creativity and the ability to work autonomously. The development of these skills was fostered by the introduction of activities throughout the occupational-specific curriculum, although there was no explicit module aimed at building them. Unlike other programs targeted at similar beneficiaries [for instance Orazio et al., 2011], there were no explicit partnerships with firms for a period of on-the-job training.

3 Data and Experimental Design

If providers face excess demand for a class (course in a given site and semester) offered by the program, they are allowed by law to choose which selection criteria to apply. Our study explores events of over-subscription in courses in four municipalities in the southern State of Santa Catarina (Chapecó, Itapiranga, São Miguel do Oeste and Xanxerê. These four municipalities were selected for two reasons: i) there was excess of demand for the program, ii) we could find the official registers of the lotteries. Consequently, we had access to the documents for the years of 2012, 2013 and 2014, which contained the list with the order and names of individuals who were selected (and of the ones who were not), in each class, for these four specific municipalities in the state.). Enrollments in each class were collected by the Santa Catarina Secretariat of Education and the institutions chose to select their candidates through a randomization criteria in 2012 (Chapecó), 2013 (all municipalities) and 2014 (Chapecó). During this period, SENAI and SENAC offered 34 classes and 29 of them had over-subscription. Each class had 35 slots to be filled and the schools kept a list of around twice the class size (around 70 subscribers per class) and followed the order of the list to fill all the open slots. It is possible that, in some cases, more than 70 individuals could have subscribed to a given class. However, information was kept only for the first randomly allocated 70 subscribers. To the extent that selecting the first 70 individuals acts as random sampling on the population of applicants, this fact won't compromise the main results of the paper. We had access to provider administrative data with offer assignments, which contained the names of individuals. In principle, an individual could subscribe for more than one class and thus increase her chances to be treated. However, in practice, very few of them actually had more than one class subscription and the position on the randomly generated waiting list. Additionally, this list contained information on gender, race and schooling.

3.1 Survey Data

We collected data using a survey from June to August of 2016. The interviews occurred at SENAI's schools in the municipalities of Chapecó and Xanxerê. For the municipalities of Itapiranga and São Miguel do Oeste, due to budget constraints, we made the survey through telephone

interviews. Answering the survey was not mandatory, but the majority of candidates who we could contact agreed to participate in the study. All interviews were conducted by a team of undergraduate students as enumerators, trained and supervised by the authors. The survey included questions on demographic and socioeconomic characteristics (age, sex, color, education degree, educational level of the parents and household condition) and labor market variables.

In our analysis of labor market attachment and work earnings, we focus on six outcome variables, which are constructed following the literature on the impacts of technical and vocational education programs [Card et al., 2011, Attanasio et al., 2011, Alz et al., 2013, Ibarra et al., 2015, Hirshleifer et al., 2016]: (i) whether the respondent had a job in the month before the survey; (ii) whether the respondent had a job or was looking for a job in the month before the survey; (iii) length of duration in months in the current most important job; (iv) whether the respondent claimed to have a signed work card (formal job); (v) monthly work earnings in the survey reference month; (vi) the last variable in logarithm.

We also applied a nationally validated instrument. The next paragraphs draw heavily on [Santos and Primi, 2014], named SENNA (Social and Emotional or non-cognitive Nationwide Assessment) to measure individuals non-cognitive skills. The creators of the instrument, developed by the Ayrton Senna Institute (“Instituto Ayrton Senna”, IAS), adopted the method proposed by [John and Srivastava, 1999] to frame the traits collected by tests within the Big Five framework, which groups personality traits in five basic dimensions: agreeableness, conscientiousness, extraversion, neuroticism and openness to experience. The authors also added a sixth dimension to these dimensions, namely, locus of control. These are latent constructs obtained by factor analysis performed on the answers to extensive questionnaires with questions about behaviors that are representative of characteristics that an individual could have. It is considered a well-accepted taxonomy of non-cognitive skills [Kautz et al., 2014] and encompasses a broad range of traits that could be valued in the labor market. The constructs used in our analysis of non-cognitive skills and the definitions are:

- *Agreeableness*: “tendency to act in a cooperative and unselfish manner [...] the agreeable or cooperative individual is characterized as being tolerant, altruistic, modest, likeable, flexible and objective (direct when dealing with people)” [p. 20 Santos and Primi, 2014];
- *Conscientiousness*: “tendency to be organized, hard working and responsible [...] the conscientious individual is characterized as being efficient, organized, autonomous, disciplined, lacking impulse and guided towards his objectives (a fighter)” [p. 19 Santos and Primi, 2014];
- *Extraversion*: “orientation of interests and energy towards the external world, people and things (instead of the internal world of subjective experience) [...] the extravert individual is characterized as being friendly, sociable, self-confident, energetic, adventurous and enthusiastic” [p. 20 Santos and Primi, 2014];
- *Neuroticism*: “the predictability and consistence of emotional reactions, without quick mood swings [...] the emotionally unstable individual is characterized as being short-tempered, introspective, impulsive and lacking in self-confidence, with a tendency towards depression and anxiety disorders” [p. 21 Santos and Primi, 2014];
- *Openness to experiences*: “tendency to be open to new aesthetic, cultural and intellectual experiences [...] the individual who is open to new experiences is characterized as being

imaginative, artistic, excitable, curious and unconventional, whilst having a wide range of interests” [p. 18 Santos and Primi, 2014];

- *Locus of Control* “how much individuals attribute current experiences either to decisions and attitudes they have taken in the past” [p. 22 Santos and Primi, 2014].

3.2 The Experiment

The study explores the over-subscription to the program that occurred in the SENAI and SENAC schools in four municipalities in the State of Santa Catarina, Brazil. The schools offer secondary technical education degree. The program offered financial support to these providers in exchange of free-of charge-enrollments of eligible individuals. They are students currently enrolled in or graduated from public high schools. They self-select into subscription and the subscriptions are made by class (course and time period in a given school) separately. The schools offered a varied range of courses such as mechanics, workplace safety, computer network, electrotechnology, food technology, and informatics, among others Table 17 shows the number of subscriptions by course and municipality. They offered 34 classes and 29 of them had over-subscription in the period of the analysis. The schools generated a random order of classification of subscribers for each class separately. In general, each class had 35 slots to be filled and a list of around twice the class size (around 70 subscribers per class) and followed the order of the list to fill all the open slots. The offer did not imply that the awarded students would accept the program. In this case, the scholarship offer followed the order given by the waiting list, until the 35 slots to be filled. This is the main reason the number of offered students is larger than the number no offer students. Thus, this is a typical over-subscription waiting list experimental design. It is possible that in some cases more than 70 individuals could have subscribed to a given class. However, information was kept only for the first randomly allocated 70 subscribers. To the extent that this acts as random sampling of the population we interpret the estimates as causal effects on the population of subscribers with typical SENAI and SENAC schools in Chapecó and Xanxerê, Itapiranga and São Miguel do Oeste. In principle, an individual could subscribe for more than one class and thus increase her chances to be treated endogenously. However, in practice it did not become a serious problem since very few of them actually had more than one class subscription.

Excess demand generated an waiting list randomized controlled trial design, which we leverage for causal identification. We use the following specification to compare outcomes of individuals across treatment offer arms:

$$y_{is} = \alpha + \tau_{ITT} W_i + X_i' \gamma + \mu_s + \varepsilon_{is} \quad (1)$$

where y_{is} denotes an outcome for individual i in lottery strata (class) s , W_i is an indicator variable that takes on the value one if individual i received an offer to participate in the program, X_i is a vector of individual i covariates (sex, age and color), μ_s are strata fixed-effects and ε_{is} is an error term. The intent-to-treat effect captures the effect of being offered the chance to participate in the program and is given by the ordinary least squares estimate of τ_{ITT} in equation (1). In the setting of a stratified randomized experiment, the ordinary least squares estimator $\hat{\tau}_{ITT}^{ols}$ is consistent for τ_{ITT} , the weighted average of within-stratum average intention-to-treat parameters [Bruhn and McKenzie, 2009, Imbens and Rubin, 2015]

One can also estimate the impact of actually completing training by replacing the offer variable W_i with a graduation dummy and instrumenting this with treatment assignment. Under the assumptions that the offer has no impact on outcomes for those who do not graduate in the course and there are no individuals who would take courses only if assigned to the control group, the instrumental variables estimator is consistent for the local average treatment effect (LATE) — the impact of graduating with a scholarship for an individual who takes up training when selected in the course-site lottery, and does not take it up otherwise. In a given sample, the estimate will be an “inflated” version of the intent-to-treat estimate. Table 4 shows that the offer affected the likelihood of graduating by 45 percentage points (std. err.= 0.041), and thus, for each outcome, the causal effect of graduating on *compliers* is approximately two times the effect of the offer on the full sample.

We present p -values for inference using heteroskedasticity-robust standard errors. Standard errors are not clustered because randomization was conducted at the individual level [Abadie et al., 2017]. Nevertheless, a natural question is whether our statistical inferences are sound, given the relatively small number of observations. As an alternative to standard t -tests to determine statistical significance, throughout the discussion of our experimental findings, we also report results from randomization inference permutation tests. These tests are non-parametric in the sense that inference does not rely on distributional assumptions about the error term in a given sample nor on asymptotic approximations as the sample size grows to infinity [Imbens and Rubin, 2015]. In practice, they rely on generating a series of placebo offer assignment vectors that mimic the underlying assignment process and computing the empirical distribution of average intention-to-treatment effects. The p -values are then obtained using the rank of the actual offer assignment vector. Since we rely on individual level randomization within classes, we use 1,000 randomly drawn offer vectors with fixed within-strata probabilities taken from the realized assignment and obtain p -values in the same way.

