FUNDAÇÃO GETULIO VARGAS

ESCOLA DE ECONOMIA DE SÃO PAULO

JULIANA CAMARGO

ENSAIOS EM ECONOMIA DA EDUCAÇÃO

São Paulo

2017

FUNDAÇÃO GETULIO VARGAS

ESCOLA DE ECONOMIA DE SÃO PAULO

JULIANA CAMARGO

ENSAIOS EM ECONOMIA DA EDUCAÇÃO

Tese apresentada ao Programa de Pós-Graduação da Escola de Economia de São Paulo da Fundação Getulio Vargas, como requisito para a obtenção do título de Doutor em Economia.

Campo de Conhecimento: Microeconomia Aplicada - Economia da Educação

Orientador: Prof. Dr. André Portela Fernandes de Souza

São Paulo

2017

Camargo, Juliana.
Ensaios em economia da educação / Juliana Camargo. - 2017.
134 f.

Orientador: André Portela Fernandes de Souza
Tese (doutorado) - Escola de Economia de São Paulo.

1. Capital humano. 2. Ensino profissional. 3. Qualificações profissionais. 4. Ensino técnico. I. Souza, André Portela Fernandes de. II. Tese (doutorado) - Escola de Economia de São Paulo. III. Título.

CDU 337

JULIANA CAMARGO

ENSAIOS EM ECONOMIA DA EDUCAÇÃO

Tese apresentada ao Programa de Pós-Graduação da Escola de Economia de São Paulo da Fundação Getulio Vargas, como requisito para a obtenção do título de Doutor em Economia. Área de Conhecimento: Microeconomia Aplicada - Economia da Educação

Data de Aprovação: 22/05/2017

Banca examinadora:

Prof. Dr. André Portela Fernandes de Souza
(Orientador)
FGV-EESP

Prof. Dr. Bruno Ferman
FGV-EESP

Prof. Dr. Daniel Domingues dos Santos
USP-FEA/RP

Prof. Dr. Enlinson Henrique Carvalho de Mattos
FGV-EESP

Prof. Dr. Naercio Aquino Menezes Filho
Insper e FEA-USP

Aos meus pais, José e Sílvia.

AGRADECIMENTOS

Em primeiro lugar, agradeço aos meus pais, José e Sílvia Camargo, por todo amor, doação, apoio, confiança e por sempre me incentivarem em minhas decisões. Além disso, agradeço por sempre terem feito o possível (e muitas vezes mais do que o que era possível, mas mesmo assim o fizeram) para me oportunizar o poder transformador da educação, e por desde pequena me ensinarem a importância da mesma. Aos meus irmãos, Joseane e Rogner Camargo, por todo amor, apoio e companheirismo.

Agradeço ao meu namorado, Bernardo Dutra, pelo amor, por sempre me transmitir muito apoio, equilíbrio e calma, e por toda amizade e companheirismo desde quando nos conhecemos.

Aos meus amigos da EESP, que foram fundamentais para eu conseguir superar toda a pressão dos estudos e tornaram o doutorado muito mais alegre e leve. Em especial, agradeço ao Ronan Cunha, pela amizade especial que construímos, por dividir praticamente todos os momentos comigo e sempre estar ao meu lado, me apoiando. Sobretudo por me ensinar a importância da respiração. À Dalila Figueiredo, pela amizade, por sempre estar presente e disposta a me ajudar, e por sua doação e apoio (sinto muitas saudades de você). À Amanda Arabage, Bernardo Modenesi, Eduardo Cenci, Rafael Moualllen, Thiago de Lucena e Lucas Finamor.

Como não poderia deixar de ser, agradeço imensamente à irmã que minha alma escolheu, Lycia Lima, por sempre enxergar o melhor em mim, pelas palavras de incentivo, por todas as vezes que você "mudou minha vida" ao longo do doutorado, por toda a preocupação, por sempre me escutar e por todos os momentos incríveis que compartilhamos. Você é uma inspiração para mim.

Agradeço muito ao professor André Portela, pelo exemplo de paixão pela docência e pela pesquisa, pelo aceite de ser meu orientador, por toda a atenção, pelas valiosas discussões ao longo da tese e pelo seu otimismo nos momentos mais difíceis. Suas contribuições para esse trabalho são ímpares.

Aos professores da banca examinadora, Bruno Ferman, Enlinson Mattos, Daniel Santos e Naercio Menezes, pelas valiosas sugestões.

Ao professor e amigo Sabino Porto Júnior, por todo exemplo pessoal e profissional, por sempre acreditar em mim e me incentivar, pelo apoio, por todos os conselhos e "Toldos Amarelos" (que agora buscamos pelo Brasil).

Aos professores da EESP, por compartilharem seu conhecimento conosco.

À CAPES, ao BID, ao SENAI-SC e à Secretaria Estadual de Educação de Santa Catarina.

ABSTRACT

The aim of this study is to provide rigorous evidence of the impact of a brazilian vocational education and training program on its beneficiaries. For this, we took advantage of the selection criteria for the program, in which the beneficiaries were selected through a lottery . Besides, our unique dataset allow us to evaluate the effect of the intervention not only on the economic dimension, but also on human capital, socio-emotional, crime and risky behavior outcomes. This is a valuable contribution since there is a lack of evidence in the literature about the impact of VET programs on these outcomes. Our estimations were based on ITT and LATE strategies. Overall, our results indicated that, at least in the short-run, there was no effect of the program on conventional labor market outcomes, such as employment and formal work probability and wages. Also, we found no effect in socio-emotional, crime and risky behavior dimensions. However, our results pointed towards a possible heterogeneity of the program, since we found positive and significant effects of the intervention on women, specially in labor market (formal work probability and wage) and socio-emotional (extraversion) outcomes.

Keywords: Vocational Education; PRONATEC; Experimental Evidence; Human Capital; Socio-emotional Skills; Crime and Risky Behavior.

RESUMO

O objetivo deste estudo é fornecer evidências rigorosas do impacto de um programa brasileiro de educação técnica de nível médio sobre seus beneficiários. Para tanto, nós aproveitamos o critério de seleção do programa, no qual os beneficiários são selecionados através de um sorteio. Além disso, nosso banco de dados único nos permite avaliar o efeito da intervenção não apenas na dimensão econômica, mas também em dimensões como capital humano, habilidades socioemocionais e crime e comportamento de risco. Essa é uma contribuição relevante, uma vez que na literatura há uma lacuna de evidências sobre o impacto de educação técnica nesses resultados. Nossas estimações foram baseadas em estratégias de ITT e LATE. De forma geral, nossos resultados indicam que, ao menos no curto prazo, não há efeito do programa nos resultados de mercado de trabalho, tai como empregabilidade, emprego formal e salários. Ademais, nós também não encontramos efeitos do programa em outras dimensões como socio-emocionais, crime e comportamento de risco. Entretanto, nossos resultados sugerem que há uma heterogeneidade nos efeitos do programa, uma vez que encontramos efeitos positivos e significantes do programa sobre as mulheres, especialmente em resultados do mercado de trabalho (emprego formal e salário) e habilidades socioemocionais (extroversão).

Palavras-Chave: Educação Técnica; PRONATEC; Evidência Experimental; Capital Humano; Habilidades Socioemocionais; Crime e Comportamento de Risco.

List of Figures

List of Tables

Contents

1 Introduction

Education and its transformative potential has been gaining power in the public policy debate in recent years. The benefits of investments, enhancing years of education or improving educational quality, are well established in the human capital literature, which highlights their positive effects in economic growth[1]. Moreover, private and social benefits of education go beyond economic returns, also affecting other dimensions, such as crime (Machin, 2011, Lochner and Moretti, 2001, Machin et al., 2011) and health (Spear, 2000). Improving access and quality of education is one of the most widely implemented policies, particularly in developing countries. However, it is important to understand the technology by which people acquire skills over the life cycle, especially for policy makers, who usually face an efficiency-equity trade-off when deciding which educational policy they will provide to population (Carneiro et al, 2010).

Vocational Education and Training (VET) plays an important role in the development of workers' abilities. These are associated to individuals' characteristics, but also with their educational decisions - more years of education can develop important skills that will improve individuals' productivity and ability, and, as a consequence, their lifetime earnings in the labor market. This kind of education can prepare students for a smoother school-to-work transition, conciliating traditional education with work experience and specific vocations and tasks (Quintini and Martin, 2006, Eichhorst et al., 2015). VET is also a way to improve opportunities to disadvantaged youth who decided not attend higher education. It provides a career during secondary education, bringing them closer to the labor market and enhancing their network even before high school graduation (Harhoff and Kane, 1997, Lerman, 2012, Neuman and Ziderman, 1999). Furthermore, it is a way to increase intergenerational mobility, reducing the weight of parents' background on education choices (Shavit and Muller, 1998).

There are different ways of delivering vocational education, and countries have adopted different structures when providing VET courses. Specially for the vocational secondary courses, targeted to smooth school-to-work transition for young people, there are many different possibilities of course designs. As a consequence, several issues and questions regarding how these courses should be delivered on the ground arise. For example, should they focus on on-the-job training or on academic and more general contents? Are there alternative pedagogies that should be implemented, like using real world challenges to promote

[1] Hanushek and Kimko (2000), Hanushek (2006), Hanushek and Woessmann (2009), Barro (1991, 1997), and Mankiw, Romer and Weil (1992).

learning? What is the optimal course load? Should they also be concerned on developing non-cognitive skills? Should contracts be more flexible allowing teachers to have a career concurrently with their teaching profession? Should VET be seen as "dead end" courses, or as a way to enable students to access higher levels of education? Examples of two different vocational education designs are the ones adopted by The United States and Germany. The first argues that specific skills become obsolete too quickly and recently have redesigned his vocational courses in secondary school, in order to provide young people more general knowledge and prepare them to adapt to new technologies. In an opposite direction, in Germany, vocational education at the secondary level still remains focused on specific skills, and often adopts apprenticeships in its courses (Hanushek et al, 2011).

In developed countries, the returns of VET programs vary across countries, depending on the program design and on the institutional arrangements for VET supply. There is evidence for positive impacts of VET programs on labor market outcomes, such as wages (Dearden et al., 2002, Bishop and Mane, 2004, Neuman and Ziderman, 1999), and employment probability (Bishop and Mane, 2004; Hanushek, 2011); although, there is also evidence of low or nonexistent economic returns for some kind of VET programs (Wößmann, 2008, Machin and Vignoles, 2005, Jenkins, Greenwood and Vignoles, 2007, Dearden et al., 2004, 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, 2008, Tansel, 1998). For Brazil in particular, some studies find positive impacts of VET on labor market variables (Almeida et al., 2014, Vasconsellos et al., 2010, Assunção and Gonzaga, 2010, Oliva, 2014, Neri, 2010, Biondi, 2015).

Regarding the technologies of skills formation, VET programs may have more potential for enhancing non-cognitive skills, since they are more malleable than cognitive skills over the life-cycle, including in adulthood. Even if it is hard to improve non-cognitive skills through vocational education, this investment may still be interesting since there is also the intergenerational channel: enhancing non-cognitive skills of young people and adults may have effect on early cognitive and non-cognitive skills of their children (Carneiro et al., 2010, Carneiro and Heckman, 2003, Cunha and Heckman, 2009, Carneiro, 2009, Woessmann, 2008). There is also evidence of a positive relationship between non-cognitive skills and academic achievement, good health outcomes, social and economic development (Almlund et al., 2011, Friedman and Kern, 2014, Kautz et al., 2014, OECD, 2015, Poropat, 2009) and labor market outcomes (Kuhn and Weinberger, 2005, Bowles, Gintis and Osborne, 2001, Nyhus and Pons, 2005, Hogan and Holland, 2003, Salgado, 1997, Barrick and Mount, 1991, Cattan, 2010). Specifically in the concomitant modality[2], where students face heavier course load,

[2] Concomitant is the modality which technical education can be attended concurrently with high school

it is not clear what is the direction of the impact of VET on general education outcomes - academic achievement, attendance rate and graduation at correct age, for example. On one hand, it could enhance cognitives skills through the knowledge acquired in the course, on the other hand, it could increase the opportunity cost, in terms of time and effort, of attending vocational education, damaging their outcomes on general education. In this study, we use experimental evidence to shed light on the impact of vet on the following outcomes for students in concomitant modality: probability to graduate at correct age and attendance rate in regular education, academic achievement in Math and Portuguese in regular education.

Moreover, the benefits of Vocational and Technical Education may go further the economic dimension. For example, there is a well established literature on the negative relationship between education, crime and risky behavior (Sabates, 2008; 2009, Sabates and Feinstein, 2008) [3]. The four main channels through which schooling might mitigate criminal participation according to this literature are: (i) income effects, (ii) time availability, (iii) patience or risk aversion, and (iv) social networks or peers of individuals. Education enhances the opportunity cost of illegal behavior, since it increases wage rates, and, consequently, the opportunity costs of crime (Lochner, 2004, Lochner and Moretti, 2001, Hjalmarsson, 2008). Besides, time spent in education may also reduce time available for participating in criminal activity (Tauchen et. al, 1994, Jacob and Lefgren, 2003, Luallen, 2006, Anderson, 2010), which could be an important channel for youth enrolled in concomitant modality of vocational education. Education may change individuals preferences, modifying their patience or risk aversion. More patient and risk-averse individuals would give more value to the possibility of future punishments (Lochner and Moretti, 2001, Oreopoulus, 2007). Finally, young people are highly influenced by their peers and environment (Chowdry et al, 2009, Center for Mental Health in Schools at UCLA, 2007), and school may provide an environment through which teenagers enhance their networks and increase their relationship with peers, who could influence their probability of not engaging in crime and risky behavior.

The main VET program carried out in Brazil was launched in 2011 by the Federal Government, and it has become one of the Brazilian largest VET programs of all time: Programa Nacional de Acesso ao Ensino Técnico e Emprego – PRONATEC (National Program of Access to Professional Education and Employment). The aim of the program is to expand access to VET in Brazil, promoting labor market opportunities to the population (Souza et al., 2015). Despite this recent effort, the number of students who decide to follow a technical track in secondary education is still a small portion of the total high school enrollments in Brazil. In 2015, the number of students enrolled in general education exclusively was around

[3] For studies in economic field, see Machin (2011), Lochner and Moretti (2001), Machin et al. (2011), Grogger (2000), Hjalmarsson (2008), Machin and Meghir (2004), Lochner (2004), Tauchen et. al (1994), Jacob and Lefgren (2003), Luallen (2006), Anderson (2010).

6.4 million, whereas the number of students enrolled in Technical education was 1.7 million - including subsequent[4] modality (INEP, 2016). When analyzed over time, however, it is clear that technical courses are attracting more students in recent years, comparing to general secondary education. In 2007, 95.2 % of the Brazilian secondary students were enrolled in general education, and 4.8 % of them were enrolled in vocational education; while in 2015, 8.7 % of the secondary students were enrolled in vocational education (considering only concomitant and integrated modalities). However, these numbers are still far from those in similar Latin American countries. The ratio of young students enrolled in vocational education in Colombia and Mexico is 28 % and 38 %, respectively. In comparison to some developed countries, the discrepancy is even higher: the share of students enrolled on vocational education in Italy is 56 %, in Austria is 71%, in Switzerland is 62%, and the OECD average is 40% (OECD, 2007 and 2016).

One of the main challenges when evaluating the impact of VET courses on labor market and other outcomes is to isolate the effects exclusively assigned to the abilities developed by the program from those that are inherent to the individual. There are unobservable characteristics that can influence individuals' decisions and generate selection bias on impact estimation. To deal with endogeneity and estimate the causal effect of PRONATEC[5], this paper explores its random assignment [6] identified in some circumstances. Furthermore, we constructed a unique dataset (using primary data collected through a survey and administrative data) to evaluate the impact of "Bolsa Formação Estudante", a voucher-type scholarship of PRONATEC, targeted to low-income students, on labor market and human capital outcomes; crime and risky behavior; and non-cognitive skills.

In this study, we contribute to the literature by providing rigorous evidence of the impact of VET programs in Brazil. We take advantage of the randomization used to select beneficiaries to the PRONATEC Program to identify its impact. Besides, our unique dataset allows us to investigate the impact of technical education on the development of non-cognitive skills, crime and risky behavior, for which there is a lack of evidence in the literature. The first challenge of the study is that this experimental design is a randomization with waiting list (Abdulkadiroglu et al., 2011, Behaghel et al., 2015, Curto & Fryer, 2014, Chaisemartin

[4] Students enrolled in the subsequent modality have already graduated secondary education, thus being only enrolled in the technical component.

[5] We are looking to a particular modality of PRONATEC, named Bolsa Formação Estudante, which is provided by a private institution, called S-System (a group of nonprofit entities that carry out private activities of public interest). Considering professional and technical education supply, the two most important entities of the S-System are SENAI (industry's national learning service) and SENAC (commerce's national learning service). From now on, we are going to refer to this specific design of the program as PRONATEC.

[6] Providers can choose their selection criteria when there is excess of demand for the program. Particularly, S-System in Santa Catarina, choose to select their beneficiaries through a lottery.

and Behaghel, 2015). To address this problem, we applied the technique proposed by Chaisemartin and Behaghel (2015), which (for each class) classify into treatment group individuals who received a random number strictly lower than the random one received by the last student who had an offer to enroll in the course. Individuals who were ordered by a random number strictly greater are assigned to control group. The authors show that in expectation both groups contain the same proportion of accepters, making them statistically comparable.

Our second challenge is the fact that our randomization is pooling several lotteries, since the randomization level was the class. Chaisemartin and Behaghel (2015) show that their results still hold when several lotteries are conducted. However, to ensure the balance of the proportion of accepters between both groups, we must use inverse probability of being invited to enroll in the course reweighting[7]. Moreover, we had a high rate of attrition (65%), largely due to missing or outdated telephone numbers. Thus, we split our administrative data in two groups: administrative data from the lottery ("Administrative 1" - which contains only, for each class, individual's name and order in the lottery - without any additional information) and administrative data with individual's information ("Administrative 2" - from S-System and Secretariat of Education of the State of Santa Catarina).

Comparing "Administrative 1" data with our sample, we found that attrition is asymmetric across treatment and control groups, while comparing "Administrative 2" data with our sample, attrition is not significantly correlated with treatment. First, as our balance checks show that there are no statistically significant differences between treatment and control groups, we make an assumption that as both groups are likely to be well balanced in observed variables, they are likely to be well balanced also in unobserved variables as well. In order to test the robustness of our estimations, we performed hypothesis tests using randomization inference[8]. Results from randomization tests for our ITT coefficients corroborated the ones we found using conventional ones. Notwithstanding, attrition can bias our estimations if unobserved characteristics are correlated with sample attrition and the outcomes analyzed in this study. For these reason, in a second exercise, we relax the hypotheses that both groups are likely to be well balanced in unobserved characteristics. We try to deal with attrition applying the nonparametric method from Lee (2009), which construct extreme bounds for the intervention effects when there are significant differences in attrition rates between treatment and control groups. Unfortunately, it leads to wide nonparametric bounds on treatment impacts, which are not very informative.