3.3 Samples, Attrition and Balancedness

We started with the full set of 29 oversubscribed lists with random orders of classification for 1,444 individuals. We were able to use valid phone numbers and interview 735 (51%) individuals successfully (237 women and 498 men). In previous versions of this draft, we also analyzed the results for a sub-sample of individuals aged 36 or below. We eliminated 63 subscribers who were above this age threshold at the time of subscription to concentrate on impacts on individuals who takes the decision to enroll as an investment in human capital in early stages of the life cycle. The results are qualitatively similar to the main sample ones and are available upon request. The main source of attrition was outdated contact information, not refusal to participate in the interview. Table 1 shows that there is no significant differences in the attrition rates of treatment and control groups. The dependent variable is an indicator whether the individual in the provider administrative list was not interviewed. We run regression for all subscribers as well as for women and men separately. There is no difference in the attrition rates of treatment and control groups within gender sub-samples either.

Tables 2 and 3 suggest that characteristics are balanced across treatment offer arms, both in the full sample and considering the sub-samples of women and men separately. Both tables present means across offer arms (columns (a)-(b)), p -values (column (c)) from a test of difference

Table 1 Attrition

Dep. Var.:	All Subsc.	Women	Men
Offer	−0.047 [0.043]	−0.057 [0.104]	−0.033 [0.048]
<i>p</i> –value	0.267	0.585	0.489
Number of observations	1,444	467	977

Notes: This table considers attrition in the full group of subscribers, and conditional on gender. Coefficients, standard errors and *p*-values are computed using specification (1), an ordinary least squares regression with course-by-site strata dummies and heteroskedasticity-robust standard errors [in brackets], using an indicator of not being in the endline sample as the dependent variable. Specification included control variables for sex, age, color and education.

Table 2 Within-Strata Balance Across Treatment Offer Arms, All Interviewed Subscribers

	(a)	(b)	(c)	(d)	(e)
	Mean	Mean	p–value	%	Number
	(No offer)	(Offer)	(w. strata)	Size	of obs.
Male	0.698	0.673	0.783	−3.6%	735
Black, 'pardo' or indigenous	0.206	0.261	0.377	26.5%	735
Age	27.25 (7.60)	23.80 (6.46)	0.949	−12.7%	735
Mother completed high-school	0.190	0.258	0.645	35.6%	734
Mother completed tertiary	0.071	0.090	0.418	26.4%	735
Father completed high-school	0.230	0.247	0.311	7.2%	734
Father completed tertiary	0.048	0.061	0.944	27.6%	735
Concomitant modality	0.246	0.473	0.072	92.2%	735
N	126	609			
	<i>F</i> –stat	<i>p</i> –value			
	0.838	0.569			

Notes: This table considers balancedness in the group of applicants surveyed at endline. Column (a) (column (b)) shows means and standard deviations (in parenthesis) for a pre-determined variable, for those not offered (offered) the opportunity to participate in the program with the scholarships in the schools and time frame considered. *p*-values in column (c) are computed for each variable using specification (1), an ordinary least squares regression with course-by-site strata dummies and heteroskedasticity-robust standard errors. Column (d) displays the unadjusted percent size of the raw difference between treatment offer arm means with respect to the control group mean, and column (e) shows the sample size. The last row shows the *F*-statistic and *p*-value from a regression of the treatment indicator on the full set of pre-determined variables and course-by-site strata dummies. Additional details on how the outcomes were created can be found in (see section "Definition of Main Variables" in Appendix).

Table 3 Within-Strata Balance Across Treatment Offer Arms, All Interviewed Subscribers by Gender Group

	Women					Men				
	(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)
	Mean (No offer)	Mean (Offer)	p-value (w. strata)	% Size	Number of obs.	Mean (No offer)	Mean (Offer)	p-value (w. strata)	% Size	Number of obs.
Black, "pardo" or indigenous	0.184	0.256	0.189	39.1%	237	0.216	0.263	0.829	22.0%	498
Age	28.21 (8.24)	22.71 (5.40)	0.109	19.5%	237	26.84 (7.32)	24.33 (6.86)	0.228	-9.3%	498
Mother completed high-school	0.184	0.256	0.909	39.1%	237	0.193	0.259	0.774	34.2%	497
Mother completed tertiary	0.053	0.095	0.974	81.4%	237	0.080 (0.272)	0.088 (0.283)	0.304	10.4%	498
Father completed high-school	0.211	0.266	0.373	26.5%	237	0.239	0.237	0.506	-0.6%	497
Father completed tertiary	0.026	0.040	0.922	52.8%	237	0.057	0.071	0.917	24.5%	498
Concomitant modality	0.237	0.548	0.989	131.3%	237	0.250	0.437	0.064	74.6%	498
N	38	199				88	410			
		<i>F</i> -stat	<i>p</i> -value				<i>F</i> -stat	<i>p</i> -value		
		1.062	0.390				0.704	0.669		

Notes: This table considers balancedness in the group of female (leftmost columns) and male (rightmost columns) applicants surveyed at endline. Column (a) (column (b)) in each panel shows means and standard deviations (in parenthesis) for a pre-determined variable in a gender group, for those not offered (offered) the opportunity to participate in the program with the scholarships in the schools and time frame considered. *p*-values in column (c) are computed for each variable using specification (1), an ordinary least squares regression with course-by-site strata dummies and heteroskedasticity-robust standard errors, and *p*-values less than 0.15 are bolded. Column (d) displays the unadjusted percent size of the raw difference between treatment offer arm means with respect to the control group mean, and column (e) shows the sample size. The last row shows the *F*-statistic and *p*-value from a regression of the treatment indicator on the full set of pre-determined variables and course-by-site strata dummies. Additional details on how the outcomes were created can be found in (see section "Definition of Main Variables" in Appendix).

between means within groups computed using regressions with lottery strata dummies, the percent size of the difference between treatment offer arms normalized by the control group mean (column (d)) and the sample size (column (e)). Some of the differences between treatment and control are large in magnitude but they become statistically insignificant at reasonable confidence levels once we rely only in within-strata variation. One could argue that the small sample size could mask deviations from randomization. However, we ran both specifications, with and with no control variables, and the results are robust for both estimations. Finally, *p*-values for the test of joint significance are not significant, suggesting that there is no systematic correlation between observables and the offer to participate in PRONATEC.

4 Main Results

Below, we provide a series of estimates of the impact of the program offer on educational investments in technical and general education as well as labor market outcomes. All tables

share a similar structure. The results for the full set of surveyed applicants are in Panel A and the results for the sub-groups of female and male surveyed applicants are in Panels B and C, respectively. This section presents the intent-to-treat estimates. The LATE estimates for the main outcomes are presented in (see section "Subscription information" in Appendix).

4.1 Educational Investments

Our results show that there was a high take-up of the program, as individuals that received an offer to participate in the program were 78 p.p.s more likely to enroll at PRONATEC, and 45 p.p.s more likely to graduate at PRONATEC at endline, as it can be observed in columns (a) and (c) of Table 4. We show in Table 13 that self selection into the program among those that have been offered the opportunity to participate is not a major concern, as there are no significant differences among observable characteristics between those who have enrolled and those who have not.

These enrollment and graduation results do not substantially differ among men and women as shown in Panels B and C of Table 4. Note that there were some leakages as some non-offer group individuals have enrolled and graduated in the program too, but with much lower incidence. Indeed, 5.3 percent of women and 10.2 percent of men from the groups that did not receive an offer have graduated in the program. These are individuals that did not get the scholarship but did enroll nonetheless in a technical course paying the tuition themselves.

One concern is that the program's scholarships could have simply crowded out the demand for similar courses. In order to investigate this possibility, we asked individuals if they have enrolled in any TVET course. The impact estimates of having a scholarship offer on the probability of enrollment in any TVET course (including PRONATEC) are presented in Column (b) of Table 4. Roughly 20% of individuals in the non-offer group enrolled in any technical education course. Nevertheless, the offer increased the probability of enrollment in any TVET course by 64.1 p.p.s. Taken all together, these results suggest that the crowding out effect is not substantial, as the program has majorly increased the overall likelihood of enrolling in a secondary technical education degree and thus seemed to have relaxed some choice constraints faced by eligible individuals who expressed interest in participating.

We also investigate the impact of the scholarship program on progression in general education. The Brazilian education system requires all individuals to have a general education high school degree. Individuals pursuing secondary technical education — for instance, through the Student Training Scholarship in a PRONATEC course — must comply with this norm, at the cost of not receiving a formal secondary technical education degree. Thus, high school graduation acts as a conditionality for obtaining the TVET degree and could reinforce the compliance with the compulsory educational standard. We estimate the actual size of its effect. Column (e) in Table 4 presents intention-to-treat effects on the probability of completing high school for individuals in the concomitant modality. We find that those offered an opportunity to enroll were 15 p.p.s (st.err.= 0.068) more likely to have graduated, a 19% increase over what would be expected in the absence of the offer. This result is driven primarily by the sample of men, who are 21.0 p.p.s (st.err.= 0.088) more likely to have a high school diploma at endline. Among women, there is already a high incidence of regular high school completion (89%). Perhaps for this reason, the effect in this sub-sample is not statistically significant.

Table 4 Average Intent-to-treat Estimates on Technical and Non-Technical Education Graduation Rates

Dep. Var.:	Technical Education			General Education		
	(a)	(b)	(c)	(d)	(e)	(f)
	Enrolled (PRONATEC)	Enrolled (any)	Graduated (PRONATEC)	Graduated (any)	Graduated High-school, concomitant	Enrolled (Tertiary)
Panel A: All						
Offer	0.777 [0.030]	0.641 [0.044]	0.452 [0.041]	0.234 [0.054]	0.147 [0.068]	-0.034 [0.051]
No offer group mean	0.016	0.198	0.087	0.437	0.774	0.389
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.000	0.000	<0.001	<0.001	0.031	0.513
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.000	0.000	<0.001	<0.001	0.033	0.253
Number of observations	735	735	735	735	319	735
Panel B: Women						
Offer	0.736 [0.055]	0.596 [0.089]	0.458 [0.070]	0.200 [0.101]	0.014 [0.074]	0.016 [0.108]
No offer group mean	0.026	0.211	0.053	0.368	0.889	0.447
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.000	0.000	<0.001	0.050	0.845	0.884
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.000	0.000	<0.001	0.022	0.429	0.413
Number of observations	237	237	237	237	118	237
Panel C: Men						
Offer	0.815 [0.034]	0.678 [0.052]	0.464 [0.051]	0.238 [0.067]	0.210 [0.088]	-0.018 [0.059]
No offer group mean	0.011	0.193	0.102	0.466	0.727	0.364
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.000	0.000	<0.001	<0.001	0.018	0.756
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.000	0.000	<0.001	<0.001	0.033	0.377
Number of observations	498	498	498	498	201	498

Notes: This table considers the impact of the offer on technical and general education investments in the full sample of surveyed applicants (Panel A) and the sub-sample of female and male surveyed applicants (Panels B and C, respectively). All variables are self-reported and were collected from June to August of 2016 (an average of 2 years after the end of the program in our sample). Estimates are from the intent-to-treat specification (1), i.e., an ordinary least squares regression with course-by-site strata dummies, and standard errors computed using heteroskedasticity-robust standard errors are shown [in brackets]. Two p -values are displayed in the bottom rows for each panel: (i) two-tailed p -values for the null hypothesis of a zero average parameter $H_0 : \tau_{ITT} = 0$, using the reported standard errors; (ii) two-tailed randomization inference p -values for the sharp null hypothesis of no effect of the offer for any individual $H_0 : \tau_{ITT,i} = 0, \forall i$, using 1,000 randomly drawn assignment vectors with fixed within-strata probabilities taken from the realized assignment. Additional information about the samples in each panel (no offer arm mean and sample size) can be found in the bottom rows and details on how the outcomes were created can be found in (see section "Definition of Main Variables" in Appendix).

The scholarships could also induce individuals not to seek post-secondary general education. Column (f) in Table 4 considers whether applicants that received the offer were more or less likely to be enrolled or have completed college at endline. The point estimates are small in magnitude and imprecisely estimated: taking the full sample (Panel A) as an example, a 95% confidence interval covers values that are empirically relevant, from approximately -15 to 10 p.p.s over a counterfactual probability of approximately 40%. We thus find no conclusive evidence that participation led to decreases in post-secondary general education enrollment.

4.2 Labor Market

Panel A in Table 5 presents evidence on the impact of the program on various measures of labor market attachment and job quality. Column (a) shows that individuals offered training were 4.3 p.p.s more likely to be employed at the time of the survey, but this difference is statistically insignificant (p -value = 0.301). We also investigate overall labor market participation, which takes into consideration whether individuals were either employed or looking for a job. Column (b) shows that lottery winners were 5.3 p.p.s more likely to participate in the labor market, even though this effect is also statistically insignificant. We also find no effects on self-reported tenure, consistently with the lack of overall impacts on employment and labor market par-

Table 5 Average Intent-to-treat Estimates on Labor Market Outcomes

Dep. Var.:	(a)	(b)	(c)	(d)	(e)	(f)
	Employment	Labor market	Tenure	Formal	Work	Wages
		participation		employment	earnings	(in logs)
Panel A: All						
PRONATEC Offer	0.043	0.053	4.258	0.051	102.688	0.046
	[0.042]	[0.034]	[4.486]	[0.051]	[108.5]	[0.050]
No offer group mean	0.802	0.865	33.008	0.690	1399.135	7.345
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.301	0.124	0.393	0.315	0.344	0.360
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.270	0.102	0.345	0.309	0.312	0.399
Number of observations	735	735	726	735	731	567
Panel B: Women						
PRONATEC Offer	0.217	0.262	16.410	0.354	513.000	0.229
	[0.093]	[0.078]	[6.663]	[0.097]	[149.1]	[0.101]
No offer group mean	0.658	0.711	24.711	0.474	873.026	7.055
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.020	0.002	0.015	<0.001	0.001	0.025
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.008	0.000	0.011	<0.001	0.001	0.018
Number of observations	237	237	235	237	237	166
Panel C: Men						
PRONATEC Offer	-0.017	-0.027	-0.542	-0.068	-39.420	-0.007
	[0.047]	[0.035]	[5.731]	[0.058]	[144.3]	[0.060]
No offer group mean	0.864	0.932	36.632	0.784	1626.318	7.445
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.717	0.416	0.736	0.252	0.899	0.810
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.825	0.487	0.926	0.254	0.783	0.906
Number of observations	498	498	491	498	494	401

Notes: This table considers the impact of the offer on labor market outcomes in the full sample of surveyed applicants (Panel A) and the sub-sample of female and male surveyed applicants (Panels B and C, respectively). All variables are self-reported and were collected from June to August of 2016 (an average of 2 years after the end of the program in our sample). Estimates are from the intent-to-treat specification (1), i.e., an ordinary least squares regression with course-by-site strata dummies, and standard errors computed using heteroskedasticity-robust standard errors are shown [in brackets]. Two p -values are displayed in the bottom rows for each panel: (i) two-tailed p -values for the null hypothesis of a zero average parameter $H_0 : \tau_{ITT} = 0$, using the reported standard errors; (ii) two-tailed randomization inference p -values for the sharp null hypothesis of no effect of the offer for any individual $H_0 : \tau_{ITT,i} = 0, \forall i$, using 1,000 randomly drawn assignment vectors with fixed within-strata probabilities taken from the realized assignment. Additional information about the samples in each panel (no offer arm mean and sample size) can be found in the bottom rows and details on how the outcomes in columns (a)-(f) were created can be found in (see section "Definition of Main Variables" in Appendix).

ticipation. Column (d) shows that lottery winners were 5.1 p.p.s more likely to be formally employed (i.e., to have a signed work card), but this effect is also imprecisely estimated in the full sample (p -value = 0.315). Columns (e) and (f) suggest that the impacts on work earnings and wages were similar to the results found for other labor market variables: self-reported work earnings increased by roughly 100 Reais or, considering the conditional distribution, 4.6%. In both cases, we are also unable to reject the null hypothesis of zero intent-to-treat parameters.

The results in the full sample of applicants hide significant gender heterogeneity, as can be seen by analyzing Panels B and C. Women offered training were 21.7 p.p.s (st. err. = 0.093) or 33% more likely to be employed at the time of the endline survey, as can be seen in column (a). This result is significant at the 0.05-level, either considering the usual heteroskedasticity-robust standard errors and p -values from randomization inference tests. Column (b) shows that the effects on labor market participation are stronger than those found on employment. Lottery winners are 26.2 p.p.s (st. err. = 0.078) more likely to participate on the labor market and this result is statistically significant at the 0.01-level, either considering the usual heteroskedasticity-robust standard errors and p -values from randomization inference tests. This finding suggests that the program has induced more women into looking for work outside home, as it has likely increased their opportunity costs.

Column (c) shows that women who won the lottery had also been for roughly 16 months longer in their current most important job. There is also a highly significant and large impact of the program on the probability of working in the formal sector, as opposed to either not working at all or working in the informal sector, suggesting that much of the gain in employment was into formal jobs. More precisely, column (d) shows that women offered training were 35.4 p.p.s (st. err. = 0.097) more likely to be formally employed, a large difference that amounts to approximately 75% of the mean of the group not offered training. Also, work earnings rose substantially because of the program, by roughly 513 Reais or almost 59%. These results are even more surprising if one notes that most of the coefficients for the sub-sample of men displayed in Panel C are imprecisely estimated, but often negative.

One possible explanation for this lack of impact may be due to regional characteristics, since Santa Catarina has the lowest unemployment rate comparing to the rest of the country. For instance, in the third trimester of 2016, unemployment rate in Santa Catarina was 5.7%. According to data from Continuous National Household Sample Survey - Continuous PNAD of 2016., while for Brazil it was 11.3% in this same period. On the other hand, we find positive and statistically significant effect of the program on the subsample of women. One can hypothesize that the unemployment rate is higher for women, about 7%, than for men, about 5%. As a consequence, there is more space to quantify the PRONATEC's impact on women's labor market outcomes. These issues regarding low unemployment rate and gender heterogeneity are aligned with the figures reported in our survey: 78% of the interviewees reported to be employed at the time of the survey, 77% in the treatment group and 80% in the control group. For women, 70% of the interviewees reported to be employed at the time of the survey - 71% and 66% in the treatment and control group, respectively. For men, 82% of the interviewees reported to be employed at the time of the survey - 81 and 86% in the treatment and control group -, respectively.