Finally, our study must consider that the program compliance was not perfect. Only

[7] Rosenbaum Rubin (1983), Frölich (2007).

[8] Fisher (1935), Young (2016), Hayes (2000). The distribution of randomization tests does not depend of asymptotic theorems or distributional assumptions. Randomization satisfies the exchangeability assumption of the test: that is, the treatment allocation can be permuted across observations.

a fraction of the individuals took up the treatment. In order to avoid the reintroduction of selection bias, we must compare all individuals initially assigned to the treatment group to all those allocated to the comparison group, regardless of their actual treatment status (Duflo et al., 2007). With imperfect compliance, the randomization generates an instrument Z_i for the treatment of interest T_i. Thus, our main estimates are an ITT, and, when using the lottery as an instrument, a LATE.

Our main results, considering our whole sample, indicate that, at least in the short-run, there is no effect of PRONATEC on labor market outcomes, such as employment and formal work probability and wages. Furthermore, we found no effect of the program on socio-emotional, crime and risky behavior outcomes. However, our results point towards a possible heterogeneity on the impact of the program, since we find positive and significant effects of PRONATEC considering only the subsample of women, specially in labor market (formal work probability and wages) and socio-emotional (extraversion) outcomes.

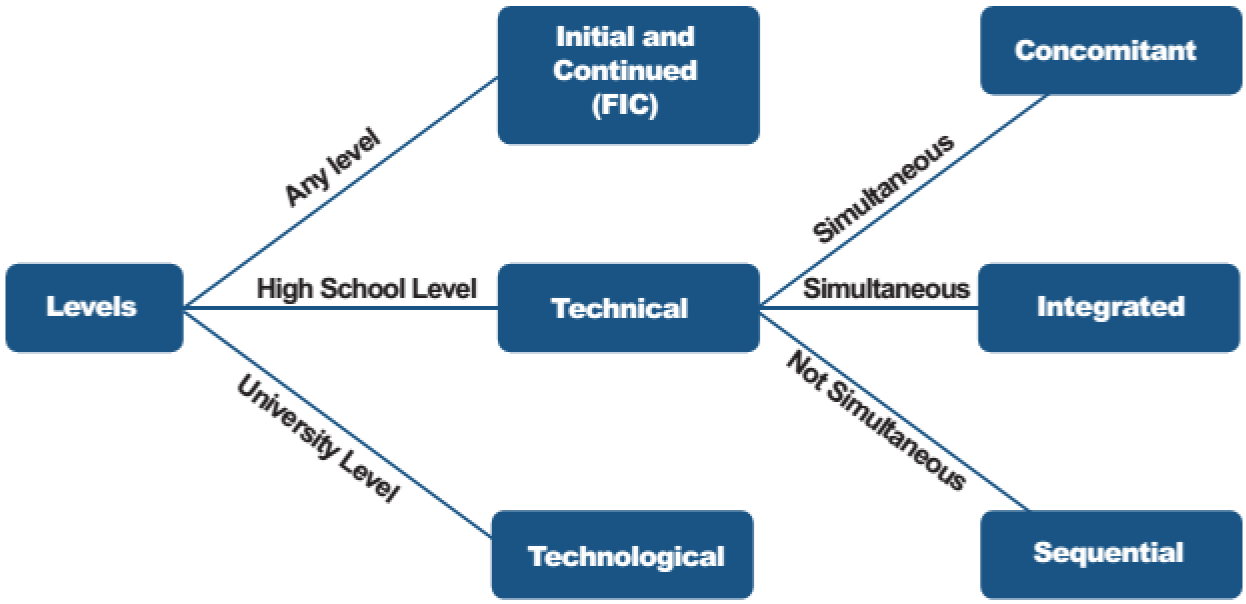
This study is organized as follows: section 2 briefly describes the PRONATEC Program. Section 3 contains the program experimental design, and presents the data. Section 4 describes our empirical strategy. Sections 5, 6 and 7 contain a briefly review of the related literature and the results for labor market and human capital, non-cognitive skills, crime and risky behavior outcomes, respectively. Section 8 presents the conclusion.

2 PRONATEC Program

Programa Nacional de Acesso ao Ensino Técnico e Emprego – PRONATEC (National Program of Access to Professional Education and Employment) began in Brazil in 2011 and became one of the largest VET programs. The main objective of the program is to increase the access to VET and foster labor market opportunities to the population. The Federal Government designed several actions to achieve this target, increasing the number of institutions providing VET courses, the number of courses and vacancies available and subsidizing technical education for vulnerable students (Souza et al., 2015). There are three different levels of VET courses offered by PRONATEC: i) initial and Continued (FIC), which can be coursed with any level of education; ii) technical education, which requires primary school level and is equivalent to an high school degree; iii) and technological education, which requires high school education to be coursed and is equivalent to university degree.

Figure 1 shows the different levels of the Vocational System, the educational degrees granted by each one and its different modalities. Technical education can be attended concurrently with high school, in concomitant and integrated modalities, or after graduating secondary education, in subsequent modality. Students enrolled in the integrated modality study both general and vocational education at the same institution. The ones attending the concomitant modality also study both general and vocational education, but each component is studied in different schools. Students enrolled in the subsequent modality have already graduated secondary education, thus being only enrolled in the technical component.

Figure 1 – Levels of Vocational System in Brazil



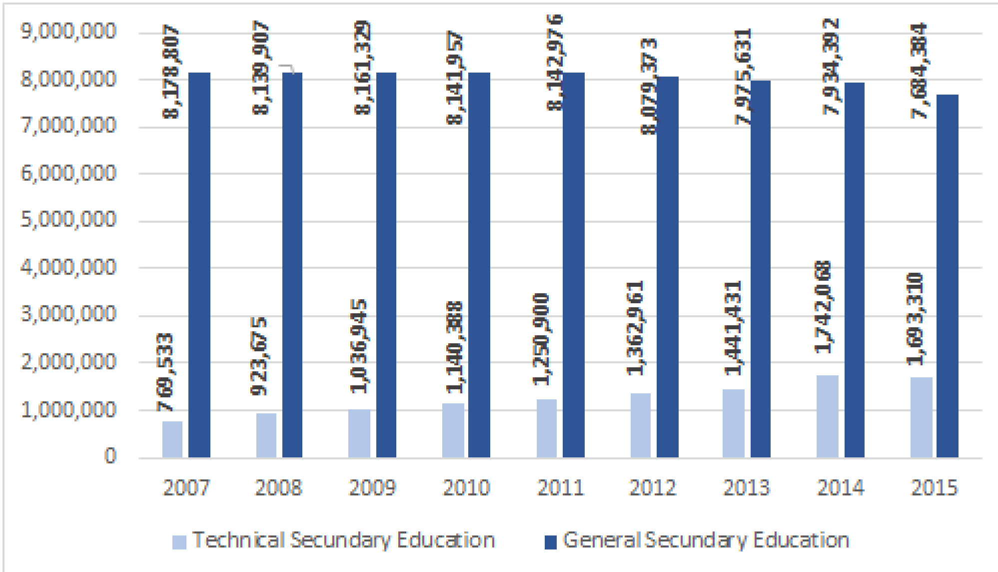
Source: Souza et al., 2015

In 2014, year in which the program reached its maximum number of enrollments, PRONATEC offered 646 types of initial and continued courses and 227 technical courses across 4,300 municipalities, and SENAI was responsible for providing 360 FIC courses and 63 technical courses across 2,237 municipalities. Moreover, SENAI was responsible for 41 % of the whole enrollments of PRONATEC - considering all kind of courses covered by the program - (SENAI, SESI and IEL- Relatório Anual 2014, MEC, 2014). The extension of FIC courses are at least 160 hours, while technical course loads are at least 600 hours. Despite this recent VET scale-up, technical education still corresponds to a small percentage of total number of enrollments in high school level in Brazil. In 2015, 7.7 million students attended general education exclusively, whereas 1.7 million students were enrolled in technical education (including subsequent modality).

Particularly in the state of Santa Catarina, where we collected information from former students of technical education, there was, in 2015, 230.7 thousand students enrolled exclusively on general education, while there was 57.6 thousand students enrolled on vocational education - including subsequent modality (INEP, 2016). However, when analyzed over time, it is clear that the number of enrollments on vocational education is proportionally increasing more than the number of enrollments on general education (Figure 2). And this

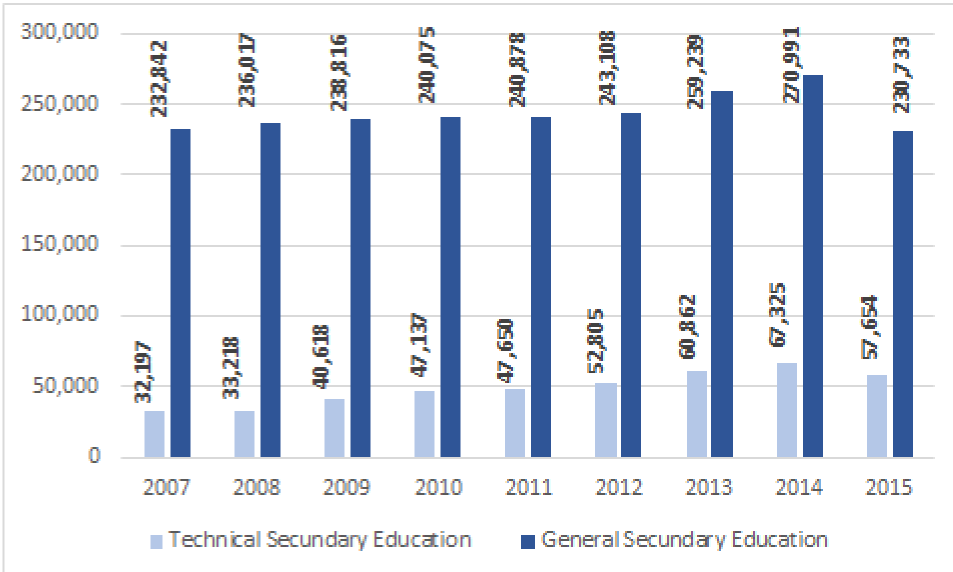
path is also verified at Santa Catarina (Figure 3).

Figure 2 – Technical VS General Secondary Education in Brazil



Source: INEP/MEC. Technical Secondary Education covers integrated, concomitant and subsequent modality. General Secondary Education is the sum of enrollments in general secondary education and teaching degree (magistério). It does not include enrollments in integrated modality of technical education, special education (educação especial) and youth and adult education (EJA)

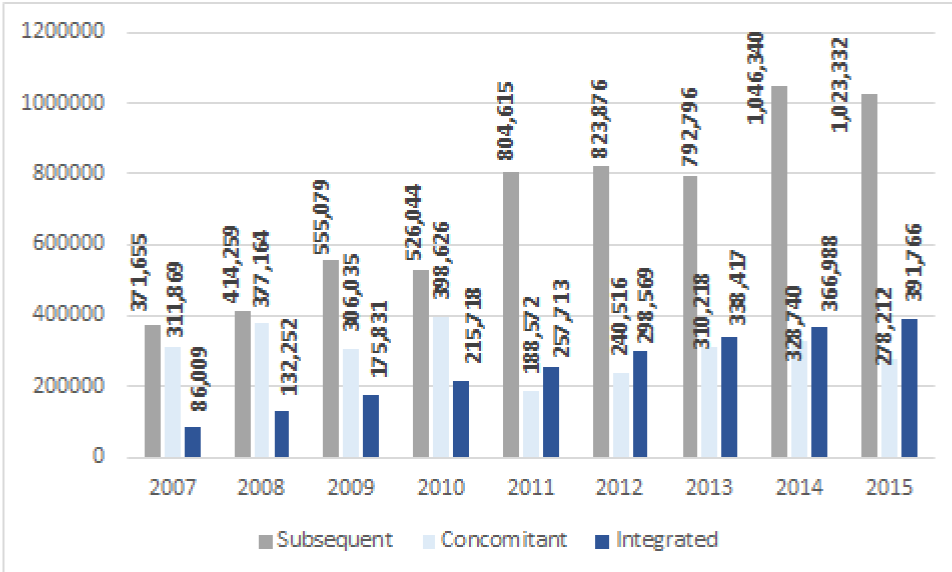
Figure 3 – Technical VS General Secondary Education in Santa Catarina



Source: INEP/MEC. Technical Secondary Education covers integrated, concomitant and subsequent modality. General Secondary Education is the sum of enrollments in general secondary education and teaching degree (magistério). It does not include enrollments in integrated modality of technical education, special education (educação especial) and youth and adult education (EJA)

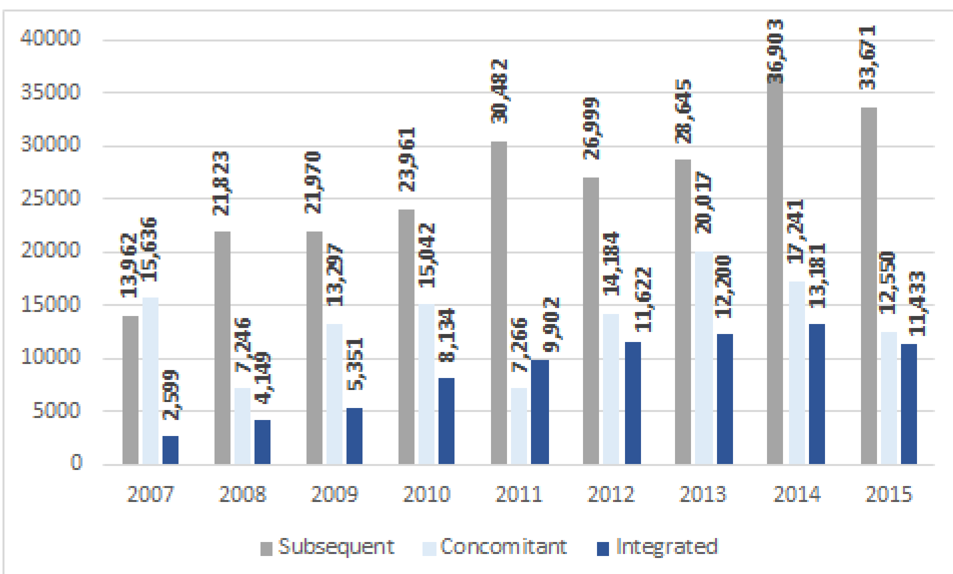
The composition of enrollments in Brazil, according to course modality, also has changed in recent years. In Figure 4 it is clear that integrated modality have proportionally increased the number of enrollments more than concomitant and subsequent ones, followed by subsequent modality, which is still the greatest modality in absolute number of enrollments. In 2007, integrated was the least demanded course modality, when compared to subsequent and concomitant ones. Meanwhile, this pattern has changed over the years: since 2011 it has more students enrolled than concomitant modality. In Santa Catarina, according to Figure 5, we note the same pattern: subsequent modality is also the most demanded, despite enrollments in integrated modality proportionally increased more than in other ones. Nevertheless, in contrast to Brazil, it still has less students than the other two modalities.

Figure 4 – Technical Education in Brazil, by course modality



Source: INEP/MEC.

Figure 5 – Technical Education in Santa Catarina, by course modality



Source: INEP/MEC.

PRONATEC is composed of six smaller programs: i) Bolsa Formação (Training Scholarship); ii) FIES Técnico (Technical FIES); iii) E-Tec Network; iv) S-System Agreement; v) Brasil Profissionalizado (Professionalized Brazil); and vi) Expansion of the Federal Network (Souza et al., 2015). This study focuses on Bolsa Formação program, a voucher-

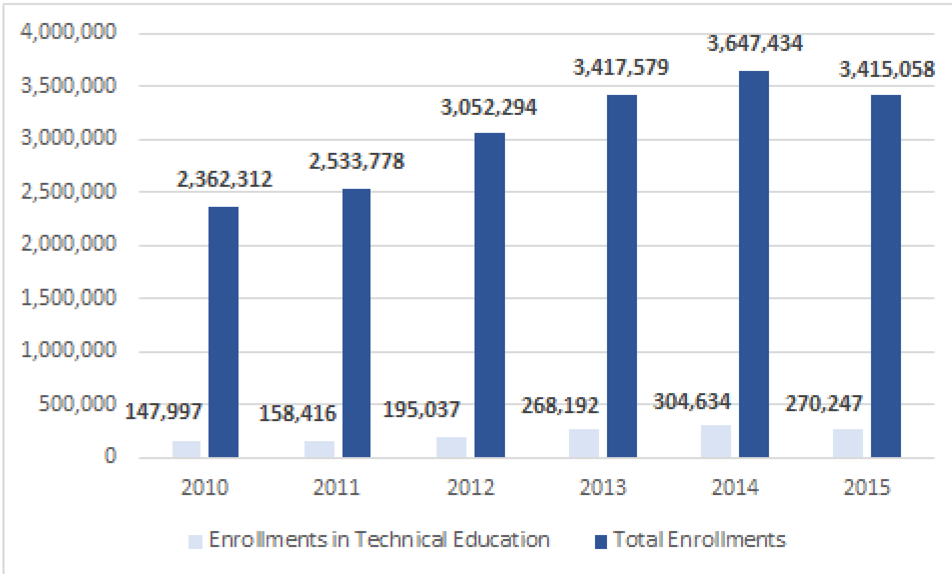
type scholarship and one of the most important PRONATEC initiatives. Particularly, we investigate the modality of the program named Bolsa Formação Estudante - targeted to low-income students enrolled in secondary school from public schools. Bolsa Formação can be provided by public or private institutions - by federal and state schools, or by National Service for Apprenticeship (S-System) [1], and courses are free of charge. These possibility of different providers is aligned with an important question about the most effective ways of delivering the program, whether by expanding the public network or whether by boosting partnerships with private institutions that already have the infrastructure and expertise on providing technical courses.

In addition, their private administration enables a wider flexibility to adjust the quality of technical courses offered in order to meet labor market demand. In this study, we aim to provide evidences of the effects of at least one of this program designs: Bolsa Formação Estudante provided by S-System institutions (National Service of Industrial Training (SENAI, Serviço Nacional de Aprendizagem Industrial) and National Service for Commercial Training (SENAC, Serviço Nacional de Aprendizagem Comercial) - the two most important providers of technical education among S-System). The main reason for this choice was their selection criteria, through which we can identify rigorously the impact of this specific intervention. The majority of the courses in our study was provided by SENAI, only two courses were from SENAC.