There are also other potential reasons why the program may affect men and women differently. Program participation may have an effect on labor market outcomes through different

channels. One possibility is that program participation increases non-cognitive skills, which may have an effect on employment. Another possibility is that program participation may act as a gateway for employment, with students being referred to job openings by the school or school instructors. Students may also develop, through the program, job specific skills, which may facilitate the students transition to the labor market. Another potential mechanism is the credentialism effect - students with a diploma from PRONATEC may have an advantage when applying for jobs. If the impact of program participation on these potential mechanisms is different among men and women, this might shed some light on the gender differentials we find on labor market outcomes. In the next section, we investigate the effect of the program on these intermediate outcomes, and investigate the plausibility of these as mechanisms through a structural equation exercise.

5 Potential Channels

The previous section documented gender heterogeneity on labor market outcomes. Our results show strong effects of the scholarship offer on women's labor market outcomes, but no effects in these outcomes either on the full sample of all subscribers or on men's subsample. TVET may impact labor market outcomes directly if the program acts through credentialism or a gateway for employment; but also indirectly, through enhancing skills - such as cognitive, non-cognitive or occupational-specific skills - which are valuable on labor market.

In this section, we examine potential mechanisms through which the program may impact labor market outcomes. First, we examine whether the program has an impact on the intermediate outcomes that may act as potential channels for the labor market effects. Second, we further explore this possibility through a structural equation model (SEM) framework, in which we assume that the offer has an effect on the potential mechanisms, which in turn have an effect on labor market outcomes.

5.1 Impacts on Intermediate Outcomes

5.1.1 Non-cognitive Skills

First possible channel we investigate is the impact of the scholarship offer on the non-cognitive skills presented in Section 3. Results are shown in Table 6. In addition to the latent constructs measures, we also created an aggregate index of non-cognitive skills (Column (a)), which is computed by using factor analysis to combine the individual's scores of the constructs presented in columns (b) - (g). As discussed before, the program offer could have had an effect on enhancing non-cognitive skills which, in turn, may have had positive impacts on individuals' labor market outcomes.

All variables displayed in Table 6 have been standardized with respect to full sample of applicants, so the coefficients can be interpreted as changes in terms of the no offer group standard deviation. The sample for non-cognitive measures differs from the number of observations for the other measures since, for this outcome, the instrument was self-reported, not interviewed by the enumerator. We investigate whether this could be an additional source of selection bias. Table 12 shows that this possible attrition is not a concern, since there is no statistically significant correlation between answering the non-cognitive module and the treatment.

Table 6 Average intent-to-treat Estimates on Standardized Indicators of Non-cognitive Skills

Dep. Var.:	(a)	(b)	(c)	(d)	(e)	(f)	(g)
	NC Index	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness to experiences	Locus of control
Panel A: All							
Offer	0.148 [0.141]	0.062 [0.147]	0.215 [0.143]	0.182 [0.161]	0.014 [0.146]	-0.030 [0.157]	-0.068 [0.157]
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.296	0.673	0.133	0.258	0.924	0.847	0.664
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.306	0.335	0.067	0.132	0.467	0.381	0.343
Number of observations	376	364	364	363	364	364	364
Panel B: Women							
Offer	0.503 [0.183]	0.360 [0.218]	0.421 [0.241]	0.690 [0.300]	0.251 [0.262]	0.284 [0.304]	-0.307 [0.283]
No offer group mean	-0.316	-0.171	-0.038	-0.378	-0.513	-0.368	0.235
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.007	0.102	0.084	0.024	0.341	0.352	0.280
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.028	0.089	0.066	0.009	0.188	0.205	0.138
Number of observations	111	106	106	105	106	106	106
Panel C: Men							
Offer	0.0517 [0.188]	-0.002 [0.189]	0.183 [0.181]	0.062 [0.189]	-0.055 [0.187]	-0.082 [0.188]	0.001 [0.195]
No offer group mean	0.110	0.056	0.012	0.123	0.167	0.120	-0.077
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.783	0.909	0.313	0.741	0.770	0.665	0.997
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.761	0.466	0.138	0.403	0.401	0.354	0.474
Number of observations	265	258	258	258	258	258	258

Notes: This table considers the impact of the offer on non-cognitive skills in the full sample of surveyed applicants (Panel A) and the sub-sample of female and male surveyed applicants (Panels B and C, respectively). All variables have been standardized with respect to the no offer group mean and standard deviation and were collected from June to August of 2016 (an average of 2 years after the end of the program in our sample). Estimates are from the intent-to-treat specification (1), i.e., an ordinary least squares regression with course-by-site strata dummies, and standard errors computed using heteroskedasticity-robust standard errors are shown in brackets. Heteroskedasticity-robust standard errors are in brackets, and two p -values are displayed in the bottom rows for each panel: (i) two-tailed p -values for the null hypothesis of a zero average parameter $H_0 : \tau_{ITT} = 0$, using the reported standard errors; (ii) two-tailed randomization inference p -values for the sharp null hypothesis of no effect of the offer for any individual $H_0 : \tau_{ITT,i} = 0, \forall i$, using 1,000 randomly drawn assignment vectors with fixed within-strata probabilities taken from the realized assignment. Additional information about the samples in each panel (no offer arm mean and sample size) can be found in the bottom rows.

Table 7 Average intent-to-treat Estimates on Possible Channels Outcomes

Dep. Var.:	Possible Channels			
	(a)	(b)	(c)	(d)
	NC Index	Occupation (course area)	Certificate	Gateway
Panel A: All				
Offer	0.148 [0.141]	0.175 [0.049]	0.286 [0.079]	-0.017 [0.028]
No offer group mean		0.165	0.270	0.059
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.296	0.000	0.000	0.529
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.306	0.001	0.000	0.467
Number of observations	376	571	587	569
Panel B: Women				
Offer	0.503 [0.183]	0.079 [0.041]	0.283 [0.201]	0.015 [0.023]
No offer group mean	-0.316	0.043	0.250	0.000
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.007	0.054	0.160	0.503
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.028	0.206	0.025	0.807
Number of observations	111	163	183	164
Panel C: Men				
Offer	0.052 [0.188]	0.210 [0.064]	0.283 [0.089]	-0.023 [0.037]
No offer group mean	0.110	0.203	0.276	0.078
(i) p -value ($H_0 : \tau_{ITT} = 0$)	0.783	0.001	0.002	0.534
(ii) p -value ($H_0 : \tau_{ITT,i} = 0, \forall i$)	0.761	0.003	0.000	0.539
Number of observations	265	408	404	405

Notes: This table considers the impact of the offer on possible channels outcomes in the full sample of surveyed applicants (Panel A) and the sub-sample of female and male surveyed applicants (Panels B and C, respectively). All variables are self-reported and were collected from June to August of 2016 (an average of 2 years after the end of the program in our sample). Estimates are from the intent-to-treat specification (1), i.e., an ordinary least squares regression with course-by-site strata dummies, and standard errors computed using heteroskedasticity-robust standard errors are shown [in brackets]. Two p -values are displayed in the bottom rows for each panel: (i) two-tailed p -values for the null hypothesis of a zero average parameter $H_0 : \tau_{ITT} = 0$, using the reported standard errors; (ii) two-tailed randomization inference p -values for the sharp null hypothesis of no effect of the offer for any individual $H_0 : \tau_{ITT,i} = 0, \forall i$, using 1,000 randomly drawn assignment vectors with fixed within-strata probabilities taken from the realized assignment. Additional information about the samples in each panel (no offer arm mean and sample size) can be found in the bottom rows and details on how the outcomes in columns (a)-(d) were created can be found in (see section "Definition of Main Variables" in Appendix).

The results show, once again, substantial gender heterogeneity driven by the positive results on women. As can be seen in column (a) of Table 6, the scholarship offer has a positive and statistically significant effect on the synthetic index of non-cognitive skills only in the sub-sample of women (0.5 standard deviations). The significance is robust for both heteroskedasticity-robust standard errors and randomization inference permutation tests. These results are stronger for the bottom of the distribution, as shown in Table 8, which reports the effects of program participation on the distribution of synthetic non-cognitive index. Program participation had a statistically significant impact on the probability of being above the bottom 20 percent of the index distribution.