The S-System was created in the 1940s in order to improve the quantity and quality of the workforce, meeting the labor market demand. Today, it is the most important private provider of technical education in Brazil, specially through SENAI (the greatest provider among all S-System institutions), and the quality of its courses is widely recognized by students and by employers. SENAI, in 2014, was responsible for 41 % of the whole enrollments on PRONATEC (taking into account all several different modalities covered by the program), and was considered the most important private provider of technical education in Latin America (SENAI, SESI and IEL- Relatório Anual, 2014). Figure 6 shows the number of enrollments at SENAI over past years. For each year, first column represents the number of enrollments only in technical education, while second column accounted for total number of enrollments at SENAI - regardless the kind of course offered and not only those provided to PRONATEC (for example, short-duration, technological, initial and continuing training courses). It seems that SENAI is an important player on VET provision: in 2015, for example, it was responsible for about 16% of the enrollments on technical education in Brazil.

[1] We are following National Institute of Studies and Research (INEP), who classified S-System as a private institution

Figure 6 – Enrollments at SENAI



Source: SENAI, SESI and IEL- Relatório Anual 2012 - 2015

VET courses are offered by S-System in concomitant and subsequent modalities; and are open to all individuals, since they meet the educational degree requirements: students must have completed at least the first year of secondary school. Moreover, to obtain a technical degree, a student must successfully complete his studies on general secondary education and on technical education. Even though more complete than a general school degree, from the educational system perspective these two are equivalent, as both allow you to access tertiary education. The course costs vary according to their area and state. Students who were awarded with Bolsa Formação can attend classes at no cost, since it is sponsored by the government.

As a consequence, there is a large number of students interested in subscribing to technical education, and this design of the program usually faces excess of demand. In this case, institutions can choose which eligibility and selection criteria to apply. They can, for example, select candidates through an entrance exam; or using age, low-income, or quotas for students from public schools as a priority criteria; they can also use ordering/randomization criteria. S-System institutions from Santa Catarina chose to select their candidates through a randomization criteria. Thus, for each class of technical course offered, when the number of candidates was bigger than the number of vacancies available, the candidates were randomly selected through a lottery. This random assignment of the vacancies allows us to rigorously identify the causal effect of the program on several outcomes, for example: labor market, human capital, risky behavior and non-cognitive skills.

3 Experimental Design and Data

Technical courses offered by Bolsa Formação Estudante have at least 800 classroom hours. The ones from our study have on average 1200 hours of course load and their course duration is 2 years.

The study was conducted in four municipalities in the state of Santa Catarina, Brazil: Chapecó, Itapiranga, São Miguel do Oeste and Xanxerê. 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 where selected (and of the ones who were not), in each class, for these four specific municipalities in the state. We conducted a survey and collected primary data in order to construct a unique database, interviewing individuals who were and the ones who were not awarded with PRONATEC scholarship for the years of 2012 (Chapecó), 2013 (all the four municipalities) and 2014 (Chapecó) at SENAI and SENAC. The survey was composed by several outcome dimensions, including socioeconomic condition, labor market outcomes, self reported crime and risky behavior participation. We also applied a non-cognitive instrument to take a measure of individuals' socio-emotional skills.

One important issue about the best way of delivering technical education is finding a balance between academic and vocational contents. It is important to not enhance the cost of vocational education for youth, in terms of time and effort, due to its extensive course load. To investigate this possible channel, in addition to the primary data collected, administrative data were used to identify, for students in concomitant modality, their academic achievement and attendance rate at general education school. We also verified if they graduated in general education at correct age.

3.1 Administrative Data

There are two main sources of data to this study: administrative and primary data. The last one was collected on a survey and is detailed in the next section. Administrative data, which contains individual level baseline information, comes from SENAI and Secretariat of Education of the State of Santa Catarina database. They provide baseline variables like age, sex, education degree and color. Moreover, it is possible, through official documents of the lottery, to identify the random number assigned to each individual, which is his invitation

to enroll in the course ordering. From this random number we constructed the variable, Z_i, which identifies students' treatment status, after the second round of enrollment, at "Administrative 1" data. At the end of these classification, we had 2007 individuals, 1475 assigned to treatment and 532 to control group. Merging individuals' names with their telephone number, available at "Administrative 2" data (from SENAI and Secretariat of Education of the State of Santa Catarina), allowed us to get in touch with most program subscribers [1]. Considering only "Administrative 2" data, we had 1419 individuals, 1,186 from treatment group and 233 from control group. However, there were some missing or outdated contacts on "Administrative 2" registries, and this was our second source of attrition in the study. The third source of attrition was individuals who we could contact but were not available, or did not want to participate on the survey.

For all individuals whose administrative information in baseline covariates was found, we carried out balance checks. We also made these tests using time invariant covariate variables from our survey dataset, in order to evaluate the balance of the sample collected. Results of these last tests are presented in the next section. Furthermore, for those individuals who were enrolled on concomitant modality, administrative data form Secretariat of Education system also provided information about their academic achievement at secondary general education, and about their attendance rate and graduation at correct age in general education. These data are used to investigate if there is a channel through which VET can damage students, enhancing their cost of attending vocational education in terms of effort and time. We estimate the impact of PRONATEC on students' academic achievement, and on their general education graduation at correct age.

First, we show descriptive statistics from "Administrative 2" data, in order to briefly display information about the profile of students who looked for vocational education supplied by the intervention. Overall, we can note in Table 1 that vocational education courses were more demanded by men (only 32% of all subscribers were women), and white[2]. Considering course modality, it seems that demand is well balanced between concomitant and subsequent modality (the number of vacancies were according to the class, regardless the modality - both concomitant and subsequent students attended the same class). Moreover, subscribers are currently on average 24 years old. Analyzing separately for each course modality, concomitant has greater proportion of women enrolled than subsequent courses, 40% versus 24%. The percentage of individuals declared as white is quite similar in both modalities. Students enrolled for concomitant courses, as expected, are younger - they are current on average 21

[1] Individuals with no information at "Administrative 2" data were our first source of attrition

[2] This variable is not very precise on administrative data, because for 37% of the individuals it was reported as "not informed" or "color not declared", it seems that this information was not well filled on administrative data, only 16% of this information was reported as other color than white

years old, while students enrolled on subsequent modality are current 28 years old.

Table 1 – Descriptive Statistics of Administrative Data with Individuals' Information, by course modality

Variables	All Subscribers		Concomitant		Subsequent	
	N	Mean	N	Mean	N	Mean
Female	1419	0.32	720	0.40	699	0.24
White	1419	0.47	720	0.45	699	0.48
Secondary School (incomplete)	1419	0.51	-	-	-	-
Age	1419	24.42	720	21.13	699	27.81

Source: author's elaboration using administrative data.

3.2 Survey Information and Descriptive Statistics of the Sample

We conducted a survey from June to August of 2016 using a team of undergraduate students surveyors. They were trained and supervised by the authors of this study. 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 large number of candidates who we could contact agreed to participate of the study. The main source of attrition was outdated telephone numbers, not participation agreement. All interviews, including the ones made by telephone, were supervised in person by the authors.

The survey was an individual level questionnaire. It included questions of demographic and socioeconomic characteristics (age, sex, color, education degree, educational level of the parents and household condition). It also contained a labor market module (wage, labor market status, formal work status, area of work); and a self reported crime and risky behavior participation. Our measure of crime in this study is the self-reported engagement in some illegal activities, such as participation in an argument or fight, sale of pirate goods, sale of drugs, sale of stolen goods (all these questions were considering time period of the last four years) and drink and drive (which could also be classified as a risky behavior). Our self-reported measures of risky behavior are: use of alcohol, cigarette, marijuana, and other drugs; use of alcohol more than twice a week, and binge drinking (5 or more alcoholic drinks for males and 4 or more alcoholic drinks for females on the same occasion)[3].

We also applied a non-cognitive instrument to measure individuals non-cognitive skills. It is an instrument for the measurement of social and emotional skills, developed by

[3] According to the definition from US National Institute on Alcohol Abuse and Alcoholism.

Ayrton Senna Institute (a non-profit organization focused on improving the quality of education), and named SENNA (Social and Emotional or Non-cognitive Nationwide Assessment). SENNA was designed to be a simple instrument, and one robust enough to be applied in large samples. At the same time, it was designed to be precise and interpretable enough, in order to be used scientifically in studies (Santos and Primi, 2014, p.29).

One important issue on evaluating non-cognitive skills is identifying which ones are relevant and the best way to measure them. One of the most applied theoretical model is the Big Five personality traits, also known as the five factor model (FFM), which groups personality traits in five basic dimensions: agreeableness, conscientiousness, extraversion, neuroticism and openness to experience. "The “Big Five” are latent constructs obtained by factor analysis performed on the answers to extensive questionnaires with varied questions about behaviors that are representative of all the personality characteristics that an individual could have" (Santos and Primi, 2014, p. 16). Agreeableness is defined as the tendency to be cooperative and unselfish. A conscientiousness person is 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)" (Santos and Primi, 2014, p. 19). Extraversion is the characteristic of being friendly, sociable, self-confident and enthusiastic. Neuroticism is defined as the consistence of emotional reactions. An unstable individual is defined as short-tempered, introspective and impulsive. Finally, an open to experience individual is characterized as someone open to new aesthetic, cultural and intellectual experiences. He is often imaginative, curious and unconventional (Santos and Primi, 2014). SENNA provides us latent measures of agreeableness, conscientiousness, extraversion, neuroticism, openness to experience and locus of control[4].

We first present descriptive statistics of the sample, using survey data. Summary statistics are quite similar to the ones from administrative data. Our sample has 718 individuals, 594 assigned to treatment and 124 to control group. It is possible to see in Table 2 that individuals are on average 24 years old, most of them are men and white. Considering family background, it seems that most of them have a disadvantage background, the highest education from either mother and father is incomplete primary school, 38% and 41%, respectively. Only 8% had mother with tertiary school degree, and considering the father this number is even smaller 4%. Splitting by course modality, students profile have some differences: concomitant students are younger, on average 19 years old, most of them are men, white and with parents who studied until secondary school; subsequent students are older, on average 29 years old, their proportion of male students is even higher, 71%, white

4 Locus of Control is defined as a measure of how much "individuals attribute current experiences either to decisions and attitudes they have taken in the past" (Santos and Primi, 2014, p.22)

and most of their parents have not even completed primary school.

Table 2 – Descriptive Statistics of the Sample

Variables	All Subscribers		Concomitant		Subsequent	
	N	Mean	N	Mean	N	Mean
Age	718	24.42	318	19.11	400	28.64
Male	718	0.68	318	0.63	400	0.71
White	718	0.73	318	0.72	400	0.74
Other Colors	718	0.27	318	0.28	400	0.26
Mother's Highest Grade						
Primary School (inc.)	718	0.38	318	0.25	400	0.49
Primary School (comp.)	718	0.23	318	0.21	400	0.24
Secondary School	718	0.28	318	0.40	400	0.19
Tertiary School	718	0.08	318	0.13	400	0.05
Not Informed	718	0.03	318	0.02	400	0.03
Father's Highest Grade						
Primary School (inc.)	718	0.41	318	0.28	400	0.51
Primary School (comp.)	718	0.22	318	0.22	400	0.21
Secondary School	718	0.26	318	0.35	400	0.20
Tertiary School	718	0.04	318	0.07	400	0.02
Not Informed	718	0.07	318	0.07	400	0.07

Source: author's elaboration using survey data.

3.3 Classification of individuals in Treatment and Control Groups

One specific characteristic of the design of this program is that participating is not mandatory. It means that individuals awarded in the lottery may not take part of it. In this case, according to the Federal Guideline for "Bolsa Formação" management (Ministry of Education, 2011, p. 31), providers can call students in the waiting list to fulfill these remaining vacancies, as long as they follow its order of classification. Municipalities chosen for this study decided to randomly select beneficiaries for the program, as well as for the waiting list. Individuals were classified according to the order through which they were awarded in the lottery. The ones classified from 1 to 35 were assigned to the beneficiary group and the ones from 36 and over were assigned to the control group. Individuals from control group still could participate of the program whether the seats were not completely filled. If an individual assigned to beneficiary group decide to not participate of it, or if suppliers could not contact him in able time, they can call the first individual assigned to the control group, that is, the 36th awarded in the lottery, and so on.

Therefore, for each class, until the program fill all vacancies, the control group classification works as a waiting list. This process is called by PRONATEC as first and second round of enrollment. In other words, for each class, n students applied to enter in a specific technical course where $v < n$ vacancies are available. We know that 35 seats are available for each class, thus $v = 35$. Each student received a random number $C_i \in [1, n]$. Individuals with $C_i \leq v$ received an offer to attend classes. If all $1 \leq C_i \leq 35$ individuals accepted the invitation, no other potential student received an offer. If d students denied, students with $v < C_i \leq v + d$ received an offer, and so on until all v vacancies were filled. Using administrative data from SENAI, we could identify, for each class, the last individual who received an offer, defined as τ. Thus, define $W_i = 1$ individuals with $C_i \leq \tau$ and $W_i = 0$ individuals with $C_i \leq \tau$ Moreover, treatment and control groups may not be statistically comparable, because the group of $W_i = 1$ individuals may have a greater proportion of accepter than the $W_i = 0$ one (Chaisemartin and Behaghel, 2015). To address this problem, we applied the technique proposed by Chaisemartin and Behaghel (2015) through which the new variable is constructed as follows: $Z_i = 1\{C_i < \tau\} - 1\{C_i = \tau\}$. This slight modification implies that in expectation both groups contain the same proportion of accepters, making them statistically comparable.

Table 3 shows the distribution of individuals assigned to Treatment and Control groups at first and second round of enrollment in our sample (due to the adjustment required by the instrument, our final sample size is of 718 individuals). First and second columns present the distribution of individuals assigned to control and treatment groups at second round, respectively. 124 individuals belong to control group in both rounds, that is, they have never been called to participate of the program in any of the rounds. 443 individuals were assigned to treatment group at first round and have been called to participate of the program. This way, they have been assigned to treatment group in both rounds. 151 students were assigned to control group at first round, but have been called to the program in order to fill remaining vacancies, so they were assigned to treatment group at second round. Thus, after second round, total number of individuals in treatment group is 594 and in control group is 124.

Table 3 – Distribution of Treatment and Control groups at First and Second Round of Enrollment

		Second Round		
		Control	Treatment	Total
First Round	Control	124	151	275
	Treatment	0	443	443
	Total	124	594	718

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

3.4 Attrition

There are 3 sources of attrition in our study, as displayed in tables 4 and 5. From the 2007 individuals listed at "Administrative 1" data, we could find baseline information and, the most important, telephone numbers for only 1,419 individuals: 1,186 were assigned to beneficiary group and 233 to control group.

The first one explains the difference of observations between our database "Administrative 1" and our final sample. This is due to the fact that PRONATECs official system only records information for individuals called in the first round, and in the second round up to the number of vacancies existing in the first round. So, for a course with 35 vacancies, the system only recorded information for the 35 students called in the first round, and up to 35 in the second round. The information for all remaining students in the second round (those that randomly received a ranking number above 70) was not recorded in the system. Therefore, the first source of attrition is related to this guidelines form the government and, hence, is not related to unobservable characteristics of the individuals. The second source of attrition is due to outdated phone numbers. The third source of attrition is explained by those students who were successfully contacted but refused to participate in the survey. It is important to notice that sources 2 and 3 of attrition are not correlated with treatment, as displayed in Table 8.

Table 4 – Number of successful interviews considering all subscribers from administrative data, by treatment status

	Treatment		Control	
	N	%	N	%
Number of interviews	594	40.27%	124	23.31%
Total number of subscribers	1,475		532	

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

Table 5 displays attrition rates considering "Administrative 2" data (subscribers we had administrative data with phone information but we could not contact or interview), by treatment status, for all subscribers, separating it according to course modality: concomitant and subsequent and for female and male - for all subscribers, concomitant and subsequent.

Table 5 – Attrition rates considering administrative data with phone information, by treatment status

	Treatment		Control	
	N	%	N	%
All subscribers	437	58.89	305	41.11
Concomitant	114	59.38	78	40.63
Subsequent	323	58.73	227	41.27
Male (all subscribers)	264	56.05	207	43.95
Female (all subscribers)	173	63.84	98	36.16
Male (concomitant)	58	53.7	50	46.3
Female (concomitant)	56	66.67	28	33.33
Male (subsequent)	206	56.75	157	43.25
Female (subsequent)	117	62.57	70	37.43

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

For each municipality, each class has his own lottery, thus, it is important to show how many individuals were interviewed in each course. Table 6 shows the number of individuals (both, treatment and control group) that were interviewed, according to their municipality and the course they were enrolled in.

Table 6 – Number of interviews by municipality and course

Course	Chapecó	Itapiranga	S. M. Oeste	Xanxerê	Total
Mechanics	88	0	0	42	130
Administration	12	0	0	0	12
Pharmaceutical	0	0	6	0	6
Computer Network	66	0	0	0	66
Workplace Safety	128	0	7	26	161
Building	19	0	0	0	19
Electrotechnology	52	41	21	0	114
Electronics	15	0	0	0	15
Informatics/Comp. Maintenance	0	28	13	18	59
Electromechanics	0	0	27	43	70
Food Technology	31	0	35	0	66
Total	411	69	109	129	718

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

The numbers presented on Tables 4 and 5 show that we must pay attention to a common challenge existing in almost all empirical evaluations: attrition. According to the literature, "random attrition will only reduce a study's statistical power; however, attrition

that is correlated with the treatment being evaluated may bias estimates" (Duflo et al., 2007, p. 58). Moreover, the independence of potential outcomes between the treatment and control groups, once ensured by the randomization, does not hold if the attrition is systematically correlated with the treatment.

As we had a large rate of attrition in our study, specially due to the difficulty of following up subscribers years later they have made their inscription in the program, it is important to verify if it was correlated with the treatment or not. There was many outdated registers, and this was the main reason why we were not able to contact all students subscribed to participate of the program. However, almost all candidates we could contact agreed in participating of the study. Therefore, our main source of attrition was outdated telephone numbers, not participation consent. We imagine that the asymmetric distribution of attrition, considering "Administrative 1" data are likely to be associated with the fact that administrative employees did not fill information for individual who received a random number too far from 35 (for example, 95), and were not likely to receive an invitation to enroll at the course.

We ran OLS estimations for "Administrative 1" (Table 7), and "Administrative 2" data (Table 8) in order to investigate if the attrition is correlated with the treatment. Dependent variable is an indicator of attrition, a dummy variable that assume value equal to one if we interviewed student i, and equal to zero otherwise. Our treatment is a dummy variable that assume value equal to one if individual i is assigned to treatment group, and equal to zero otherwise.

Finally, our randomization is pooling several lotteries [5], since the randomization was made for each class. Chaisemartin and Behaghel (2015) show that their results still hold when several lotteries are conducted. However, to ensure the balance of the proportion of accepters between both groups, we must use inverse probability of been called to enroll in the course reweighting[6].

Considering "Administrative 1" data, Table 7 shows that there is a statistically significant correlation between attrition and the treatment. However, considering "Administrative 2" data, Table 8 displays that there is no statistically significant correlation between attrition and the treatment. Besides, as we can see in Tables 9 and 10 it seems that both treatment and control groups (using "Administrative 2" data), are likely to be well balanced in observed characteristics (using inverse probability of being called reweighting in our balance checks).

[5] Behaghel et al. (2015).

[6] Rosenbaum & Rubin (1983), Frölich (2007).

Table 7 – OLS estimation for correlation between attrition and treatment status ("Administrative 1" data, using all subscribers)

VARIABLES	Attrition
$Z_i = 1$	0.200***
	(0.0295)
Constant	0.169**
	(0.0842)
Observations	2,007
R-squared	0.083
Lottery Dummies	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Thus, even if there is a significant correlation between attrition and treatment (considering "Administrative 1" data), treatment and control groups in both data ("Administrative 2" and survey data from our sample) are likely to be well balanced in observed variables, than, we make an assumption that, they are likely to be well balanced even in unobserved variables as well. We test the robustness of our estimations using this assumption using hypothesis tests through randomization inference[7]. Results from randomization tests[8] corroborated the ones we found using conventional ones. That is, outcomes with treatment coefficients that are significant on conventional tests remained statistically significant using randomization tests.

Nonetheless, attrition can bias our estimations if unobserved characteristics are correlated with sample attrition and the outcomes analyzed in this study. For these reason, we relax the hypotheses that both groups are likely to be well balanced in unobserved characteristics. We try to deal with attrition applying the nonparametric method from Lee (2009), which construct extreme bounds for the intervention effects when there are significant differences in attrition rates between treatment and control groups. Unfortunately, it leads to wide nonparametric bounds on treatment impacts, which are not very informative. Results for this estimation are also reported at the Appendix (Table 51).

[7] Fisher (1935), Young (2016), Hayes (2000). An advantage of this test is that his distribution does not depend of asymptotic theorems or distributional assumptions. Furthermore, randomization satisfies the exchangeability assumption of the test: that is, the treatment allocation can be permuted across observations.

[8] We exchange within strata one thousand different placebo groups and estimate by OLS the intention to treat effect regression. We apply a one-tailed hypothesis test.

Table 8 – OLS estimation for correlation between attrition and treatment status ("Administrative 2" data)

VARIABLES	Attrition
$Z_i = 1$	-0.0122
	(0.0525)
Age	0.00133
	(0.00332)
Male	0.0758**
	(0.0385)
White	0.0617
	(0.0562)
Color Not Declared	-0.0744
	(0.0605)
Secondary School (incomplete)	-0.00689
	(0.0467)
Tertiary School	-0.0675
	(0.178)
Constant	0.471**
	(0.229)
Observations	1,408
R-squared	0.094
Lottery Dummies	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 9 presents results for our balance check using administrative data from SENAI. Dependent variable is an indicator of the treatment status, a dummy variable that assume value equal to one if the individual was assigned to treatment group and equal to zero otherwise, using the variable Z_i to identify treatment and control groups, described in the subsection above. Besides, our estimation included lottery dummies and the inverse probability of Z_i=1 reweighting proposed by Chaisemartin and Behaghel (2015) in order to ensure that both groups are statistically comparable, even when polling several lotteries[9]. Results show that only for one variable there was statistically significant differences between treatment and control groups. It seems that both groups are statistically comparable.

9 For all estimations, unless noted, we used this inverse probability of being called to enroll in the course reweighting, proposed by Chaisemartin and Behaghel (2015).

Table 9 – Linear probability model regressions for conditional mean difference test ("Administrative 2" data), by course modality

VARIABLES	Z_i=1
Male	-0.0326
	(0.0439)
White	-0.0820**
	(0.0322)
Secondary School (incomplete)	-0.0250
	(0.0386)
Tertiary School	0.148
	(0.106)
Age	0.00283
	(0.00293)
Constant	1.154***
	(0.0868)
Observations	1,413
R-squared	0.147
Lottery Dummies	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Finally, we ran balance ckecks using observed characteristics from our sample data. Table 10 presents our balance checks through OLS regressions - a linear probability model to estimate a conditional mean difference test: dependent variable is an indicator of the treatment status, a dummy variable that assume value equal to one if the individual was assigned to treatment group and equal to zero otherwise. We use time invariant variables from survey data, and they are less noisy than the "Administrative 2" ones, since we do not have a way to verify how much employees were concerned to precisely fill those information on dataset. There are no significant differences between treatment and control groups, thus, it seems that both groups are likely to be statistically comparable.

Table 10 – Linear probability model regressions for conditional mean difference test (sample data), by course modality

VARIABLES	$Z_i = 1$
Male	-0.0137
	(0.0527)
Age	-0.00186
	(0.00307)
White	0.000261
	(0.0287)
Mother's Highest Education	
Primary School	0.0592
	(0.0586)
Secondary School	0.0700
	(0.0586)
Do not Know	0.0343
	(0.0986)
Father's Highest Education	
Primary School	-0.0290
	(0.0702)
Secondary School	-0.0805
	(0.0697)
Do not Know	0.0349
	(0.0721)
Constant	1.013***
	(0.102)
Observations	718
R-squared	0.142
Lottery Dummies	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

4 Empirical Strategy

The main challenge on estimating the impact of Vocational Education and Training is to identify only the effects resulting from the abilities acquired from this specific kind of Secondary Education. This study provides rigorous evaluation of these effects, since the treatment was randomly assigned. However, the evaluation of the voucher-type scholarship of PRONATEC's Bolsa Formação must consider that the program compliance was not perfect. Only a fraction of the individuals that could have enrolled took up the treatment. In addition, this intervention was designed as a randomization with waiting list, and as a consequence, some individuals that initially were allocated to the comparison group received the treatment. There are several challenges our study must address to properly identify the causal effect of the intervention.

First, in order to avoid the reintroduction of selection bias, we must compare all individuals initially assigned to the treatment group to all those allocated to the comparison group, whatever their actual treatment status (Duflo et al., 2007). In this context of imperfect compliance the randomization generates an instrument Z_i for the treatment of interest T_i. According to the theoretical framework used by Angrist and Imbens (1994; 1995), under the assumptions of independence, exclusion restriction, existence of a first stage and monotonicity, the Local Average Treatment Effect (LATE) is the effect of treatment on the population of compliers. These are defined as those who, in the absence of the randomly assigned instrument, would not have been treated but are induced to receive treatment by the assigned instrument. Denote Z_i the variable that is randomly assigned, while T_i remains the treatment of interest. Define Y_{0i} the potential outcome for an individual if $Z_i = 0$, and Y_{1i} the potential outcome for an individual if $Z_i = 1$.

Random assignment implies that the difference $E[Y_{0i}|Z = 1] - E[Y_{0i}|Z = 0]$ is equal to zero, and that the difference $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$ is equal to the causal effect of Z_i. However, it is not equal to the effect of the treatment , T_i, since Z_i is not equal to T_i. Because Z_i has been chosen to at least affect the treatment, this difference is called the *Intention to Treat estimate* (ITT). On the other hand, if one is interested in the effect of the intervention (T_i) itself it is necessary to recover the LATE from the the ITT estimates (Duflo et al, 2007, Angrist and Imbens, 1994; 1995). In our study, we will estimate both ITT and LATE effects.

4.1 Intention to Treat Estimate

The Intention to Treat is important, because it shows the impact of this design of PRONATEC's Bolsa Formação as a public policy as a whole: it provides evidences of how the possibility of enrolling on a secondary technical education provided by a private institution impacts, on average, the potential beneficiaries. It is an interesting measure, specially for policy makers, since, in this kind of intervention design, they cannot control students' decisions of attending the technical course once they were appointed to the treatment group. Policy makers can only decide to offer or not this kind of education, but they cannot control beneficiaries behavior as a response for the policy supply.

Given the random assignment of the treatment, the ITT empirical model of our study is given by the following regression:

$$Y_i = \beta_0 + \beta_1 Z_i + X_i'\gamma + W_i'\phi + \varepsilon_i, \tag{4.1}$$

where Y_i is the individual i labor market, human capital, risky behavior and socio-emotional outcomes, Z_i is a dummy variable equal to one for individuals who were randomly assigned to treatment group and zero otherwise, X_i is a vector of individual i covariates and W_i is a vector of dummy variables which identifies individual i class lottery[1]. And β_1 is the coefficient which captures the average ITT effect in our empirical model.

4.2 Instrumental Variable Estimate

4.2.1 First Stage

The first stage of our estimation, which captures the causal effect of Z_i on T_i is given by:

$$T_i = \pi_0 + \pi_1 Z_i + X_i'\gamma + W_i'\alpha + \xi_i \tag{4.2}$$

where

T_i is the individual i treatment status, Z_i is a dummy variable equal to one for individuals who were randomly assigned to treatment group and X_i is a vector of individual i covariates and W_i is a vector of dummy variables which identifies individual i class lottery.

[1] As we are pooling several lotteries and the randomization was made by class level, we added lottery dummies in all estimations.

4.2.2 Second Stage

The second stage of our LATE estimation is given by the following equation:

$$Y_i = \beta_0 + \beta_1 \widehat{T}_i + X_i'\phi + W_i'\delta + \eta_i, \tag{4.3}$$

where Y_i is the individual i labor market, human capital, risky behavior and socio-emotional outcomes, $\widehat{T}_i$ is the predicted value of the individual i treatment status, X_i is a vector of individual i covariates and W_i is a vector of dummy variables which identifies individual i class lottery. β_1 is the coefficient which captures the average LATE effect in our empirical model.

Furthermore, as we discussed in the previous subsection, Classification of individuals in Treatment and Control Groups, another challenge for our study is that this experimental design is a randomization with waiting list (Abdulkadiroglu et al., 2011, Behaghel et al., 2015, Curto Fryer, 2014, Chaisemartin and Behaghel, 2015). This implies that treatment and control groups may not be statistically comparable, because there is a correlation between receiving an invitation and being an accepter. Thus, comparing the mean outcome of individuals receiving and not receiving an invitation estimates the effect of receiving an invitation plus a selection bias term, resulting from the asymmetric proportion of accepter between both groups. Therefore, we applied the technique proposed by Chaisemartin and Behaghel (2015) through which the new variable identifying treatment and control groups is constructed as follows: $Z_i = 1\{C_i < \tau\} - 1\{C_i = \tau\}$. This slight modification implies that in expectation both groups contain the same proportion of accepters, making them statistically comparable. Using the variable Z_i to identify treatment and control groups allow us to estimate the "intention to treat" and the "local average treatment effect" effects. Moreover, as our randomization is pooling several lotteries (since the randomization was made for each class), the authors show that we must use inverse probability of being called reweighting to ensure that the results still hold even when polling several lotteries[2].

In addition, we use nonparametric method from Lee (2009), which construct extreme bounds for the intervention effects, for two purposes: first, we use this method to try to deal with the possible bias coming from attrition. Second we use those nonparametric extreme bounds to address the sample selection problem that arises when estimating the effect of training programs on wages (Lee, 2009, Heckman, 1979). We can observe wages only for employed individuals, and even randomization can not ensure that both groups of beneficiaries and control individuals are comparable conditional on being employed. Therefore, under the assumptions of independence and monotonicity, the trimming procedure leads to bounds on

[2] Rosenbaum Rubin (1983), Frölich (2007).

average treatment effects for those whose sample selection status has been affected by the treatment.

Finally, we test the robustness of our ITT treatment coefficients, obtained through conventional statistical inference, which tests the null hypothesis of no average treatment effect, through randomization tests[3]. The null hypothesis in randomized tests is different from the one of no average treatment effect tested in conventional statistical inference. According to Young (2016), the null hypothesis of no treatment effect in randomized tests is that there is no effect for all i, i.e. "the experiment has absolutely no effect on any participant. This is not a null of zero average treatment effect, it is a null of no effect whatsoever on any participant" (Young, 2016, p. 26). Results from randomization tests[4] (Tables 15 to 38) corroborated the ones we found using conventional ones. That is, outcomes with treatment coefficients that are significant on conventional tests remained statistically significant using randomization tests[5]. One advantage of Fisherian tests is that only them can provide an exact test of mean difference for finite sample regardless the knowledge of the distribution of disturbances.

[3] Fisher (1935), Young (2016), Hayes (2000). An advantage of this test is that his distribution does not depend of asymptotic theorems or distributional assumptions. Furthermore, randomization satisfies the exchangeability assumption of the test: that is, the treatment allocation can be permuted across observations.

[4] We exchange within strata one thousand different placebo groups and estimate by OLS the intention to treat effect regression. We apply a one-tailed hypothesis test.

[5] Also known as permutation or Fisherian tests.

5 Labor Market and Human Capital Outcomes

The next subsection briefly review the related literature about the effect of VET programs on labor market outcomes and is followed by the subsection presenting our results for these outcomes.

5.1 Related Literature

Labor market is one of the most widespread dimension analyzed when assessing the effects of Vocational Educational and Training programs. There are well documented evidences not only for developed, but also for developing countries. There are different ways of delivering vocational education, and countries have adopted different structures when providing VET courses, which yield different returns and outcomes. Particularly for the vocational secondary courses, targeted to smooth the school-to-work transition for young people, there are many different possibilities of course designs, and, as a consequence, several issues and questions regarding how these courses should be delivered on the ground.

In developed countries, there is evidence of the impact, controlling for selection, of VET programs on labor market outcomes, such as wage (Dearden et al., 2002, Neuman and Ziderman, 1999). Dearden et al. (2002), using matching approach, found that, for men, the return of vocational education is around 15%, and for women it is 19.8%. Neuman and Ziderman (2001) found a significant impact of 10.3 % on wage, but only for those working in the same area which they have made the course. Hanushek et al. (2011), using a longitudinal dataset and difference-in-difference with country fixed effect approach, found that in the short-run vocational students exhibit greater positive impact on labor market outcomes (wage and employment probability). However, these results are diminishing over the life-cycle, and over 50 years, observed labor market outcomes from general education are greater than the vocational ones. Although Malamud and Pop-Eleches (2010), using regression discontinuity design, analyzed the effects of secondary vocational education in Romania. They found no statistically significant impact on employment probability and wage. Greenwood et al. (2007) and Dearden et al. (2004), despite not controlling for selection bias, provide evidences of negative impact of vocational education on average wage for UK.

There are also evidences of the benefits of VET programs in developing countries, especially on earnings. Tansel (1998), in his study for Turkey, found that when controlling

for observed characteristics and sample selection, there are positive impacts of vocational education on wages and employment probability for men. Attanasio et al. (2011), in their evaluation of a randomized training program in Colombia, found positive effects on earnings and employment probability, especially for women (their wage is 18% higher, and the employment probability is 5 percentage points higher). Card et al. (2011), used a randomized experiment to evaluate the impact of a training program in Dominican Republic. As well, they found that the program has no significant impact on employment probability and has positive impact on earnings per month (10%). Particularly for Brazil, there are some studies that have already found positive impacts of VET on labor market variables (Almeida et al., 2014, Vasconsellos et al., 2010, Assunção and Gonzaga, 2010, Oliva, 2014, Neri, 2010, Biondi, 2015, Silva, Gukovas and Caruso, 2015).

Almeida et al. (2014), Vasconsellos et al. (2010), Assunção and Gonzaga (2010), Neri (2010), and Biondi (2015), using data from National Household Sample Survey - PNAD - IBGE, from 2007, and propensity score matching approach, found a positive impact of VET programs on wages. For Almeida et al. (2014), Vasconsellos et al.(2010), Assunção and Gonzaga (2010), wages are on average 9.7%, 12.5% and 15.9% higher than the general education ones, respectively. Neri (2010) found a positive effect of 15.1 % on wage and Biondi (2015) also found positive impact of secondary technical education (13.9%) on wage when considering individuals with at most general secondary degree. Silva, Gukovas and Caruso (2015), using SENAI's administrative data, and data from Annual Social Reports (RAIS), estimated the effect of technical courses provided by SENAI. They found positive effects of the courses (4%) on wage.

Oliva (2014), using a longitudinal dataset from Centro Paula Souza (State-level public network of technical education), and difference-in-difference with fixed effects approach, found positive impact of secondary vocational education on wage (7.8%) when comparing to individuals who did not pass in the entrance test to attend vocational education. He also found positive effects on employment and formal work probabilities: 3.47 and 2.7 percentage points, respectively. Neri (2010), Assunção and Gonzaga (2010), and Vasconsellos et al. (2010) also found positive effect of VET on employment probability.

Next section presents our results on measuring the impact of Vocational Educational and Training (PRONATEC's Bolsa Formação Estudante provided by S-System) on labor market and human capital outcomes. We separated results in two main subsections: Intention to Treat and Instrumental Variable. In each subsection, we show results to four labor market outcomes - employment probability, formal work probability, wage and probability to work in the course area; and for human capital outcomes - probability to graduate technical education at SENAI, probability to graduate any technical education, probability

to graduate other technical education, probability to graduate regular education at correct age, attendance rates and academic achievement on regular education, and probability to access tertiary education. For each outcome, we ran estimations for subsamples of concomitant and subsequent modalities and analyzed separately estimations for men and women.

5.2 Results

Next subsections display Intention to Treat results for labor market and human capital outcomes, and are followed by the subsections showing results for Instrumental Variable estimations on labor market and human capital outcomes.

5.2.1 Intention to Treat - Labor Market Outcomes

We carry out Intention to Treat estimations on four main labor market outcomes: employment probability, formal work probability, wage and probability to work in the course area. Our specification includes control variables for sex, age, color and parental education. It also contains lottery dummies, which identifies the class which individual i was enrolled in to participate of the lottery, and inverse probability of $Z_i = 1$ reweighting. Robust standard errors are reported in parentheses. We also split the sample by course modality and sex to analyze some possible heterogeneities on the effect of the program. First column of Table describes the dependent variable that is being analyzed in each estimation, second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women.