Table 8 Average Intent-to-treat Estimates on Probability of being above Quantiles of Standardized Index of Non-cognitive Skills

	(a)	(b)	(c)
	All	Women	Men
Probability of being above 20%			
Offer	0.0621	0.116	0.0446
	[0.0319]	[0.0561]	[0.0384]
p-value (w. strata)	0.052	0.038	0.246
Probability of being above 40%			
Offer	0.0524	0.0652	0.0507
	[0.0413]	[0.0779]	[0.0485]
p-value (w. strata)	0.205	0.402	0.296
Probability of being above 60%			
Offer	0.0224	0.0615	0.0228
	[0.0461]	[0.0844]	[0.0548]
p-value (w. strata)	0.627	0.466	0.678
Probability of being above 80%			
Offer	0.0337	-0.0655	0.0725
	[0.0474]	[0.0871]	[0.0559]
p-value (w. strata)	0.477	0.452	0.194
P-value joint	0.3442	0.0502	0.4181
Number of observations	735	237	498

Notes: This table considers the impact of the offer on the probability of being above the 2nd, 4th, 6th and 8th deciles of the standardized non-cognitive index in the full sample of surveyed applicants (column (a)) and the sub-sample of female and male surveyed applicants (columns (b) and (c), respectively). The estimates are from simultaneous regressions of dummies for being above each decile on offer status, controls for color, age, gender and course-by-site strata dummies. Standard errors computed using heteroskedasticity-robust standard errors are shown [in brackets]. Two-tailed p -values for the null hypothesis of a zero average parameter $H_0 : \tau_{ITT} = 0$ are displayed in the bottom rows for each decile estimation. The p -value at the bottom of the table is for the null hypothesis that all coefficients are equal to zero.

We also perform an exercise in Table 9 to try to investigate heterogeneity by observables which may be proxies of non-cognitive skills at baseline. We use parents' educational level as a potential proxy for the level of non-cognitive skills of participants before the program. However, no statistically significant effects are found.

When analyzing each dimension separately, gender heterogeneity persists. Women who won the scholarship lottery scored approximately 0.36 standard deviations higher in the agreeableness indicator (column (b)). The coefficient is not significant, but large, specially when compared to what happened to men who received an offer. Columns (c), (d), (e) and (f) show that the same can be said about the conscientiousness, extraversion, neuroticism and openness to experiences. In all cases, there is no evidence at all that the program was associated with improvements in any of the indicators for men, while women fared at least 0.2 standard deviations better in all of the Big Five indicators. In the case of the extraversion, which describes personality characteristics like sociability and self-confidence, the index rose sharply and

Table 9 Possible Heterogeneous Effects

	(a)	(b)	(c)
Dep. Var.:	Employment	Employment	Employment
	All	Women	Men
Offer	-0.000 [0.045]	0.241 [0.111]	-0.099 [0.047]
<i>p</i> -value	0.999	0.030	0.038
Mother completed high-school	-0.242 [0.096]	-0.131 [0.220]	-0.292 [0.098]
<i>p</i> -value	0.012	0.554	0.003
Father completed high-school	0.152 [0.105]	0.046 [0.230]	0.192 [0.111]
<i>p</i> -value	0.149	0.841	0.084
Offer*Mother completed high-school	-0.004 [0.092]	0.064 [0.235]	-0.043 [0.092]
<i>p</i> -value	0.968	0.785	0.643
Offer*Father completed high-school	-0.009 [0.101]	-0.185 [0.247]	0.090 [0.102]
<i>p</i> -value	0.928	0.455	0.379
Constant	0.779 [0.093]	0.373 [0.194]	1.004 [0.099]
<i>p</i> -value	0.000	0.056	0.000
Number of observations	735	237	498

Notes: This table considers possible heterogeneous effects in the full group of subscribers, and conditional on gender. Coefficients, standard errors and *p*-values are computed using an ordinary least squares regression with course-by-site strata dummies and heteroskedasticity-robust standard errors [in brackets], using an indicator of being employed in the endline sample as the dependent variable. Specification included control variables for sex, age, and color. Parents' educational level are used as a proxy for non-cognitive skills.

significantly for women, by 0.69σ (st. err. = 0.30), and significance is robust to the use of randomization inference permutation tests.

All in all, this suggests that non-cognitive skills may be malleable — especially for women — to the investments channeled through the program in dimensions that are important for labor market success. The relationship between non-cognitive skills and labor market outcomes has been documented by other studies in the literature. Personality traits are found to be linked to labor productivity ([Bowles et al., 2001]). Conscientiousness, for instance, has been found to be largely associated to wage and other labor market outcomes ([Almlund et al., 2011], [Nyhus and Pons, 2005], [Hogan and Holland, 2003], [Salgado, 1997], [Barrick and Mount, 1991]). Similarly, there is evidence of a positive relationship between extraversion and labor market outcomes ([Cattan, 2010], [Heckman, 2007]).

5.1.2 Job-Specific Skills

A second possible channel we investigate is the impact of the program offer on job-specific skills. We assume that different courses offered by PRONATEC program could enhance individuals' specific skills, which are required in some occupations, increasing their probability

of being employed in these occupations. Thus, we use as a proxy for job specific skills the variable which indicates whether the individual is employed in the same area of the course undertaken through PRONATEC. Table 7 (Column (b)) reports the ITT estimates on the probability of having an occupation in the course area (conditional on being employed). The effect of the scholarship offer is positive and statistically significant for this outcome (17.5 p.p.s - p -value = 0.000). Panels B and C show that the receiving an offer to participate of the program acts differently than expected through this channel, having a higher impact on men. Women in the offer group were 7.9 p.p.s (p -value = 0.054) more likely to be employed in the course area at the time of the survey, whereas for men this probability is 21.0 percentage points (p -value = 0.001) higher for the ones who received a scholarship offer. Therefore, this finding does not allow us to shed light on the gender differences observed on our labor market results, which are hence unlikely to have occurred through this channel.

5.1.3 Credentialism

We also investigate whether the program acts differently on men and women through a credentialism mechanism, via which individuals with a diploma from PRONATEC may have an advantage when applying for jobs. Table 7 documents the results of our ITT estimates on the probability of having an certificate from PRONATEC (Column (c)). The increase on the probabilities of the offer group are similar between women and men (panel B and C: 28.3 p.p.s) and the effect is statistically significant for randomization inference p -values (0.025 and 0.000, respectively). However, considering heteroskedasticity-robust standard errors, the effect is statistically significant only for the sub-sample of men (0.160 and 0.002, respectively). Therefore, this is unlikely to be a channel through which labor market effects arise.

5.1.4 Gateway

Another possible mechanism through which TVET may impact labor market outcomes is the gateway effect. Institutions offering PRONATEC may act as a channel for job opportunities with students being referred to job openings by the school or school instructors. Conditional on being employed, we asked in the survey if the PRONATEC provider was responsible for connecting them to their current job opportunity. Column (d) of Table 7 shows that there are no statistically significant effects of the scholarship offer on the probability of having a job indication, so this is unlikely to be a relevant mechanism.

5.1.5 Psychological Effect

Finally, we investigate whether the gender heterogeneity in our labor market results could have been channeled by a “psychological effect” triggered by the program’s waiting list randomization design. Since there was oversubscription to the program, students who won the lottery could potentially feel special and empowered for having been selected for a desired spot, even knowing that the spots were randomly assigned. We took advantage of another dataset to estimate the effect of a TVET program with no waiting list selection criteria to be able to provide a comparison with our experimental results. We assume that this psychological effect would be associated with a sense of empowerment which could have a positive impact on labor market. Thus, if the psychological effect is relevant, our results for the waiting list sample should be

greater than those that did not have a waiting list, using the same ordinary least squares (OLS) specification in both estimations. Since we did not have enough observations to split up these new dataset in sub-samples of women and men, we ran the OLS specification for the full sample including a variable which is an iteration of treatment and gender variables.

For this additional dataset, we also interviewed students attending vocational education in the integrated modality, which is provided by the state level high schools (EEBs) in the state of Santa Catarina. This modality integrates in the same curriculum general and vocational education. However, in this case, there is no excess demand for these courses, as students choose to attend either general or vocational education when enrolling for these high schools. For each EEB randomly selected, we compared labor market outcomes of the students who graduated from the vocational track with the ones from the general education graduated students (similar in observable characteristics). Table 10 reports the results of our OLS estimation. Column (a) shows the results for the sample of vocational education provided by EEBs, in which the treatment is defined as choosing the vocational education track. Column (b) shows the results for the PRONATEC experimental sample, including course-by-site strata dummies, in which the treatment is defined as receiving an offer to the scholarship. We find no evidence corroborating the hypothesis of the existence of a “psychological effect”.

Table 10 OLS Estimates of the Correlation between the Treatment and Labor Market Outcomes

Dep. Var.:	(a)	(b)
	Employment	Employment
	EEB	PRONATEC
Treatment	0.101 [0.077]	-0.007 [0.046]
<i>p</i> -value	0.189	0.883
Female	-0.228 [0.094]	-0.225 [0.092]
<i>p</i> -value	0.016	0.015
Age	-0.003 [0.017]	0.001 [0.003]
<i>p</i> -value	0.860	0.644
Treatment*Female	0.183 [0.118]	0.158 [0.096]
<i>p</i> -value	0.123	0.100
Constant	0.829 [0.345]	0.780 [0.086]
<i>p</i> -value	0.017	0.000
Number of observations	204	735

Notes: This table considers correlation between attending a vocational course with no waiting list design (column (a)) or receiving an scholarship offer with a waiting list design (column (b)) and labor market outcomes. Specification in column (b) includes course-by-site strata dummies. Coefficients, standard errors and *p*-values are computed using an ordinary least squares regression with heteroskedasticity-robust standard errors [in brackets], using an indicator of being employed in the endline sample as the dependent variable. All specifications include control variables for sex and age.

5.2 Structural Equation Model (SEM)

In this section, we develop a structural model to investigate further the relative importance of non-cognitive skills to explain our findings. We assume the following causal channels:

- (i) The scholarship offer has a direct effect on PRONATEC enrollment (equation (2));
- (ii) Enrollment on PRONATEC has positive effect in both PRONATEC graduation (equation (3)), and on non-cognitive skills (equation (4));
- (iii) Both PRONATEC graduation and non-cognitive skills impact labor market outcomes directly (equation (5)).

We interpret the effect of graduation on labor market outcomes, conditioned on non-cognitive skills, as the direct joint effect of job specific skills, credentialism and gateway effects. Thus, we can isolate the non-cognitive effect from all the other direct effects. This structural model is represented by the following system of linear equations:

$$Enrolled = \alpha_0 + \alpha_1 Offer + \alpha_2 Age + \alpha_3 Agem + X' \gamma + \varepsilon \quad (2)$$

$$Graduated = \beta_0 + \beta_1 Enrolled + \beta_2 Age + \beta_3 Agem + \beta_4 Ageit + X' \pi + \mu \quad (3)$$

$$NCIndex = \gamma_0 + \gamma_1 Enrolled + \gamma_2 Age + \gamma_3 Agem + \gamma_4 Ageit + \gamma_5 Agem^2 + X' \lambda + \nu \quad (4)$$

$$Employed = \theta_0 + \theta_1 Graduated + \theta_2 NCIndex + \theta_3 Age + \theta_4 Agem + X' \omega + \zeta \quad (5)$$

where *Enrolled* denotes the probability of enrollment in a PRONATEC course, *Graduated* denotes the probability of graduation in a PRONATEC course, *NCIndex* denotes the non-cognitive skill index and *Employed* denotes the probability of being employed. *Age* is the individuals' ages at the moment of the survey, *Agem* is the individuals' ages at the moment of the offer to participate of the program, *Ageit* is an interaction of the *Age* and *Agem* variables, *Agem*² is the squared *Agem* variable and *X* is a vector of individuals' covariates (sex, color and mother's education). ε , μ , ν , ν and ζ are error terms.

This system of linear equations impose the following exclusions restrictions:

- (i) The lottery is an instrument for PRONATEC enrollment and does not affect any other outcome directly;
- (ii) Research on the technology of skill formation indicates that acquirement of skills are heterogeneous over the life cycle [Cunha et al., 2010]. Also, non-cognitive skills may be more malleable and less subject to "sensitive periods of learning". In this sense, non-cognitive skills development by vocational education may be correlated with the age at which students enrolled in the course. Thus, we assume that the interaction of age at the moment of the offer and the current age affects graduation and non-cognitive skills directly. Moreover, we assume *Agem*² as an additional instrument for non-cognitive skills only.

We estimate the parameters of the equation system (2)-(5) in two stages. First, we estimate the reduced form through the SUR model with errors that are correlated across equations for a

given individual, uncorrelated across individuals, with zero mean and homoskedastic. Second, we identify the estimators of structural parameters and obtain their standard errors through the delta method.

Table 11 shows the results of the estimation. Standard Errors are reported in brackets. The direct effect of graduating on a TVET course on labor market outcomes is given by θ_1 . The direct effect of non-cognitive skills is given by θ_2 . The effect of PRONATEC enrollment on labor market outcome is given by $\partial Employed / \partial Enrolled = (\theta_1 * \beta_1) + (\theta_2 * \gamma_1)$. The first term is the effect of enrollment through graduation (joint effects of all other channels) and the second term is the effect of enrollment through non-cognitive skills.

Table 11, column (b) shows the results for women. The total impact of PRONATEC enrollment on the probability of being employed is 55.6 p.p.s. Its impact through graduation is 12.9 p.p.s and through non-cognitive skills is 42.6 p.p.s. Thus, our findings suggest that the impact of PRONATEC on women employment is mostly driven by the non-cognitive channel.

Table 11 Structural Equation Model Parameters

Dep. Var.:	(a)	(b)	(c)
	All	Women	Men
α_1	0.801 [0.062]	0.669 [0.125]	0.875 [0.071]
p -value	0.000	0.000	0.000
β_1	0.627 [0.086]	0.800 [0.173]	0.573 [0.100]
p -value	0.000	0.000	0.000
γ_1	0.144 [0.102]	0.964 [0.178]	-0.022 [0.124]
p -value	0.160	0.000	0.862
θ_1	0.043 [0.039]	0.034 [0.076]	-0.119 [0.045]
p -value	0.266	0.652	0.007
θ_2	0.176 [0.003]	0.477 [0.010]	0.022 [0.003]
p -value	0.000	0.000	0.000
$\theta_1 * \beta_1$	0.029	0.129	-0.086
$\theta_2 * \gamma_1$	0.034	0.426	0.003
$\theta_1 * \beta_1 + \theta_2 * \gamma_1$	0.063	0.556	-0.082
Number of observations	364	107	257

Notes: This table considers the impact of the offer on labor market outcomes through a structural equations model framework in the full sample of surveyed applicants (column (a)) and the sub-sample of female and male surveyed applicants (columns (b) and (c), respectively). The estimates are from simultaneous regressions of dummies for being enrolled and for graduating in PRONATEC, non-cognitive index and dummies for being employed. Controls for color, age, gender, mother education level and course-by-site strata dummies. Standard errors computed by delta method are shown [in brackets]. Two-tailed p -values for the null hypothesis of a zero average parameter $H_0 : \tau_{ITT} = 0$ are displayed in the bottom rows each parameter.

Table 12 Attrition (answered non-cognitive module), All Inter-viewed Subscribers and by Gender Group

	(a)	(b)	(c)
	All	Women	Men
Answered			
Offer	-0.015 [0.050]	0.099 [0.090]	-0.058 [0.062]
p-value (w. strata)	0.766	0.273	0.354
Number of observations	735	237	498

Notes: This table considers attrition in the full group of subscribers, and conditional on gender. Coefficients, standard errors and *p*-values are computed using specification (1), an ordinary least squares regression with course-by-site strata dummies and heteroskedasticity-robust standard errors [in brackets], using an indicator of not answering the non-cognitive module as the dependent variable. Specification included control variables for sex, age and color.

Table 13 Within-Strata Balance Across Enrollment Arms, All Interviewed Subscribers who received an offer

	(a)	(b)	(c)	(d)	(e)
	Mean (No Enrollment)	Mean (Enrollment)	<i>p</i> -value (w. strata)	% Size	Number of obs.
Male	0.622	0.687	0.631	10.4%	609
Black, "pardo" or indigenous	0.299	0.251	0.518	-16.1%	609
Age	23.559	23.867	0.461	1.3%	609
Mother completed high-school	0.270	0.255	0.910	-5.4%	609
Mother completed tertiary	0.087	0.091	0.477	5.4%	609
Father completed high-school	0.222	0.253	0.347	13.9%	609
Father completed tertiary	0.055	0.062	0.507	12.9%	609
Concomitant modality	0.457	0.477	0.0126	4.5%	609
N	127	482			
	<i>F-stat</i>	<i>p-value</i>			
	1.240	0.278			

Notes: This table considers balancedness in the group of applicants who received an scholarship offer surveyed at endline. Column (a) (column (b)) shows means and standard deviations (in parenthesis) for a pre-determined variable, for those not enrolled (enrolled) in the program with the scholarships in the schools and time frame considered. *p*-values in column (c) are computed for each variable using specification (1), an ordinary least squares regression with course-by-site strata dummies and heteroskedasticity-robust standard errors. Column (d) displays the unadjusted percent size of the raw difference between enrollment arm means with respect to the control group mean, and column (e) shows the sample size. The last row shows the *F*-statistic and *p*-value from a regression of the treatment indicator on the full set of pre-determined variables and course-by-site strata dummies. Additional details on how the outcomes were created can be found in (see section "Definition of Main Variables" in Appendix).

6 Conclusion

The ways through which TVET programs may affect the trajectories of youth are multiple. Investing in occupational skills that are demanded in the labor market can improve formal employability and work earnings in a context in which these specific skills are seen by firms as requirements that are not fulfilled in general education schools. TVET may also provide

the setting where non-cognitive skills such as organization, hard-working ability and self-confidence are prone to be developed.

We find that these dimensions were differentially affected among men and women in the setting of a scholarship program that targets current and former high school students of the public educational system and grants them the opportunity to attend TVET courses free of charge. The program offer has increased enrollment and graduation for both men and women. Women experienced large gains in labor market outcomes (employment, labor market participation, formality and earnings) and non-cognitive skills (conscientiousness and, particularly, extraversion). The results for men suggest null effects on most of the dimensions we analyze.

The heterogeneity we find on labor market outcomes has been pointed out in systematic reviews about technical and vocational education in developing countries [Blattman and Ralston, 2015]. We bring novel information of heterogeneous effects on non-cognitive skills that are arguably valued in the labor market. In this paper, we also perform exercises to explore potential mechanisms through which the labor market effects for women arise. Our findings suggest that gender differences on labor market outcomes could have emerged through the enhancement of non-cognitive skills. These results rely on a number of hypothesis, and must be interpreted with caution. Therefore, future work in this area is crucial and additional research could shed light on the interactions among the different outcome dimensions.

Declarations

Availability of data and material

The datasets generated and/or analysed during the current study are not publicly available due to ethical concerns. It is a unique dataset generated from the primary data collection which contains identification of participants at an individual level. But dataset are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The Coordination for the Improvement of Higher Level -or Education- Personnel (CAPES) and the Inter-American Development Bank (IDB) provided financial support in different stages of the research.