Table 11 shows estimation results for ITT effect of PRONATEC on labor market outcomes. Dependent variables are listed in first column. For each outcome, we also report control means. Number of observations for each subsample is reported on the bottom of the Table. For wage variable, we were able to observe information only for individuals who were employed, thus, the number of observations for this outcome is different from the others, and it is reported on the penultimate line of the Table. Results indicate that probably there is an heterogeneity on the effect of the intervention. Women are likely to be the most benefited by the program. These results are aligned to ones found by Attanasio et al. (2011), and Friedlander et al. (1997).

We find impact of PRONATEC on employment probability only for the subsample of women in subsequent modality. The probability to be employed is 18.5 percentage points higher, significant at the 10% level, for this specific subsample. For this outcome, results are similar to ones from Card et al. (2011), who found no effect of the randomized program in Dominican Republic on employment probability. These lack of impact may be due to regional characteristics, since Santa Catarina has the lowest unemployment rate comparing to the rest of the country. In the second trimester of 2016, unemployment rate in Santa Catarina was 6,7%, while for Brazil it was 11,3% in the same period. Although, we find no

impact of the intervention on employment probability, we expect to capture PRONATEC's effects on outcomes that may reflect a measure of job quality, such as, individuals working in the area of the course which they were enrolled in, formal work, and earnings.

Considering formal work probability, we find effect only for the subsample of women. For the whole subsample of women, there is an increase of 32.3 percentage points, significant at the 1% level, on formal work probability. The effect for women on subsequent modality is 22.3 percentage points, but it is not statistically significant. For women in concomitant modality, there is a positive and significant impact of 31.7 percentage points on formal work probability, significant at the 10% level. There is no significant effect for male subsample. And we also find no impact on formal work probability when analyzing full sample.

In order to analyze the effect of the program on individuals' labor market placement, we look to its impact on probability to work in the course area. We asked even for students who were not offered to attend classes if they are currently working in the area of the technical course for which they applied to. As expected, beneficiaries have greater probability to be currently working in the course area, when comparing to the ones who were not offered technical education. For the full sample, the probability is 11.5 percentage points higher, significant at the 5% level. For concomitant students, the probability is 13.5 percentage points higher, significant at the 5% level. For subsequent modality, the probability is 10.9 percentage points higher and significant at the 10% level. Separating the sample by sex, these positive effect is statistically significant only for women: the probability is 8.5 percentage points higher, significant at the 5% level.

Finally, the last dependent variable reported is wage. Table 11 shows estimation results for wage[1]. As well as most of labor market outcomes, we find significant effect only for women. Overall beneficiaries women presented an average wage 32.9% higher than the control group ones (significant at the 1% level). However, we must address the sample selection problem that arises when estimating the effect of training programs on wages. We can observe wages only for employed individuals, and even randomization can not ensure that both groups of beneficiaries and control are comparable conditional on being employed (Lee, 2009, Heckman, 1979). We applied the method developed by Lee (2009) to deal with this sample selection problem, bounding the treatment effects. Under the assumptions of independence and monotonicity, the trimming procedure leads to bounds on average treatment effects for those whose sample selection status has been affected by the treatment.

We can not include covariates in our estimation of the nonparametric extreme bounds for the program effect on wages. In order to compare results between the conventional estimation and the extreme bounds for the ITT effect of PRONATEC on wages, we carry

[1] Logarithm of wage.

out conventional ITT estimation using only wage as dependent variable and the treatment status dummy as explanatory variable. Table 12 displays results for this estimation. The effect of the program is still statistically significant for the subsample of women.

Table 11 – ITT estimation on Labor Market Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Employment Probability	0.00875	-0.0566	0.0285	-0.0411	-0.0450	-0.0355	0.133	-0.171	0.185*
	(0.0429)	(0.103)	(0.0443)	(0.0443)	(0.128)	(0.0495)	(0.0902)	(0.180)	(0.0979)
Control mean	0.8128	0.7842	0.8261	0.8564	0.7774	0.8947	0.7311	0.7977	0.7023
Formal Work Probability	0.0638	0.141	0.0489	-0.0804	-0.0455	-0.0894	0.323***	0.317*	0.223
	(0.0593)	(0.117)	(0.0670)	(0.0589)	(0.156)	(0.0650)	(0.104)	(0.162)	(0.138)
Control mean	0.6366	0.5079	0.6965	0.7404	0.6727	0.7732	0.4425	0.1752	0.5582
Probability to Work in Course Area	0.115**	0.135**	0.109*	0.113	0.228*	0.0881	0.0854**	0.00215	0.124*
	(0.0505)	(0.0678)	(0.0597)	(0.0739)	(0.119)	(0.0829)	(0.0379)	(0.0551)	(0.0627)
Control mean	0.1156	0.0583	0.1423	0.1775	0.0872	0.2212	0.0000	0.0000	0.0000
Observations	718	318	400	487	202	285	231	116	115
Ln (Wage)	0.0706	0.109	0.0554	-0.00622	-0.0533	0.0169	0.329***	0.291*	0.259*
	(0.0546)	(0.131)	(0.0586)	(0.0613)	(0.173)	(0.0640)	(0.123)	(0.156)	(0.155)
Control mean	1584.43	1281.30	1690.15	1775.15	1550.69	1847.61	1140.33	758.06	1298.40
Observations	556	216	340	392	145	247	164	71	93
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 13 presents results for the lower and upper bounds of the treatment effect of PRONATEC on wages. It shows the estimates of the bounds, their standard errors and p-values. In addition, last two columns show trimming proportion and the number of observations for each subsample. Although not statistically significant, results for lower bound corroborate the ones obtained by Lee(2009) and Chen and Flores (2012). They include zero and negative effects. According to Lee (2009), "these lower bound is based on an extreme and unintuitive assumption - that wage outcomes are perfectly negatively correlated with the propensity to be employed" (Lee, 2009, p. 30), which is implausible for most standard models of labor supply. That is, "those on the margin of being employed will have lowest wages, not the highest wages" (Lee, 2009, p. 30). This negative estimation is specially observed for male subsample, and its point estimation is greater for concomitant modality.

One possible explanation is that students are lacking labor market experience, comparing to the control ones, because they are attending vocational education, specially ones enrolled on concomitant modality, which face a higher course load and have no time to acquire work experience, since they have to allocate their time between general and vocational education. Table 14 shows that students of concomitant modality have on average almost 8 months less labor market experience when comparing to students assigned in control group. This difference is also verified for male students: treatment students have 31 months of labor market experience, while control students have nearly 39. According to Lee (2009) 8 months of less labor market experience could result in nearly 5.8 % wage disadvantage even four years after randomization (for the large majority of our sample (573), randomization occurred in 2013).

Considering only the time period after graduating vocational course, treatment students still have nearly 4 months, on average, less labor market experience than control ones (17 and 13 months, respectively). One possible explanation for the lack of impact of PRONATEC on wages is that we are likely capturing short-run effects of PRONATEC on wages, where students are still gathering their benefits of human capital investments. Particularly for males this disadvantage can be higher, since they face greater experience gap.

Table 12 – ITT estimation on Ln(wage), by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Ln (Wage)	0.0864	0.0779	0.122*	0.0243	-0.0842	0.0874	0.214*	0.389**	0.179
	(0.0660)	(0.148)	(0.0678)	(0.0632)	(0.147)	(0.0692)	(0.125)	(0.150)	(0.140)
Constant	7.257***	7.021***	7.339***	7.399***	7.255***	7.446***	6.925***	6.567***	7.073***
	(0.0614)	(0.145)	(0.0608)	(0.0563)	(0.142)	(0.0596)	(0.119)	(0.142)	(0.130)
Observations	556	216	340	392	145	247	164	71	93
R-squared	0.004	0.003	0.009	0.000	0.004	0.005	0.029	0.096	0.024

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 13 – Bounds on Treatment Effects for ln(wage), by course modality and sex

		Lower Bound	Std. Error	P>z	Upper Bound	Std. Error	P>z	Trimming Proportion	Obs.
Full Sample	All Subscribers	0.064	0.085	0.451	0.109	0.071	0.127	0.019	556
	Concomitant	-0.030	0.149	0.839	0.187	0.149	0.210	0.125	216
	Subsequent	0.072	0.079	0.363	0.171	0.078	0.028	0.053	340
Men	All Subscribers	-0.031	0.068	0.647	0.073	0.065	0.261	0.057	392
	Concomitant	-0.158	0.180	0.379	-0.084	0.145	0.562	0.072	145
	Subsequent	0.059	0.079	0.459	0.122	0.075	0.104	0.034	247
Women	All Subscribers	0.166	0.132	0.211	0.279	0.126	0.026	0.056	164
	Concomitant	0.303	0.172	0.079	0.537	0.186	0.004	0.215	71
	Subsequent	0.016	0.146	0.915	0.316	0.136	0.020	0.220	93

Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Upper bound estimation shows positive and significant effects for the subsample of subsequent modality (17.1% - significant at the 5% level). For women as a whole, there is a positive and significant effect: 27.9%, significant at the 5% level. When splitting the subsample of women by course modality, we find an effect of 53.7% in concomitant modality (significant at the 1% level), and of 31.6% in subsequent modality (significant at the 5% level).

Table 14 – Average Labor Market Experience (months) for concomitant modality

Total Labor Market Experience (months)			Labor Market Experience after the course graduation (months)		
Full Sample	Control	37.96	Full Sample	Control	17.04
	Treatment	30.53		Treatment	13.43
Men	Control	38.95	Men	Control	18.45
	Treatment	31.00		Treatment	13.07
Women	Control	34.50	Women	Control	12.33
	Treatment	29.77		Treatment	13.96

Source: author's elaboration using survey data.

Moreover, robustness of our ITT estimations is corroborated through randomization tests. Table 15 summarizes our randomization P-values (considering our full sample, the subsample of women and the subsample of women in subsequent modality), which is the probability we would see such an extreme test-statistic if the null hypothesis is true. Under the null hypothesis of no effect for all i, we note that, for those coefficients that we find significant treatment effect, results show that they are also able to reject the null hypotheses of no treatment effect anywhere.

Table 15 – Randomization Test

	Randomization P-value		
	Full sample	Women	Women (subsq.)
Employment Probability	0.724	0.051	0.034
Formal Work Probability	0.216	0.000	0.04
Ln (Wage)	0.311	0.000	0.071
Probability to Work in Course Area	0.037	0.124	0.081

Source: author's elaboration.

5.2.2 Intention to Treat - Human Capital Outcomes

We also carry out Intention to Treat estimations on seven human capital outcomes: probability to graduate technical education at SENAI, probability to graduate any technical education, probability to graduate other technical education, probability to graduate regular school at correct age, attendance rates on general education, academic achievement on general education, and probability to access tertiary education. As well as regressions for labor market outcomes, our specification includes control variables like sex, age, color and parental education. It also contains lottery dummies, which identifies the class which individual i was enrolled in and inverse probability of being invited to enroll in the course reweighting. Robust standard errors are reported in parentheses. We also split the sample by course modality and sex to analyze some possible heterogeneities on the effect of the program. First column of Table describes the dependent variable that is being analyzed in each estimation, second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women.

Table 16 shows results for the four Human Capital outcomes applied to students in both modalities: concomitant and subsequent. As expected, for those who technical education was offered, the intervention had a positive and statistically significant effect on the probability to graduate technical education at SENAI and on the probability to graduate any technical education. Overall, the probability to graduate technical education at SENAI is 46.1 percentage points higher for treatment group when compared to control group, significant at the 1% level.

We also look to the probability to graduate in any technical education, because the possibility to attend vocational education may have influenced students' decisions to enroll in this kind of education even if they have not received an offer to participate of the program. They could have decided to self-fund their course at SENAI or some other institution. Moreover, for those who technical education was offered, they can have continued their education, attending another vocational course after graduating the one they have been awarded by the program. Considering the full sample, there is a positive and statistically significant impact of the program in this outcome. The probability to graduate any technical education is 14.4 percentage points higher for treatment group, significant at the 5% level. But the effect is not statistically significant for all subsamples. For example, in the subsample of women, it has no significant impact. Considering the probability to graduate other technical education (not including the program course), the probability is higher to individuals in control group than to ones in treatment group. Overall, the probability for beneficiaries is 17 percentage points

smaller, when compared to control group. It seems that students from treatment group are looking for another technical course less than the control ones. It is understandable, since the last ones demonstrated interest in attending this kind of education, making their enrollment. As they were not awarded with the program, 47.4% of them made some other technical course. And this result is quite similar for all subsamples. On the other hand, it indicates that beneficiaries are less likely to complement their education with another technical course. This result is connected with our last outcome, through which we investigate if beneficiary students continue their education through accessing tertiary education.

Finally, we investigate the effect of vocational education on the probability to access tertiary education. Despite the main objective of Vocational Education and Training is to facilitate school-to-work transition for its students, it also can provide them the opportunity, or even the motivation, to continue their investment in years of education, not being necessarily "dead end courses". Sometimes a student who could not pay for college decides not to attend tertiary education. But completing a technical education provides him a vocation, and the income needed to access college. The row of the last dependent variable on Table 16 shows results for our ITT effect estimation on the probability to access tertiary education. They indicate that there is no effect of PRONATEC on the probability to access tertiary education. That is, there is no statistical difference in the probability to access tertiary education between treatment and control groups. Despite the negative signal of the coefficient, 40% of the students in treatment group are attending tertiary education, while in control group about 47% are. One possible interpretation is that students in treatment group are likely to be accessing tertiary education as well as the ones in control group.

In order to investigate in which way vocational educational may influence regular education, for students attending concomitant modality, we carry out estimations of the effect of technical education on the probability to graduate general education at correct age, attendance rates and academic achievement at general education for the subsample of concomitant students. For academic achievement, we use student's grades from Math and Portuguese in their senior year of high school in regular education. Grades are standardized by the grade distribution of the control group. We also split the sample by gender.

Table 17 presents results for ITT estimation of the impact of the program on these outcomes. Considering probability to graduate general education at correct age, we find a negative and significant effect in this outcome. The probability is 11.5 percentage points smaller, significant at the 10% level. For attendance rate at general education, we find no significant effect neither for all subscribers, neither splitting the subsample by sex. Table 17 also shows results for ITT estimation of the effect of technical education on students' academic achievement on general education. To address heterogeneity across different teachers'

grading, we added school fixed-effect in estimations (the ideal should be adding classroom fixed effect, but we had not enough sample size in order to do that) [2]. Results indicate that there is a negative impact of the program on academic achievement for the subsample of men, in both subjects. For Math, their achievement is 18.1% standard deviation of the distribution of the outcome smaller (significant at the 10% level), when comparing to control group. For Portuguese the effect is negative in 18.2% standard deviation of the distribution of the outcome (significant at the 5% level). For women, point estimation indicates a negative relation between attending technical education and their achievement in Math, while for Portuguese it indicates a positive relation, despite estimations for both subjects are not statistically significant. These results draw the attention for the issue regarding the optimal balancing between academic and technical content, and if this extensive course load is warranted. Results point towards that students on concomitant modality may have been damaged in essential subjects (Math and Portuguese) on general education.

[2] Despite adding classroom fixed-effect in estimation, the dispersion problem still remains. Unfortunately, we had no information about classroom's averages to try to deal with dispersion issue.

Table 16 – ITT estimation on Human Capital Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Probability to Graduate Technical	0.461***	0.443***	0.480***	0.469***	0.390**	0.516***	0.485***	0.679***	0.410***
Education at SENAI	(0.0421)	(0.0976)	(0.0497)	(0.0525)	(0.159)	(0.0617)	(0.0683)	(0.139)	(0.0876)
Control mean	0.0898	0.1284	0.0719	0.1096	0.1920	0.0697	0.0529	0.0000	0.0758
Probability to Graduate Any	0.144**	0.00256	0.210***	0.157**	0.00677	0.242***	0.146	0.0342	0.126
Technical Education	(0.0608)	(0.105)	(0.0705)	(0.0728)	(0.146)	(0.0809)	(0.106)	(0.182)	(0.120)
Control mean	0.5639	0.7735	0.4662	0.5899	0.7953	0.4905	0.5153	0.7297	0.4225
Probability to Graduate Other	-0.170***	-0.242*	-0.119	-0.161**	-0.136	-0.119	-0.217**	-0.493***	-0.179
Technical Education	(0.0656)	(0.124)	(0.0778)	(0.0789)	(0.168)	(0.0908)	(0.0994)	(0.142)	(0.115)
Control mean	0.4741	0.6451	0.3943	0.4803	0.6033	0.4208	0.4624	0.7297	0.3467
Probability to Access Tertiary	-0.0539	-0.0951	-0.0543	0.00426	0.00431	-0.00856	-0.0620	-0.120	-0.0544
Education	(0.0533)	(0.105)	(0.0598)	(0.0565)	(0.142)	(0.0652)	(0.113)	(0.120)	(0.138)
Control mean	0.4688	0.6961	0.3628	0.4170	0.6552	0.3017	0.5655	0.7788	0.4732
Observations	718	318	400	487	202	285	231	116	115
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 17 – ITT estimation on Human Capital Outcomes Exclusively for Concomitant Modality, by sex

DEPENDENT VARIABLES	Concomitant		
	All Subscribers	Men	Women
Prob. to Graduate Regular Education	-0.115*	-0.123	-0.0720
at Correct Age	(0.0654)	(0.102)	(0.102)
Control mean	0.9499	0.9475	0.9549
Attendance Rate at Regular Education	-0.817	-0.329	3.066
	(3.337)	(5.261)	(4.025)
Control mean	89.8670	90.3021	88.9886
Observations	311	198	113
Lottery Dummies	yes	yes	yes
Math Grades on Regular School	-0.0793	-0.181*	-0.103
	(0.0836)	(0.107)	(0.161)
Portuguese Grades on Regular School	-0.0843	-0.182**	0.0834
	(0.0552)	(0.0722)	(0.115)
Observations	290	183	107
Lottery Dummies	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education.
To address the imbalance created by pooling lottery strata, all equations
contain inverse probability reweighting.

Furthermore, robustness of our ITT estimations for human capital outcomes is corroborated through randomization tests. Table 18 summarizes our randomization P-values (considering our full sample, the subsample of women and the subsample of women in subsequent modality), which is the probability we would see such an extreme test-statistic if the null hypothesis is true. Under the null hypothesis of no effect for all i, we note that, for those coefficients that we find significant treatment effect, results show that they are also able to reject the null hypotheses of no treatment effect anywhere.

Table 18 – Randomization Test

	Randomization P-value		
	Full sample	Women	Women (subsq.)
Prob. to Grad. Tech. Ed. at SENAI	0.000	0.000	0.006
Prob. to Grad. Any Tech. Ed.	0.028	0.232	0.432
Prob. to Grad. Other Tech. Ed.	0.003	0.026	0.122
Probability to Access Tertiary	0.475	0.948	0.944
Prob. to Grad. Regular Ed. at Correct Age	0.596	0.245	-
Attendance Rate at Regular Ed.	0.879	0.432	-
Math Grades on Regular Ed.	0.355	0.917	-
Portuguese Grades on Regular Ed.	0.037	0.903	-

Source: author's elaboration.

5.2.3 Instrumental Variable - Labor Market Outcomes

Another challenge in our study is to address the imperfect compliance problem. In this case, the randomization can provide an instrumental variable to measure the impact of PRONATEC on labor market and human capital outcomes of the ones who were enrolled on vocational education because they were awarded in the lottery. We carry out instrumental variable estimations on four main labor market outcomes: employment probability, formal work probability, wage and probability to work in the course area. As well as OLS estimations, our specification includes control variables like sex, age, color and parental education. It also contains lottery dummies, which identifies the class which individual i was enrolled in and inverse probability of Z_i=1 reweighting. Robust standard errors are reported in parentheses and the F-test of excluded instrument is reported in the penultimate row. We also split the sample by course modality and sex to analyze some possible heterogeneities on the effect of the program. First column of Table lists our dependent variables. Second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women.

Table 19 displays results for First Stage estimation for each different subsample. It is possible to note that the magnitude of the coefficient is high, and it is statistically significant (at the 1% level) for all subsamples. These indicate that there is a strong correlation between to receive an offer to enroll in the program's course and the probability to actually enroll in technical education course provided by the program.

Table 20 shows LATE estimation results for Labor Market Outcomes. Mainly, like results found on ITT estimation, we find effect for the subsample of women. Considering the employment probability outcome, we find effect only for the subsample of women in subsequent modality. The probability in this subsample is 25.6 percentage points higher for the treatment group, significant at the 5 % level. There are also positive effects on formal work probability, for the subsample of women. As displayed in column 8, this probability is 43.2 percentage points higher, significant at the 1 % level. The last labor market outcome which we can observe information for all individuals in the sample is probability to work in course area. There is a positive and significant effect of PRONATEC on it. As we can see in column 2, the probability is 15.2 percentage points higher to the treatment group, when compared to control group. And it is significant at the 5 % level. As expected, the results found in IV estimations are quite similar to the ones in previous section (ITT estimation). However, the magnitude of the coefficients is greater in Instrumental Variable approach.

Table 19 – First Stage of IV estimations, by course modality and sex

A: First-Stage of IV estimation for Labor Market Outcomes	Full Sample			Men			Women		
DEPENDENT VARIABLES	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Z_i=1	0.759***	0.755***	0.733***	0.794***	0.786***	0.794***	0.747***	0.745***	0.723***
	(0.0438)	(0.0695)	(0.0591)	(0.0391)	(0.0972)	(0.0462)	(0.0560)	(0.147)	(0.0713)
Constant	0.197	-0.148	0.188	0.260*	-0.276	-0.725***	0.149	-0.612	0.395
	(0.124)	(0.568)	(0.169)	(0.136)	(0.711)	(0.151)	(0.111)	(1.087)	(0.290)
Observations	718	318	400	487	202	285	231	116	115
R-squared	0.361	0.395	0.400	0.378	0.429	0.412	0.406	0.436	0.487
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Parental Education Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
B: First-Stage of IV estimation for Ln (wage)									
Z_i=1	0.756***	0.734***	0.731***	0.792***	0.700***	0.801***	0.757***	0.660***	0.733***
	(0.0493)	(0.0852)	(0.0643)	(0.0422)	(0.126)	(0.0487)	(0.0731)	(0.204)	(0.0777)
Constant	0.317**	-0.514	0.350**	0.281*	0.121	-0.700***	-0.114	-0.648	0.0417
	(0.153)	(0.763)	(0.174)	(0.152)	(0.905)	(0.171)	(0.248)	(1.562)	(0.303)
Observations	556	216	340	392	145	247	164	71	93
R-squared	0.363	0.413	0.390	0.398	0.464	0.424	0.410	0.502	0.513
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Parental Education Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 20 – IV estimation on Labor Market Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Employment Probability	0.0115	-0.0750	0.0389	-0.0518	-0.0573	-0.0447	0.179	-0.230	0.256**
	(0.0549)	(0.127)	(0.0580)	(0.0533)	(0.148)	(0.0585)	(0.114)	(0.217)	(0.121)
Control mean	0.8128	0.7842	0.8261	0.8564	0.7774	0.8947	0.7311	0.7977	0.7023
Formal Work Probability	0.0840	0.186	0.0667	-0.101	-0.0580	-0.113	0.432***	0.425*	0.309*
	(0.0780)	(0.149)	(0.0897)	(0.0700)	(0.179)	(0.0757)	(0.136)	(0.241)	(0.170)
Control mean	0.6366	0.5079	0.6965	0.7404	0.6727	0.7732	0.4425	0.1752	0.5582
Probability to Work in Course Area	0.152**	0.179**	0.149*	0.143	0.290**	0.111	0.114**	0.00288	0.171**
	(0.0660)	(0.0848)	(0.0802)	(0.0900)	(0.140)	(0.0992)	(0.0460)	(0.0647)	(0.0744)
Control mean	0.1156	0.0583	0.1423	0.1775	0.0872	0.2212	0.0000	0.0000	0.0000
F-test of excluded instrument	301.01***	117.94***	154.08***	413.07***	65.38***	295.42***	178.05***	25.76***	102.69***
Observations	718	318	400	487	202	285	231	116	115
Ln (Wage)	0.0934	0.149	0.0758	-0.00786	-0.0761	0.0211	0.435***	0.441*	0.354*
	(0.0689)	(0.169)	(0.0750)	(0.0738)	(0.212)	(0.0747)	(0.157)	(0.226)	(0.187)
Control mean	1584.43	1281.30	1690.15	1775.15	1550.69	1847.61	1140.33	758.06	1298.40
F-test of excluded instrument	235.13***	74.23***	129.26***	352.30***	30.80***	270.56***	107.08***	10.46***	88.93***
Observations	556	216	340	392	145	247	164	71	93
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

The last labor market outcome analyzed on Table 20 is wage[3]. We also only find impact for the female subsample for this outcome. For the subsample of women, there is impact in the whole subsample and also for both, concomitant and subsequent modality. The overall impact of PRONATEC on women's wage is an increase of on average 43.5%(significant at the 1% level). Considering the subsample of females who attended concomitant modality, PRONATEC has a positive effect of on average 44.1 % (significant at the 10% level) on wage, when compared to control group. For the subsample of subsequent modality, wages from beneficiaries women are on average 35.4% (significant at the 5% level) higher when compared to control group. Results of IV estimation on wage are also pretty similar to the ones found for ITT estimation, but the magnitude of the coefficients is larger. Despite our known sample selection problem for wage outcome, as we discussed on ITT section, upper bounds of ITT estimation corroborated the results of a positive and significant effect in the subsample of women. Thus, it seems that our findings points toward positive and significant effects of PRONATEC on women's wages.

Overall, IV estimations results on labor market outcomes corroborate the ones found on ITT estimation. Besides, the magnitude of all IV point estimations is larger. The results for labor market outcomes indicate an possible heterogeneity on the effect of the program: women are likely to benefit more from PRONATEC. There are impacts on employment probability, formal work probability and on wages only for the subsample of women. The only outcome presenting effect for men and for the full sample is probability to work in the course area. These findings are aligned to the ones found by Attanasio et al (2011), Friedlander et al. (1997), and Nõpo et al. (2007). Oliva (2014) found greater impact of vocational eduction on employment probability for women, while on wage he found greater impact for men. Despite we have found positive impact on wage only for the subsample of women, the magnitude of these effect, for this subsample, is greater than the ones found by Almeida et al. (2014), Vasconsellos et al. (2010), Assunção and Gonzaga (2010), Neri (2010), and Biondi (2015).

[3] Logarithm of wage.

Table 21 – Number of interviews by course and sex

	Women		Men		Total
	N	%	N	%	
Mechanics	7	5%	123	95%	130
Administration	11	92%	1	8%	12
Pharmaceutical	6	100%	0	0%	6
Computer Network	20	30%	46	70%	66
Workplace Safety	116	72%	45	28%	161
Building	9	47%	10	53%	19
Electrotechnology	0	0%	114	100%	114
Electronics	0	0%	15	100%	15
Informatics/Comp. Maintenance	10	17%	49	83%	59
Electromechanics	4	6%	66	94%	70
Food Technology	48	73%	18	27%	66
Total	231	32%	487	68%	718

Source: author's elaboration using survey data.

One possible explanation for this heterogeneity could be the labor market experience. As we discussed before, for the subsample of women, labor market experience gap between treatment and control groups is smaller, and considering time period after course graduation, women on treatment group present, on average, even greater labor market experience than the control group ones. Another possible channel explaining this heterogeneity is through course area. Maybe there are some specific courses that are more demanded by women, and the returns to these course areas could be greater than the ones demanded by men. Table 21, above, displays the number of interviews we made according to course and sex. It is possible to note that the courses with greatest proportion of women are: administration, pharmaceutical, workplace safety and food technology.

5.2.4 Instrumental Variable - Human Capital Outcomes

Similarly to the ITT estimations, we also carry out IV regressions on seven human capital outcomes: probability to graduate technical education, probability to graduate any technical education, probability to graduate other technical education, probability to graduate general school in correct age, attendance rates, academic achievement on general education, and probability to access tertiary education. As well as the IV regression for labor market outcomes, our specification includes control variables like sex, age, color and parental education. It also contains lottery dummies, which identifies the class which individual i was enrolled in and inverse probability of Z_i=1 reweighting. Robust standard errors are reported in parentheses and the F-test of excluded instrument is reported in the third-last row. We also split the sample by course modality and sex to analyze some possible heterogeneity on the effect of the program. First column of Table displays our dependent variables. Second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women.

Table 22 shows IV estimation results for human capital outcomes applied to students in both modalities: concomitant and subsequent. In general, results are very similar to the ITT estimation ones. They indicate that, for those who technical education was offered, the intervention has a positive and statistically significant effect on the probability to graduate technical education at SENAI and on the probability to graduate any technical education. Overall, the probability to graduate technical education at SENAI is 60.7 percentage points higher for individuals in treatment group. The probability to graduate any technical education is 19 percentage points (significant at the 5% level) higher to treatment group, when compared to control group. For the subsample of women, there is no statistically significant

effect in this outcome.

Considering the probability to graduate other technical education but the program one, the probability is higher to individuals in control group than to ones in treatment group. Overall, the probability for beneficiaries is 22.4 percentage points smaller (significant at the 5% level), when compared to control group. The magnitude of this impact is greater for concomitant modality: 32 percentage points smaller (significant at the 5% level). For the subsequent modality, the effect is not statistically significant. And this result is quite similar for all subsamples. Finally, we observe no effect of vocational education on the probability to access tertiary education, for none subsample.

Table 22 – IV estimation on Human Capital Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Probability to Graduate Technical	0.607***	0.586***	0.655***	0.591***	0.496***	0.650***	0.650***	0.912***	0.567***
Education at SENAI	(0.0560)	(0.110)	(0.0687)	(0.0608)	(0.164)	(0.0699)	(0.0729)	(0.120)	(0.0985)
Control mean	0.0898	0.1284	0.0719	0.1096	0.1920	0.0697	0.0529	0.0000	0.0758
Probability to Graduate Any	0.190**	0.00340	0.287***	0.197**	0.00862	0.305***	0.196	0.0459	0.174
Technical Education	(0.0753)	(0.130)	(0.0854)	(0.0867)	(0.168)	(0.0939)	(0.128)	(0.211)	(0.146)
Control mean	0.5639	0.7735	0.4662	0.5899	0.7953	0.4905	0.5153	0.7297	0.4225
Probability to Graduate Other	-0.224**	-0.320**	-0.162	-0.203**	-0.173	-0.150	-0.290**	-0.662***	-0.248*
Technical Education	(0.0871)	(0.150)	(0.105)	(0.0962)	(0.189)	(0.108)	(0.127)	(0.210)	(0.143)
Control mean	0.4741	0.6451	0.3943	0.4803	0.6033	0.4208	0.4624	0.7297	0.3467
Probability to Access Tertiary	-0.0710	-0.126	-0.0740	0.00537	0.00549	-0.0108	-0.0830	-0.161	-0.0753
Education	(0.0696)	(0.132)	(0.0799)	(0.0683)	(0.164)	(0.0775)	(0.143)	(0.154)	(0.171)
Control mean	0.4688	0.6961	0.3628	0.4170	0.6552	0.3017	0.5655	0.7788	0.4732
F-test of excluded instrument	301.01***	117.94***	154.08***	413.07***	65.38***	295.42***	178.05***	25.76***	102.69***
Observations	718	318	400	487	202	285	231	116	115
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Moreover, in order to investigate in which way vocational educational may influence regular education, for students attending concomitant modality, we make estimations for the effect of technical education on the probability to graduate general education at correct age, attendance rates and academic achievement at general education for the subsample of concomitant students. We also split concomitant sample by sex. First, 23 show results for the First Stage of IV estimation. It is possible to note that the magnitude of the coefficients are high, and that they are statistically significant at the 1% level. These indicate that there is a strong correlation between to receive an offer to enroll in the program's course and the probability to actually enroll in technical education course provided by the program. As displayed in Table 24, we find no statistically significant effect of the program on the attendance rate at general education. We find a negative and significant (at the 10% level) effect of the program on the probability to graduate regular school at correct age. The probability is 15.3 percentage points smaller for treatment group when compared to control group.

Table 23 – First Stage of IV estimations for Human Capital Outcomes (Exclusively for Concomitant Modality), by sex

A: First-Stage of IV estimation for Human Capital Outcomes

		Concomitant	
VARIABLES	Full Sample	Men	Women
Z_i=1	0.755***	0.786***	0.745***
	(0.0695)	(0.0972)	(0.147)
Constant	-0.148	-0.276	-0.612
	(0.568)	(0.711)	(1.087)
Observations	318	202	116
R-squared	0.395	0.429	0.436

B: First-Stage of IV estimation for Academic Achievement

VARIABLES	Full Sample	Men	Women
Z_i=1	0.786***	0.862***	0.691***
	(0.0692)	(0.0736)	(0.188)
Constant	-0.340	-0.274	-0.581
	(0.589)	(0.780)	(1.195)
Observations	290	183	107
R-squared	0.394	0.461	0.415

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education.
To address the imbalance created by pooling lottery strata, all equations
contain inverse probability reweighting.

Table 24 – IV estimation on Human Capital Outcomes Exclusively for Concomitant Modality, by sex

DEPENDENT VARIABLES	Concomitant		
	All Subscribers	Men	Women
Probability to Graduate Regular	-0.153*	-0.156	-0.0966
Education at Correct Age	(0.0883)	(0.131)	(0.118)
Control mean	0.9499	0.9475	0.9549
Attendance Rate at Regular Education	-1.078	-0.403	4.305
	(4.159)	(5.863)	(4.777)
Control mean	89.8670	90.3021	88.9886
F-test of excluded instrument	117.94***	65.38***	25.76***
Observations	311	198	113
Math Grades on Regular School	-0.0955	-0.220**	-0.185
	(0.0844)	(0.0999)	(0.220)
Portuguese Grades on Regular School	-0.102*	-0.221***	0.150
	(0.0553)	(0.0660)	(0.150)
F-test of excluded instrument	88.76***	49.70***	4.77***
Observations	290	183	107
Lottery Dummies	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Finally, the last two outcomes displayed by Table 24 are related to students academic achievement on general education, using their grades in Math and Portuguese. We also added school fixed-effect in these estimations, as discussed in ITT subsection. Results indicate that there is a negative impact of the program on academic achievement for the subsample of men, in both subjects. For Math, their achievement is 22% standard deviation of the distribution of the outcome smaller (significant at the 5% level), when compared to control group. In Portuguese, it is 22.1% standard deviation of the distribution of the outcome smaller (significant at the 1% level). For women, point estimation indicates a negative relation between attending technical education and their achievement in Math, while for Portuguese it indicates a positive relation, despite the estimations for both subjects are not statistically significant. Results point towards that students on concomitant modality may have been damaged in essential subjects (Math and Portuguese) on general education, as well as the ones found in ITT estimations. Maybe this issue should be discussed by policy makers with more attention, rethinking the optimal balancing between academic and technical content, and its extensive course load, since for concomitant modality it is mandatory to attend general and vocational education concomitantly. This arrangement could create an additional barrier

for low-income students, increasing the opportunity cost of this courses in terms of effort and time.

6 Non-cognitive skills

The next subsection briefly review the related literature about non-cognitive skills and is followed by the subsection presenting our results for non-cognitive skills outcomes.

6.1 Related Literature

Understanding the way people acquire skills over the life cycle, as well as their economic return is one important issue, specially for public policy. However, the debate about vocational education and training is controversy, since international literature found evidence of low economic returns for some kind of VET programs. Nonetheless, public policy must consider that the investments in education impact not only economic dimensions, their benefits go beyond economic returns. For example, reducing crime, improving parenting skills and health outcomes (Borghans et al., 2008, Heckman et al., 2007, Dohmen and Falk, 2010). Unlike when investing in early childhood, when sponsoring VET programs, which are targeted to young people and adults, governments must face the efficiency-equity trade-off. Literature argues that returns to investment in human capital are declining over the life-cycle and that it is hard to enhance cognitive skills if they have not been developed until the first 10 years of life (Carneiro et al, 2010).

However, VET programs can also develop individuals' non-cognitive skills, for which there are evidences indicating that they are more malleable, including in adulthood. Thus, it is very important to comprehend the technology of cognitive and non-cognitive skills acquirement and which are their sensitive periods over the life-cycle, specially for policies like VET. Besides, enhancing non-cognitive skills of young people and adults can have an intergenerational impact, since affecting their non-cognitive skills may affect early cognitive and non-cognitive skills of their children (Carneiro et al, 2010; Carneiro and Heckman, 2003; Cunha and Heckman, 2009; Carneiro, 2009; Wößmann, 2008).

Therefore, there are evidences that socio-emotional skills, such as persistence, responsibility and teamwork, are so important as cognitive skills for improving individuals' academic, professional and personal outcomes (Santos and Primi, 2014, Ikesako and Miyamoto, 2015, Falch et al., 2012, Heckman et al., 2000, Gensowski, 2014, Piatek and Pinger, 2010). Non-cognitive skills, as well as the cognitive ones, can be developed over lifetime. During late childhood and youth, school, friends and community play an important role in enhancing this kind of abilities (Valdivia, 2016; OECD, 2015; Ikesako and Miyamoto, 2015).