Authors' contributions

JC coordinated the primary data collection, analyzed and interpreted the data regarding the labor market and non-cognitive skills outcomes. FR performed the quality assurance of the estimations and was a major contributor in writing the manuscript. LL was a major contributor in writing the manuscript. APS was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This manuscript is a revised version of the main chapters of Camargo's Ph.D. thesis at São Paulo School of Economics. She would like to thank CAPES for financial support in different stages of the research, and the Inter-American Development Bank, SENAI-SC and the Secretaria Estadual de Educação de Santa Catarina for the collaboration that made this impact evaluation possible. The authors would like to thank the Ayrton Senna Institute for kindly providing their validated instruments for non-cognitive skills, Daniel Santos and Ricardo Primi for providing data support, and seminar participants at São Paulo School of Economics, the 2017 Brazilian Econometric Society meeting, the 2018 European Economic Association meeting and the 2018 Latin America and Caribbean Economic Association meeting for helpful comments and suggestions that substantially improved the earlier versions of this manuscript.

References

- Abadie, A.; Athey, S.; Imbens, G. W.; and Wooldridge, J. (2017): When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.
- Almeida, A. M. F.; Giovine, M. A.; Alves, M. T. G.; and Ziegler, S. (2017): Private education in argentina and brazil. *Educação e Pesquisa*, 43(4):939–956.
- Almeida, R.; Amaral, N.; and de Felício, F. (2015): *Assessing Advances and Challenges in Technical Education in Brazil*. World Bank Publications.
- Almeida, R.; Anazawa, L.; Menezes-Filho, N.; and Vasconcellos, L. (2014): Retornos da educação profissional e técnica no brasil. *The World Bank*.
- Almlund, M.; Duckworth, A. L.; Heckman, J.; and Kautz, T. (2011): Personality psychology and economics. In *Handbook of the Economics of Education*, volume 4, pages 1–181. Elsevier.
- Alzúa, M. L.; Cruces, G.; and Erazo, C. L. (2013): Youth training programs beyond employment. evidence from a randomized controlled trial. *La Plata, Buenos Aires: CEDLAS.(Mimeo.)*
- Amorim, V.; Arabage, A.; Camargo, J.; Lima, L.; Lucena, T.; Ripani, L.; Soares, S.; and Souza, A. (2015): A stocktake of vocational education and training cycle in brazil - from school entry to labor market transition. Technical report, Inter-American Development Bank (IDB) and Getulio Vargas Foundation.
- Assunção, J. and Gonzaga, G. (2010): Educação profissional no brasil: inserção e retorno. *SENAI. Série*.
- Assunção, J. and Gonzaga, G. (2015): Educação profissional no brasil: inserção e retorno. Technical report, Serviço Nacional de Aprendizagem Industrial (SENAI). Unidade de Prospectiva do Trabalho (UNITRAB).
- Attanasio, O.; Guarn, A.; Medina, C.; and Meghir, C. (2017): Vocational training for disadvantaged youth in colombia: A long-term follow-up. *American Economic Journal: Applied Economics*, 9(2):131–143.
- Attanasio, O.; Kugler, A.; and Meghir, C. (2011): Subsidizing vocational training for disadvantaged youth in colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 3(3): 188–220.
- Barrick, M. R. and Mount, M. K. (1991): The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26.
- Bergemann, A. and Van den Berg, G. J. (2008): Active labor market policy effects for women in europe—a survey. *Annales d'Economie et de Statistique*, pages 385–408.
- Bertrand, M.; Crépon, B.; Corsetto, L.; Fajardo-Gonzalez, J.; Macias, C.; Schickele, A.; and Shammout, N. (2017): Skills training programs. Technical report, J-PAL Abdul Latif Jameel Poverty Action Lab.
- Bishop, J. H. and Mane, F. (2005): Raising academic standards and vocational concentrators: Are they better off or worse off? *Education Economics*, 13(2):171–187.
- Blattman, C. and Ralston, L. (2015): Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs. Technical report, SSRN: <https://ssrn.com/abstract=2622220>.
- Bowles, S.; Gintis, H.; and Osborne, M. (2001): Incentive-enhancing preferences: Personality, behavior, and earnings. *The American Economic Review*, 91(2):155–158.
- Bruhn, M. and McKenzie, D. (2009): In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4):200–232.
- Caliendo, M. and Künn, S. (2015): Getting back into the labor market: The effects of start-up subsidies for unemployed females. *Journal of Population Economics*, 28(4):1005–1043.
- Card, D.; Ibararán, P.; Regalia, F.; Rosas-Shady, D.; and Soares, Y. (2011): The labor market impacts of youth training in the dominican republic. *Journal of Labor Economics*, 29(2):267–300.
- Card, D.; Kluve, J.; and Weber, A. (2010): Active labour market policy evaluations: A meta-analysis. *The economic journal*, 120(548).
- Card, D.; Kluve, J.; and Weber, A. (2015): What works? a meta analysis of recent active labor market program evaluations. Technical report, National Bureau of Economic Research.
- Card, D. (1999): The causal effect of education on earnings. *Handbook of labor economics*, 3:1801–1863.
- Cattan, S. (2010): Heterogeneity and selection in the labor market. *PhD thesis, University of Chicago*.
- Costa, G. L. M. (2013): O ensino médio no brasil: desafios à matrícula e ao trabalho docente. *Revista brasileira de estudos pedagógicos*, 94(236).
- Cunha, F.; Heckman, J. J.; and Schennach, S. M. (2010): Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.
- Dearden, L.; McIntosh, S.; Myck, M.; and Vignoles, A. (2002): The returns to academic and vocational qualifications in britain. *Bulletin of economic research*, 54(3):249–274.
- FGV/CLEAR (2017): New skills at work: Educação para o trabalho como motor para o desenvolvimento e inclusão. Technical report, FGV EESP Clear.
- Hanushek, E. A.; Schwerdt, G.; Woessmann, L.; and Zhang, L. (2017): General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, 52(1):48–87.

- Heckman, J. J. (2007): The economics, technology, and neuroscience of human capability formation. *Proceedings of the national Academy of Sciences*, 104(33):13250–13255.
- Hirshleifer, S.; McKenzie, D.; Almeida, R.; and Ridao-Cano, C. (2016): The impact of vocational training for the unemployed: experimental evidence from turkey. *The Economic Journal*, 126(597):2115–2146.
- Hogan, J. and Holland, B. (2003): Using theory to evaluate personality and job-performance relations: a socioanalytic perspective.
- Ibarrarán, P.; Kluge, J.; Ripani, L.; and Rosas, D. (2015): Experimental evidence on the long-term impacts of a youth training program. *IZA Working Paper*.
- Imbens, G. W. and Rubin, D. B. (2015): *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Jenkins, A.; Greenwood, C.; and Vignoles, A. (2007): *The returns to qualifications in England: updating the evidence base on level 2 and level 3 vocational qualifications*. Centre for the Economics of Education, London School of Economics and Political Science.
- John, O. P. and Srivastava, S. (1999): The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138.
- Kane, T. J. and Rouse, C. E. (1995): Comment on w. norton grubb: "the varied economic returns to postsecondary education: New evidence from the class of 1972". *Journal of Human Resources*, pages 205–221.
- Kautz, T.; Heckman, J. J.; Diris, R.; Ter Weel, B.; and Borghans, L. (2014): Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, National Bureau of Economic Research.
- Kugler, A.; Kugler, M.; Saavedra, J.; and Prada, L. O. H. (2015): Long-term direct and spillover effects of job training: experimental evidence from colombia. Technical report, National Bureau of Economic Research.
- Machin, S. and Vignoles, A. (2005): *What's the Good of Education?: The Economics of Education in the UK*. Princeton University Press.
- Malamud, O. and Pop-Eleches, C. (2010): General education versus vocational training: Evidence from an economy in transition. *The review of economics and statistics*, 92(1):43–60.
- Neri, M. C. (2010): A educação profissional e você no mercado de trabalho. *Rio de Janeiro: FGV/CPS*.
- Neuman, S. and Ziderman, A. (1999): Vocational education in israel: wage effects of the voced-occupation match. *Journal of human resources*, pages 407–420.
- Nyhus, E. K. and Pons, E. (2005): The effects of personality on earnings. *Journal of Economic Psychology*, 26(3):363–384.
- Oliva, B. T. (2014): *Três ensaios de economia da educação*. PhD thesis.
- Salgado, J. F. (1997): The five factor model of personality and job performance in the european community.
- Santos, D. and Primi, R. (2014): Social and emotional development and school learning: a measurement proposal in support of public policy. *São Paulo, Brazil: Ayrton Senna Institute*.
- Tansel, A. (1998): General versus vocational high schools and labor market outcomes in turkey. In *Economic Research Forum Working Paper*, number 9905.
- Vasconcellos, L.; Lima, F. C.; Fernandes, J. G.; and Menezes-Filho, N. (2010): Avaliação econômica do ensino médio profissional. *Fundação Itaú Social, Brasília*.
- Woessmann, L. (2008): Efficiency and equity of european education and training policies. *International Tax and Public Finance*, 15(2):199–230.