There are several well documented evidences of the relationship between non-cognitive skills and academic achievement, good health outcomes, social and economic development (Jencks, 1979, Almlund et al., 2011, Friedman and Kern, 2014, Kautz et al., 2014, OECD, 2015, Poropat, 2009, Gensowski, 2014). Particularly for the relationship between non-cognitive skills and labor market outcomes, Kuhn e Weinberger (2005) made a research with employers, in the United States, and found that the five most valued skills are (by order of importance): communication skills, motivation/initiative, teamwork, leadership and academic achievement. According to Bowles, Gintis e Osborne (2001), employers focus on some personality traits because they facilitate the production of effort and are linked to labor productivity. Conscientiousness is largely associated to wage and other labor market outcomes in the literature (Almlund et al., 2011, Nyhus and Pons, 2005, Hogan and Holland, 2003, Salgado, 1997, Barrick and Mount, 1991). Furthermore, there is evidence of positive relationship between extraversion and labor market outcomes (Cattan, 2010, Heckman et al., 2006).

Moreover, there are well established evidences, in the literature of psychology and criminology, about the relationship between personality measures and crime/delinquency (Caspi et al., 1994, Agnew et al., 2002, Agan, 2011). There are some evidences, in the literature of economics, about the effect of non-cognitive skills on crime (Hill et al, 2011, Agan, 2011, Heckman et al., 2005, Cunha et al., 2010). Blattman et al. (2017), through an experimental intervention targeted to foster self-control and self-image of at-risk young men aged 18 to 35. They found that these characteristics are malleable in adulthood and enhancing them have a significant effect in reducing crime and violence. Heckman (2006) found that non-cognitive skills affect risky behavior outcomes, such as probability of smoking, using marijuana, and being a teenager mother, using the US NLSY data. Chiteji (2010), using the US PSID data, found that personality traits, such as self-efficacy has a negative relation with alcohol use and a positive relation with physical exercise. Mendolia and Walker (2013) used matching methods to investigate the effect of personality traits, such as locus of control, self-esteem and conscientiousness, at age 15-16, on risky behaviours: alcohol consumption; cannabis and other drugs use; unprotected and early sexual activity. They found a negative effect of non-cognitive skills on risky behaviors.

6.2 Results

First, we investigate whether PRONATEC has an impact enhancing beneficiaries' non-cognitive skills. We estimate Intention to Treat and LATE effects using our individual's latent measure of socio-emotional skill. Since vocational education aims to prepare students

for their entrance in labor market, which is increasingly demanding skills that are linked to personality traits, such as leadership, teamwork, motivation/initiative, we expect that this kind of education enhances beneficiaries' non-cognitive skills. After that, we take advantage of our unique dataset to investigate the relationship between non-cognitive skills and labor market, human capital, crime and risky behavior outcomes existing in our dataset. We do not attempt to estimate the causal effect of socio-emotional skills on those outcomes, we only aim to provide evidences of their correlations.

Before estimating the Intention to Treat and Instrumental Variable effects of PRONATEC on non-cognitive skills, we must carry out balance check for the subsample of individuals who answered our socio-emotional questionnaire. As some interviews were made by telephone, not all respondents of the previous dimensions have answered the non-cognitive one.

Table 25 presents the result of our balance check through OLS regression - a linear probability model to estimate a conditional mean difference on observed characteristics test: dependent variable is an indicator of the treatment status, a dummy variable that assumes value equal to one if the individual was assigned to treatment group and equal to zero otherwise. This classification was made using the variable Z_i, described at Section 3 - Experimental Design. Moreover, our estimation included lottery dummies and the inverse probability of Z_i=1 reweighting proposed by Chaisemartin and Behaghel (2017) [1]. Table 25 shows that there are no significant differences in observed variables between treatment and control groups. This result indicates that both groups are likely to be well balanced in observed variables.

[1] In order to ensure that both groups are statistically comparable, even when polling several lotteries.

Table 25 – Linear probability model regressions for conditional mean difference test (non-cognitive sub-sample), by course modality

VARIABLES	Z_i=1
Male	-0.0915
	(0.0641)
Age	-0.000409
	(0.00434)
White	0.00603
	(0.0368)
Mother's Highest Education	
Primary School	0.0963
	(0.0879)
Secondary School	0.0631
	(0.0873)
Do not Know	0.104
	(0.132)
Father's Highest Education	
Primary School	0.104
	(0.114)
Secondary School	0.0109
	(0.116)
Do not Know	0.132
	(0.121)
Constant	-0.105
	(0.162)
Observations	366
R-squared	0.304
Lottery Dummies	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

6.2.1 Intention to Treat - Non-cognitive Skills

We carry out Intention to Treat estimations separately for each of our six non-cognitive skills outcomes: agreeableness, conscientiousness, extraversion, neuroticism, locus of control, and openness to experiences. Socio-emotional scores are standardized by the score distribution of the control group. Each estimation includes control variables for sex, age, color and parental education. It also has lottery dummies, which identifies the class which individual i was enrolled in to participate of the lottery, and inverse probability of $Z_i = 1$ reweighting. Robust standard errors are reported in parentheses. We also split the sample by course modality and sex to analyze some possible heterogeneities on the effect of the program. First column of Table describes the dependent variable that is being analyzed in each estimation, second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women. For each outcome, we also report control means. Number of observations for each subsample is reported on the bottom of the Table.

Table 26 displays our results for Intention to Treat estimations. Overall, similarly to our labor market outcomes, women seem to benefit more of the program on non-cognitive outcomes. For the subsample of women, we find positive and statistically significant effects of PRONATEC on agreeableness (for women in subsequent modality), and on extraversion (for the whole subsample of women and for women in concomitant modality). The positive effect on agreeableness is 72.9% (significant at the 5% level) standard deviation of the distribution of the outcome. For extraversion, there is also a positive impact. It is 66.8% standard deviation of the distribution of the outcome (for the whole subsample of women), and 110.5% for the subsample of concomitant modality (significant at the 5% and 1% level, respectively). We find no statistically significant effect on the subsample of men, and neither considering the full sample of non-cognitive skills respondents.

Considering the socio-emotional outcomes for which there is a significant positive impact of PRONATEC, there are evidences, in the literature, of their relationship with labor market, human capital, and risky behavior outcomes. According to Santos and Primi (2014), agreeableness is likely to have a positive impact in group activities and on educational outcomes. Furthermore, conscientiousness and agreeableness are the most predictive of criminality (OECD, 2014). Particularly for extraversion, Heckman et al. (2006) found strong relation between wages, self-esteem (which is a facet of extraversion) and locus of control. They also found that this two latent measures of non-cognitive skills are also positively related to schooling decisions.

Table 26 – ITT estimation on Non-cognitive Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Agreeableness	0.0467	-0.0612	0.184	-0.0333	-0.113	0.0947	0.236	-0.192	0.729**
	(0.143)	(0.295)	(0.182)	(0.173)	(0.423)	(0.234)	(0.222)	(0.365)	(0.330)
Conscientiousness	0.0878	0.330	0.00622	0.0307	0.452	-0.111	0.284	0.0973	0.581*
	(0.151)	(0.270)	(0.183)	(0.190)	(0.385)	(0.226)	(0.256)	(0.359)	(0.330)
Extraversion	0.173	0.457	0.0325	0.0841	0.145	0.0887	0.668**	1.105***	0.352
	(0.154)	(0.280)	(0.175)	(0.182)	(0.433)	(0.214)	(0.260)	(0.326)	(0.418)
Neuroticism	-0.0278	0.252	-0.0933	-0.143	0.132	-0.157	0.251	0.564	0.311
	(0.168)	(0.304)	(0.205)	(0.219)	(0.405)	(0.255)	(0.327)	(0.496)	(0.444)
Locus of Control	0.0147	-0.0724	-0.106	0.0972	-0.279	0.0767	-0.163	-0.207	-0.415
	(0.174)	(0.299)	(0.215)	(0.231)	(0.450)	(0.269)	(0.300)	(0.415)	(0.409)
Openness to Experiences	0.000739	0.129	-0.000641	-0.0525	-0.0773	-0.0280	0.175	0.317	0.257
	(0.161)	(0.304)	(0.197)	(0.172)	(0.433)	(0.233)	(0.335)	(0.410)	(0.496)
Observations	366	189	177	258	129	129	108	60	48
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Besides that, robustness of our ITT estimations for socio-emotional outcomes is corroborated through randomization tests. Table 27 summarizes our randomization P-values (considering our full sample, the subsample of women and the subsample of women in subsequent modality), which is the probability we would see such an extreme test-statistic if the null hypothesis is true. Under the null hypothesis of no effect for all i, we note that, for those coefficients that we find significant treatment effect, results show that they are also able to reject the null hypotheses of no treatment effect anywhere. Particularly, we see no significance for any coefficient, considering our full sample of non-cognitive skills, and we note that the coefficient of extraversion is statistically significant for the subsample of women.

Table 27 – Randomization Test

	Randomization P-value		
	Full sample	Women	Women (subsq.)
Agreeableness	0.828	0.438	0.011
Conscientiousness	0.720	0.359	0.133
Extraversion	0.489	0.01	0.360
Neuroticism	0.816	0.434	0.416
Locus of Control	0.891	0.631	0.357
Openness to Experiences	0.955	0.574	0.631

Source: author's elaboration.

6.2.2 Instrumental Variable

This section explores the impact of PRONATEC through instrumental variable estimations on non-cognitive outcomes. The specification includes control variables like sex, age, color and parental education. It also contains lottery dummies, which identifies the class which individual i was enrolled in and inverse probability of Z_i=1 reweighting. Robust standard errors are reported in parentheses and the F-test of excluded instrument is reported in the penultimate row. We also split the sample by course modality and sex to analyze some possible heterogeneities on the effect of the program. First column of Table lists our dependent variables. Second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women.

First, Table 28 displays results for First Stage estimation for each different subsample of our non-cognitive skills sample. It is possible to note that the magnitude of the coefficient is high, and it is statistically significant, at the 1% level, for all subsamples. These indicate that there is a strong correlation between to receive an offer to enroll in the program's course and the probability to actually enroll in technical education course provided by the program.

Table 29 shows results of our IV estimations of the impact of PRONATEC on non-cognitive outcomes. Our findings are quite similar to the ones found in ITT estimation. The effect is positive and statistically significant mostly on the subsample of women, and just for three of the six personality traits: agreeableness and extraversion. There is also a positive and statistically significant (at the 10%level) effect on extraversion for the concomitant modality (considering the full sample). The effect on agreeableness is 172.2% standard deviation of the distribution of the outcome, significant at the 5% level. For extraversion, considering the whole subsample of women, the effect is 95.8% standard deviation of the distribution of the outcome (significant at the 1% level), and for the subsample of women in concomitant modality the effect is 128.1% standard deviation of the distribution of the outcome (significant at the 1% level). The impact of PRONATEC on extraversion is also significant in the subsample that covers all subscribers in concomitant modality. It is 58.3% standard deviation of the distribution of the outcome (significant at the 10% level). There is also a positive effect of PRONATEC on conscientiousness. It is 137.3% standard deviation of the distribution of the outcome, significant at the 10% level.

Table 28 – First Stage of IV estimations for Non-cognitive Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Z_i=1	0.722***	0.784***	0.605***	0.779***	0.785***	0.709***	0.674***	0.863***	0.423***
	(0.0715)	(0.0900)	(0.104)	(0.0606)	(0.130)	(0.0862)	(0.108)	(0.153)	(0.152)
Constant	0.266*	0.129	-0.561**	0.129	0.0592	-0.596**	0.382*	0.189	0.294
	(0.147)	(0.697)	(0.251)	(0.170)	(0.950)	(0.232)	(0.218)	(1.286)	(0.423)
Observations	366	189	177	258	129	129	108	60	48
R-squared	0.321	0.392	0.381	0.372	0.412	0.425	0.416	0.483	0.607

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 29 – IV estimation on Non-cognitive Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Agreeableness	0.0647	-0.0780	0.303	-0.0427	-0.143	0.134	0.349	-0.222	1.722**
	(0.190)	(0.344)	(0.289)	(0.208)	(0.478)	(0.299)	(0.305)	(0.345)	(0.855)
Conscientiousness	0.122	0.421	0.0103	0.0395	0.576	-0.157	0.421	0.113	1.373*
	(0.197)	(0.307)	(0.274)	(0.228)	(0.404)	(0.285)	(0.367)	(0.340)	(0.798)
Extraversion	0.237	0.583*	0.0527	0.108	0.184	0.125	0.958***	1.281***	0.826
	(0.197)	(0.315)	(0.256)	(0.218)	(0.480)	(0.271)	(0.325)	(0.338)	(0.866)
Neuroticism	-0.0612	0.352	-0.213	-0.195	0.179	-0.250	0.435	0.698	0.868
	(0.218)	(0.344)	(0.316)	(0.258)	(0.440)	(0.321)	(0.456)	(0.483)	(0.934)
Locus of Control	0.0264	-0.0889	-0.130	0.124	-0.373	0.132	-0.258	-0.227	-0.946
	(0.221)	(0.330)	(0.318)	(0.263)	(0.476)	(0.334)	(0.408)	(0.391)	(0.871)
Openness to Experiences	0.0264	0.166	0.00878	-0.0432	-0.0743	-0.0468	0.288	0.383	0.664
	(0.202)	(0.354)	(0.292)	(0.204)	(0.517)	(0.296)	(0.421)	(0.392)	(0.995)
F-test of excluded instrument	101.92***	75.95***	33.59***	165.05***	36.37***	67.65***	39.25***	31.58***	7.72***
Observations	366	189	177	258	129	129	108	60	48
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

6.2.3 Evidences of Correlation between Non-cognitive Skills and Labor Market Outcomes

"Our failure to take into account the fact that skill is intrinsically a multidimensional object is not only nonsensical, but also misguides both our research and the design of social policy" (Carneiro et al., 2007, p.1). The importance of socio-emotional skills for a range of outcomes has gaining attention by the economic literature in recent years. There are increasingly evidences that non-cognitive skills are as much important (or even more important, in some cases) as the cognitive ones. Specifically on labor market outcomes, they are strong determinants of wages and employment status, for example.

Heckman et al. (2000) found that GED recipients, despite of being intelligent people, lack personality skills, such as discipline and motivation. The authors also found that GED recipients were more likely to exhibit delinquent behavior in adolescence, such as getting into fights, and engaging in crime. As a result, they are more penalized in the labor market, having lower wages than high school dropouts.

Heckman et al. (2006) showed that self-esteem and locus of control strongly affect employment status, work experience, occupational choice and wages. Duncan and Dunifon (1998) also found a strong relation between motivational traits and wages. Moreover, recent surveys reported that employers identify socio-emotional skills as key competences for the workplace (Kuhn and Weinberger, 2005). According to the survey made by the former authors, the five most important personal qualities indicated by employers are: communication skills, motivation/initiative, teamwork skills, leadership skills, and academic achievement/GPA.

In this subsection, we focus on investigating the correlation between non-cognitive skills and the labor market outcomes exiting in our dataset, such as employment probability, formal work probability and wage. We aim to contribute to the literature, providing evidences of the correlation between those variables in our dataset. We carry out OLS estimations in order to analyze these relationship in our dataset. Columns 2, 3 and 4 display results for employment probability (for the full sample of non-cognitive respondents, and splitting the sample by gender). Columns 5, 6 and 7 show results for formal work probability and 8, 9 and 10 for wage. All six latent measurements of non-cognitive skills were added as explanatory variables and non-cognitive scores are standardized by the score distribution. We also include control variables such as gender, age, color and parental education. Robust standard errors are reported in parentheses.

Table 30 displays results for the OLS regression investigating the correlation between labor market outcomes and non-cognitive skills. For employment probability, we find a

negative and significant correlation between this outcome and conscientiousness, only for women. We also find a negative and statistically significant correlation between employment probability and locus of control. These results go in the opposite direction of the ones reported by the related literature. For formal work probability, we find positive and statistically significant correlation between this outcome and agreeableness. And this positive relation is also significant for the subsample of men. Formal work probability is also positively correlated to conscientiousness, for men. Formal work probability is negatively related to openness to experiences, considering the full sample and the subsample of men. Finally, results for wage point that it is positively correlated to conscientiousness (full sample and men) and negatively related to neuroticism (full sample, men, and women), locus of control (full sample and men), and openness to experiences (men).

Results for conscientiousness corroborate the existing evidence in the literature, but the other ones go in the opposite direction. Despite not statistically significant, for women, correlations of agreeableness and extraversion (which are likely to be developed by PRONATEC) have mixed directions. Agreeableness seems to have a positive relation with employment and formal work probabilities, and a negative relation with wages. Extraversion seems to have a negative relation with employment probability and a positive relation with formal work probability and wage. The positive relation between extraversion and wages (despite not statistically significant) is in line with the findings of Heckman et al. (2006).

In the next subsection we investigate the relation between non-cognitive skills and human capital outcomes, such as academic achievement, probability to graduate general school at correct age and probability to access tertiary education.