Appendix A. Definition of Main Variables

All information used in this analysis was originally collected for the purpose of evaluating the PRONATEC program in Chapecó, Itapiranga, São Miguel do Oeste and Xanxerê. In what follows we consider how the main variables used in the analysis were defined after data collection.

A.1 Education, Technical and General

- *Enrolled (PRONATEC)*: Indicator variable that takes the value of one if the person reported having enrolled in a PRONATEC course;
- *Enrolled (any)*: Indicator variable that takes the value of one if the person reported having enrolled in a PRONATEC course or any other technical education program;
- *Graduated (PRONATEC)*: Indicator variable that takes the value of one if the person reported having graduated in a PRONATEC course;
- *Graduated (any)*: Indicator variable that takes the value of one if the person reported having graduated in a PRONATEC course or any other technical education program;
- *Graduated (High-school, concomitant)*: Indicator variable that takes the value of one if the person reported having finished high-school — it is only defined for the concomitant modality, which is the only one allowed for individuals enrolled in school;
- *Entered (Tertiary)*: Indicator variable that takes the value of one if the person reported having started tertiary at some point.

A.2 Labor Market

- *Employment*: Indicator variable that takes the value of one if the person reported to have a job in the month before the survey or zero if the person reports being unemployed or out of the labor force;
- *Labor market participation*: Indicator variable that takes the value of one if the person reported either to have a job or to be looking for a job in the month before the survey or zero if the person reports being out of the labor force;
- *Formal employment*: Indicator variable that takes the value of one if the individual claims to have a signed work card (carteira assinada) in the current most important employment” we imputed zeros for all individuals who report being either unemployed or out of the labor force in the month before the survey;
- *Days per month*: Censored continuous variable defined as the number of days worked in an usual week in the current most important job, multiplied by four — we imputed zeros for all individuals who report being unemployed or out of the labor force for more than one month at the survey date;
- *Hours per week*: Censored continuous variable defined as the number of hours worked in an usual day in the current most important job, multiplied by five — we imputed zeros for all individuals who report being unemployed or out of the labor force for more than one month at the survey date;
- *Tenure*: Censored continuous variable defined as the length of duration of employment in months in the current most important job — we imputed zeros for all individuals who

report being unemployed or out of the labor force for more than one month at the survey date;

- *Work earnings*: Monthly wage or earnings in the in the current most important employment” we imputed zeros for all individuals who report being either unemployed or out of the labor force in the month before the survey. The variable in logs is computed as the natural logarithm of the latter variable.
- *Worked in course area*: Indicator variable that takes the value of one if the person reported to have a job in the course area for which she applied in the month before the survey or zero if the person reports being unemployed or out of the labor force.

A.3 LATE Estimates

Table A1 LATE Estimates on Labor Market Outcomes

Dep. Var.:	(a)	(b)	(c)	(d)	(e)	(f)
	Employment	Labor market participation	Tenure	Formal employment	Work earnings	Wages (in logs)
Panel A: All						
Offer	0.056 [0.053]	0.068 [0.043]	5.460 [5.613]	0.066 [0.064]	132.700 [136.9]	0.058 [0.061]
No offer group mean	0.802	0.865	33.008	0.690	1399.135	7.345
p -value ($H_0 : \tau_{ITT} = 0$)	0.291	0.112	0.331	0.304	0.332	0.344
Number of observations	735	735	726	735	731	567
Panel B: Women						
Offer	0.295 [0.120]	0.356 [0.104]	22.31 [8.549]	0.481 [0.129]	697.3 [197.2]	0.309 [0.131]
No offer group mean	0.658	0.711	24.711	0.474	873.026	7.055
p -value ($H_0 : \tau_{ITT} = 0$)	0.014	0.001	0.009	0.000	0.000	0.018
Number of observations	237	237	235	237	237	166
Panel C: Men						
Offer	-0.021 [0.055]	-0.033 [0.041]	-0.661 [6.758]	-0.083 [0.068]	-48.630 [172.3]	-0.008 [0.069]
No offer group mean	0.864	0.932	36.632	0.784	1626.318	7.445
p -value ($H_0 : \tau_{ITT} = 0$)	0.708	0.424	0.922	0.225	0.778	0.905
Number of observations	498	498	491	498	494	401

Notes: This table considers the impact of participation on labor market outcomes in the full sample of surveyed applicants (Panel A) and the sub-sample of female and male surveyed applicants (Panels B and C, respectively). All variables are self-reported and were collected from June to August of 2016 (an average of 2 years after the end of the program in our sample). Estimates are from two-stage least squares regressions with course-by-site strata dummies, and standard errors computed using heteroskedasticity-robust standard errors [in brackets]. Additional information about the samples in each panel (no offer arm mean, p -values and sample size) can be found in the bottom rows and details on how the outcomes were created can be found in Appendix 7.

Table A2 LATE Estimates on Standardized Indicators of Non-cognitive Skills

Dep. Var.:	(a)	(b)	(c)	(d)	(e)	(f)	(g)
	NC Index	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness to experiences	Locus of control
Panel A: All							
Offer	0.187 [0.172]	0.132 [0.177]	0.250 [0.169]	0.253 [0.187]	0.057 [0.177]	-0.022 [0.189]	-0.071 [0.185]
p -value ($H_0: \tau_{ITT} = 0$)	0.277	0.458	0.139	0.176	0.749	0.909	0.703
Number of observations	376	376	376	376	376	376	376
Panel B: Women							
Offer	0.697 [0.240]	0.590 [0.277]	0.548 [0.294]	0.856 [0.334]	0.279 [0.327]	0.444 [0.350]	-0.468 [0.344]
No offer group mean	-0.316	-0.171	-0.038	-0.378	-0.513	-0.368	0.235
p -value ($H_0: \tau_{ITT} = 0$)	0.004	0.033	0.062	0.011	0.394	0.205	0.174
Number of observations	111	111	111	111	111	111	111
Panel C: Men							
Offer	0.060 [0.207]	0.007 [0.208]	0.176 [0.197]	0.108 [0.205]	0.000 [0.210]	-0.085 [0.210]	0.047 [0.212]
No offer group mean	0.110	0.056	0.012	0.123	0.167	0.120	-0.077
p -value ($H_0: \tau_{ITT} = 0$)	0.771	0.975	0.371	0.598	1.000	0.684	0.827
Number of observations	265	265	265	265	265	265	265

Notes: This table considers the impact of participation on non-cognitive skills outcomes in the full sample of surveyed applicants (Panel A) and the sub-sample of female and male surveyed applicants (Panels B and C, respectively). All variables are self-reported and were collected from June to August of 2016 (an average of 2 years after the end of the program in our sample). Estimates are from two-stage least squares regressions with course-by-site strata dummies, and standard errors computed using heteroskedasticity-robust standard errors [in brackets]. Additional information about the samples in each panel (no offer arm mean, p -values and sample size) can be found in the bottom rows and details on how the outcomes were created can be found in Appendix 7.

Table A3 LATE Estimates on Possible Channels Outcomes

Dep. Var.:	Possible Channels			
	(a)	(b)	(c)	(d)
	NC Index	Occupation (course area)	Certificate	Gateway
Panel A: All				
Offer	0.187 [0.172]	0.217 [0.058]	0.334 [0.090]	-0.022 [0.034]
	0.277	0.000	0.000	0.515
No offer group mean		0.165	0.270	0.059
p -value ($H_0 : \tau_{ITT} = 0$)	0.296	0.000	0.000	0.529
Number of observations	376	571	587	569
Panel B: Women				
Offer	0.697 [0.240]	0.108 [0.050]	0.318 [0.205]	0.020 [0.028]
	0.004	0.032	0.121	0.465
No offer group mean	-0.316	0.043	0.250	0.000
p -value ($H_0 : \tau_{ITT} = 0$)	0.007	0.054	0.160	0.503
Number of observations	111	163	183	164
Panel C: Men				
Offer	0.060 [0.207]	0.247 [0.072]	0.32 [0.093]	-0.027 [0.042]
	0.771	0.001	0.001	0.517
No offer group mean	0.110	0.203	0.276	0.078
p -value ($H_0 : \tau_{ITT} = 0$)	0.783	0.001	0.002	0.534
Number of observations	265	408	404	405

Notes: This table considers the impact of participation on possible channels outcomes in the full sample of surveyed applicants (Panel A) and the sub-sample of female and male surveyed applicants (Panels B and C, respectively). All variables are self-reported and were collected from June to August of 2016 (an average of 2 years after the end of the program in our sample). Estimates are from two-stage least squares regressions with course-by-site strata dummies, and standard errors computed using heteroskedasticity-robust standard errors [in brackets]. Additional information about the samples in each panel (no offer arm mean, p -values and sample size) can be found in the bottom rows and details on how the outcomes were created can be found in Appendix 7.

A.4 Subscription information

Table A4 Number of subscriptions by municipality and course

Course	Chapecó	Itapiranga	S. M. Oeste	Xanxerê	Total
Mechanics	201	0	0	77	278
Administration	8	0	0	0	8
Pharmaceutical	0	0	5	0	5
Computer Network	163	49	30	38	280
Workplace Safety	223	0	28	42	293
Building	31	0	0	0	31
Electrotechnology	113	76	35	0	224
Electronics	46	0	0	0	46
Electromechanics	0	0	62	61	123
Food Technology	92	0	64	0	156
Total	877	125	224	218	1,444

Source: author's elaboration using survey and administrative data.