Table 30 – OLS estimation for correlation between labor market outcomes and non-cognitive skills

	Employment Probability			Formal Work Probability			Ln (Wage)		
VARIABLES	Full Sample	Men	Women	Full Sample	Men	Women	Full Sample	Men	Women
Agreeableness	0.0356	0.0314	0.0324	0.0702**	0.0868**	0.00229	0.00192	0.0362	-0.0493
	(0.0290)	(0.0338)	(0.0609)	(0.0321)	(0.0366)	(0.0681)	(0.0300)	(0.0385)	(0.0527)
Conscientiousness	-0.0332	0.0136	-0.133**	0.0361	0.0784**	-0.0427	0.0712**	0.0743*	0.0474
	(0.0328)	(0.0366)	(0.0635)	(0.0342)	(0.0394)	(0.0697)	(0.0356)	(0.0436)	(0.0673)
Extraversion	0.000786	0.0132	-0.0770	0.0310	0.0217	0.0128	0.0479	0.0484	0.0280
	(0.0289)	(0.0324)	(0.0647)	(0.0314)	(0.0377)	(0.0653)	(0.0374)	(0.0454)	(0.0611)
Neuroticism	0.0116	-0.0172	0.0380	-0.0291	-0.0600	0.00173	-0.170***	-0.212***	-0.108*
	(0.0313)	(0.0363)	(0.0606)	(0.0343)	(0.0423)	(0.0617)	(0.0333)	(0.0389)	(0.0615)
Locus of Control	-0.0766**	-0.0582*	-0.112*	-0.0509	-0.0376	-0.108	-0.0782**	-0.112***	-0.0563
	(0.0302)	(0.0326)	(0.0669)	(0.0337)	(0.0385)	(0.0653)	(0.0332)	(0.0386)	(0.0537)
Openness to Experiences	-0.0283	-0.0428	0.0595	-0.0796**	-0.105***	0.00182	-0.0386	-0.0835**	0.0474
	(0.0321)	(0.0385)	(0.0713)	(0.0332)	(0.0398)	(0.0743)	(0.0311)	(0.0381)	(0.0634)
Male	0.149***			0.187***			0.381***		
	(0.0565)			(0.0592)			(0.0593)		
Age	0.00327	-0.00363	0.0155*	0.00539	-0.00177	0.0177**	0.0308***	0.0295***	0.0276***
	(0.00409)	(0.00462)	(0.00801)	(0.00463)	(0.00555)	(0.00810)	(0.00457)	(0.00501)	(0.00780)
White	0.0339	0.0558	-0.0141	0.0639	0.0954	-0.0426	-0.00106	0.0272	-0.119
	(0.0508)	(0.0569)	(0.117)	(0.0552)	(0.0630)	(0.116)	(0.0563)	(0.0658)	(0.115)
Constant	0.625***	0.955***	0.369	0.426***	0.778***	0.254	6.283***	6.684***	6.511***
	(0.123)	(0.128)	(0.245)	(0.138)	(0.154)	(0.247)	(0.129)	(0.140)	(0.259)
Observations	365	258	107	365	258	107	272	204	68
R-squared	0.130	0.149	0.172	0.146	0.163	0.210	0.349	0.327	0.395
Parental Education Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.2.4 Evidences of Correlation between Non-cognitive Skills and Human Capital Outcomes

In this subsection, we focus on investigating the correlation between non-cognitive skills and the human capital outcomes exiting in our dataset, such as probability to graduate general education at correct age, probability to access tertiary education, attendence rate at general education, and academic achievement. We ran OLS estimations in order to analyze this relationship in our dataset. Columns 2, 3 and 4 display results for employment probability (for the full sample of non-cognitive respondents, and splitting by gender). Columns 5, 6 and 7 show results for formal work probability and 8, 9 and 10 for wage. All six measurements of non-cognitive skills were added as explanatory variables and non-cognitive scores are standardized by the score distribution. We also included control variables such as gender, age, color and parental education. Robust standard errors are reported in parentheses.

According to the literature, there are well established evidences of the relation between socio-emotional skills and education outcomes. Duckworth and Seligman (2005) found that self-discipline is a strong predictor of academic performance for adolescents. Heckman et al. (2006) also showed that non-cognitive skills have a positive impact on educational attainment.

Table 31 displays results for the OLS regression investigating the correlation between human capital outcomes and non-cognitive skills. For probability to access tertiary education, we observe a positive and significant correlation between this outcome and extraversion (considering the full sample), and openness to experiences, only for women. We also find a negative and statistically significant correlation between the probability to access tertiary education and neuroticism, for women. The probability to graduate in general school is positively correlated with agreeableness (women), neuroticism (men), and openness to experiences (full sample and men). Moreover, probability to graduate general education at correct age is negatively correlated with extraversion (full sample and men). For attendance rate at general education, we note a positive and statistically significant relation between this outcome and agreeableness, for the subsample of women, neuroticism (full sample and men) and openness to experiences (full sample and men). We also note a negative and statistically significant correlation between attendance rate at general education and extraversion (full sample and men, for women this relation is also negative, despite not statistically significant).

Table 32 shows results for the OLS regression investigating the correlation between academic achievement (for those who we were able to obtain their regular education grades) and non-cognitive skills. We investigate the relation of socio-emotional skills with individual's grades of Math and Portuguese. Grades are standardized by the grade distribution. As

we have a small sample size, we do not added school fixed effect in this estimation. Socio-emotional scores are standardized by the score distribution. We also included control variables such as gender, age, color and parental education. Robust standard errors are reported in parentheses. We ran separate regressions for Math and Portuguese, and we also split our sample by gender.

We find a positive and statistically significant correlation between conscientiousness and math, and this relation is stronger in the subsample of men. We also note a negative and significant relation between math and neuroticism for women. For student's achievement in Portuguese, we find a positive relation with conscientiousness and openness to experiences. Agreeableness is positively correlated with student's achievement in Portuguese for women, while for men this correlation is negative.

In summary, we find evidences of positive relation between agreeableness and probability to graduate general education at correct age and attendance rate (for women). For achievement in Portuguese, the correlation with agreeableness is positive for women and negative for men. Conscientiousness is positively correlated with achievement in Math and Portuguese. Extraversion is positively correlated with probability to access tertiary education; and negatively correlated with probability to graduate general education at correct age, and attendance rate at general education. Neuroticism is negatively correlated with probability to access tertiary education and achievement in Portuguese (for women); and positively correlated with probability to graduate general education at correct age (for men) and attendance rate at general school. Openness to experiences has a positive correlation with probability to access tertiary education (women), probability to graduate general education at correct age, attendance rate at general education and achievement in Portuguese. We find no statistically significant correlation between locus of control and human capital outcomes.

According to Santos and Primi (2014), evidences in the literature suggest that agreeableness may be important in determining educational results, conscientiousness is the most associated with success in learning. Extraversion usually presents effects statistically equal to zero in education, although it could be relevant to young people's schooling decisions. There are evidences of the association between Neuroticism and years of study. Finally, evidences point that openness to new experiences is strongly related to education outcomes, such as highest educational degree and achievement. Our results for agreeableness, conscientiousness, and openness to experiences seems to be strongly in line with the existing evidences in the literature.

The last subsection focus on investigating the relation between non-cognitive skills, crime and risky behavior outcomes, such as participation in an argument or fight, sale of pirate goods, sale of drugs, sale of stolen goods, drink and drive, use of alcohol, cigarette,

marijuana, and other drugs, use of alcohol more than twice a week, and binge drinking. The relationship between those outcomes and non-cognitive skills is also widely documented in the literature.

Table 31 – OLS estimation for correlation between human capital outcomes and non-cognitive skills

	Probability to Access Tertiary Education			Prob. to Graduate Gen. Educ. at Correct Age			Attendance Rate at Gen. Educ.		
VARIABLES	Full Sample	Men	Women	Full Sample	Men	Women	Full Sample	Men	Women
Agreeableness	-0.0215	-0.0360	0.0772	0.00566	-0.0307	0.111**	0.230	-1.843	7.881*
	(0.0318)	(0.0392)	(0.0576)	(0.0257)	(0.0302)	(0.0498)	(2.148)	(2.590)	(4.050)
Conscientiousness	0.0186	0.0444	-0.0284	0.0165	0.0251	0.0255	-1.334	-0.672	-0.881
	(0.0321)	(0.0396)	(0.0585)	(0.0263)	(0.0326)	(0.0485)	(2.252)	(2.734)	(4.446)
Extraversion	0.0549*	0.0343	0.0667	-0.0607**	-0.0674**	-0.0313	-5.266**	-6.179**	-2.024
	(0.0325)	(0.0389)	(0.0652)	(0.0258)	(0.0312)	(0.0530)	(2.192)	(2.569)	(4.903)
Neuroticism	0.00596	0.0356	-0.105*	0.0377	0.0673*	-0.0495	5.018**	7.642***	-2.667
	(0.0352)	(0.0449)	(0.0614)	(0.0276)	(0.0344)	(0.0492)	(2.353)	(2.817)	(4.538)
Locus of Control	0.0306	0.0365	0.0239	0.0280	0.0391	0.0271	0.786	1.772	-0.289
	(0.0323)	(0.0401)	(0.0607)	(0.0271)	(0.0328)	(0.0513)	(2.305)	(2.855)	(4.539)
Openness to Experiences	0.0471	0.0374	0.115*	0.0476*	0.0556*	0.0567	3.799*	4.292*	4.516
	(0.0319)	(0.0384)	(0.0612)	(0.0247)	(0.0314)	(0.0467)	(1.986)	(2.375)	(4.173)
Male	-0.255***			-0.0732			-7.050*		
	(0.0571)			(0.0483)			(4.168)		
Age	-0.0107***	-0.0111**	-0.0113	-0.0456***	-0.0426***	-0.0501***	-4.553***	-4.346***	-4.862***
	(0.00372)	(0.00444)	(0.00707)	(0.00434)	(0.00505)	(0.00835)	(0.425)	(0.489)	(0.816)
White	0.0451	0.0516	0.00836	0.0778*	0.0639	0.0996	8.464**	6.911	11.77
	(0.0549)	(0.0642)	(0.110)	(0.0454)	(0.0539)	(0.0893)	(3.960)	(4.645)	(7.810)
Constant	0.695***	0.446***	0.681***	1.433***	1.275***	1.492***	142.8***	130.3***	144.5***
	(0.118)	(0.138)	(0.216)	(0.126)	(0.143)	(0.252)	(11.82)	(13.33)	(24.22)
Observations	365	258	107	172	115	57	172	115	57
R-squared	0.167	0.116	0.244	0.462	0.481	0.501	0.542	0.564	0.558
Parental Education Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 32 – OLS estimation for correlation between academic achievement and non-cognitive skills

	Math			Portuguese		
VARIABLES	Full Sample	Men	Women	Full Sample	Men	Women
Agreeableness	0.00624	-0.134	0.270	0.0561	-0.212*	0.606*
	(0.112)	(0.121)	(0.260)	(0.118)	(0.108)	(0.338)
Conscientiousness	0.225**	0.340**	0.114	0.172*	0.290***	0.202
	(0.112)	(0.146)	(0.189)	(0.103)	(0.0978)	(0.240)
Extraversion	-0.112	-0.153	0.115	-0.146	-0.183	0.0417
	(0.112)	(0.145)	(0.210)	(0.107)	(0.129)	(0.255)
Neuroticism	-0.154	-0.0412	-0.360*	-0.0449	-0.00322	-0.381
	(0.118)	(0.171)	(0.192)	(0.126)	(0.142)	(0.260)
Locus of Control	0.0245	0.0517	0.0812	0.0816	0.0421	0.225
	(0.139)	(0.125)	(0.289)	(0.176)	(0.0970)	(0.379)
Openness to Experiences	0.138	0.182	0.0408	0.229**	0.256**	0.231
	(0.104)	(0.134)	(0.141)	(0.101)	(0.122)	(0.145)
Male	-0.101			-0.414*		
	(0.185)			(0.212)		
Age	-0.0426	0.00889	-0.145	-0.0390	0.0217	-0.0442
	(0.0782)	(0.0972)	(0.156)	(0.0743)	(0.0678)	(0.199)
White	0.000194	0.0286	0.228	-0.00991	-0.0240	0.178
	(0.149)	(0.201)	(0.269)	(0.159)	(0.177)	(0.341)
Constant	0.758	-0.442	2.535	0.804	-0.586	0.192
	(1.583)	(1.908)	(3.093)	(1.504)	(1.342)	(3.922)
Observations	172	115	57	172	115	57
R-squared	0.092	0.110	0.297	0.134	0.191	0.339
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

6.2.5 Evidences of Correlation between Non-cognitive Skills, Crime and Risky Behavior Outcomes

Finally, we look for the relationship between non-cognitive skills, crime and risky behavior. This relation has long been discussed in psychology and criminology literature: association between personality measures and crime, delinquency and risky behavior (Capsi et al., 1994, Agnew et al, 2002, Pratt and Cullen, 2000, Agan, 2010, Spear, 2000). Economists have recently started to develop research investigating the effect of non-cognitive skills on crime (Hill et. al, 2011, Agan, 2011).

Heckman et al. (2006), found negative association between non-cognitive skills and smoking by age 18, imprisonment, marital status, pregnancy at adolescence and participation in illegal activities. Carneiro et al. (2007) found that socio-emotional skills influence smoking at age 16, truancy before age 16, school dropout, involvement with crime and adolescence pregnancy. Chiteji (2010), showed that personality traits, such as self-efficacy, has a negative relation with alcohol use and a positive relation with physical exercise. Mendolia and Walker (2013) also found a negative effect of non-cognitive skills on risky behaviors.

Tables 33 to 35 show the results for our OLS estimations. All six measurements of non-cognitive skills were added as explanatory variables and non-cognitive scores are standardized by the score distribution. We also included control variables such as gender, age, color and parental education. Robust standard errors are reported in parentheses. Table 33 displays the correlation between non-cognitive skills and crime outcomes, such as probability to be involved in an argument or in a fight, probability to be involved selling pirate goods, probability to be involved selling drugs and probability to be involved selling stolen goods. We find a negative and statistically significant correlation between probability to be involved in an argument and neuroticism. For probability to be involved in selling pirate goods, there is a positive correlation with agreeableness. We find no significant correlation for the other two outcomes: probability to be involved in selling drugs and stolen goods, possibly because we have too few variation among individuals for those variables.

Table 34 shows OLS results for correlation between non-cognitive skills and risky behavior outcomes. In this Table we focus on use of substances. For probability to use alcohol, we find a positive association with extraversion. This outcome is also negatively correlated to neuroticism. For the probability to use cigarette, we observe a negative correlation with conscientiousness and a positive relation with locus of control (for men). Considering the probability to use marijuana, we find a negative correlation with conscientiousness and a positive association to openness to experiences. There is no statistically significant correlation between the socio-emotional measures and the probability to use other drugs.

Table 35 displays OLS correlations for the remaining risky behavior outcomes. For the probability of binge drinking, there is a positive association between this outcome and extraversion. We also note a negative and statistically significant correlation between probability to binge drinking and neuroticism. For the probability to drink and drive, there is a positive association with extraversion and a negative correlation with neuroticism. Moreover, we observe a negative correlation between the probability to drink and drive and openness to experiences (only for the subsample of men). Considering the probability to use alcohol more than twice a week, there is a negative association with neuroticism and openness to experiences.

In line with the literature, for most of the latent measures of non-cognitive outcomes, we find negative association between them and our crime and risky behavior outcomes. Some personality traits have positive correlation with crime and risky behavior outcomes: agreeableness is positively correlated to probability to be involved in selling pirate goods. Extraversion is positively correlated to use of alcohol, probability to drink and drive, and binge drinking. Locus of control is positively correlated to use of cigarette; and openness to experiences is positively correlated with use of marijuana.

Table 33 – OLS estimation for correlation between crime outcomes and non-cognitive skills

	Prob. to be involved in argument			Prob. to be involved in selling pirate goods			Probability to be involved in selling Drugs			Probability to be involved in selling stolen goods		
VARIABLES	Full Sample	Men	Women	Full Sample	Men	Women	Full Sample	Men	Women	Full Sample	Men	Women
Agreeableness	0.0250	0.0181	0.0144	0.0130*	0.0110	0.0146	0.000601	2.25e-05	0	-0.00456	-0.00730	0
	(0.0225)	(0.0300)	(0.0338)	(0.00754)	(0.00893)	(0.0148)	(0.00107)	(0.00152)	(0)	(0.00611)	(0.00918)	(0)
Conscientiousness	-0.0180	-0.0346	0.0298	-0.0180	-0.0168	-0.0179	-0.00767	-0.0116	0	0.00702	0.0123	0
	(0.0252)	(0.0325)	(0.0327)	(0.0118)	(0.0151)	(0.0148)	(0.00760)	(0.0114)	(0)	(0.00538)	(0.00793)	(0)
Extraversion	-0.0231	-0.0187	-0.0135	0.00599	0.00931	3.77e-05	0.00324	0.00621	0	-0.00513	-0.00617	0
	(0.0210)	(0.0278)	(0.0285)	(0.0108)	(0.00850)	(0.0341)	(0.00327)	(0.00618)	(0)	(0.00870)	(0.0125)	(0)
Neuroticism	-0.0538**	-0.0526*	-0.0543	-0.00623	-0.0102	0.00346	-0.00172	-0.00153	0	0.00493	0.00707	0
	(0.0253)	(0.0314)	(0.0425)	(0.00828)	(0.00748)	(0.0221)	(0.00192)	(0.00228)	(0)	(0.00440)	(0.00674)	(0)
Locus of Control	0.0148	0.0171	0.000863	-8.71e-05	0.000187	-0.000609	-1.92e-05	0.000383	0	0.00988	0.0152	0
	(0.0201)	(0.0246)	(0.0324)	(0.00662)	(0.00722)	(0.0129)	(0.000777)	(0.00143)	(0)	(0.00777)	(0.0113)	(0)
Openness to Experiences	0.0251	0.0380	-0.00254	0.00283	-0.000270	0.00266	0.00445	0.00628	0	0.00221	0.00394	0
	(0.0219)	(0.0302)	(0.0224)	(0.0114)	(0.0116)	(0.0303)	(0.00446)	(0.00631)	(0)	(0.00880)	(0.0116)	(0)
Male	0.0930***			-0.00598			0.00273			0.0144*		
	(0.0307)			(0.0179)			(0.00299)			(0.00780)		
Age	-0.00145	-0.00222	-0.00148	-0.00120**	-0.00108	-0.00110	0.000454	0.000700	0	-0.000299	-0.000254	0
	(0.00205)	(0.00301)	(0.00285)	(0.000604)	(0.000774)	(0.00141)	(0.000456)	(0.000700)	(0)	(0.000593)	(0.00102)	(0)
White	-0.0341	-0.0135	-0.109	0.00814	0.0167	-0.00397	-0.00768	-0.0100	0	-0.0104	-0.0128	0
	(0.0374)	(0.0446)	(0.0741)	(0.0116)	(0.0111)	(0.0286)	(0.00767)	(0.0100)	(0)	(0.0143)	(0.0194)	(0)
Constant	0.0699	0.166*	0.123	0.0301	0.0237	0.0299	6.58e-07	-0.000436	0	0.0204	0.0359	0
	(0.0711)	(0.0957)	(0.0974)	(0.0224)	(0.0169)	(0.0580)	(0.00280)	(0.00453)	(0)	(0.0253)	(0.0417)	(0)
Observations	365	258	107	365	258	107	365	258	107	365	258	107
R-squared	0.103	0.121	0.122	0.058	0.061	0.130	0.038	0.054		0.040	0.047	
Parental Education Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 34 – OLS estimation for correlation between risky behavior outcomes and non-cognitive skills

	Probability to Use Alcohol			Probability to Use Cigarette			Probability to Use Marijuana			Probability to Use Other Drugs		
VARIABLES	Full Sample	Men	Women	Full Sample	Men	Women	Full Sample	Men	Women	Full Sample	Men	Women
Agreeableness	-0.0345	-0.0169	-0.0434	0.0342	0.0428	0.0135	0.00681	-0.00289	0.0243	0.0112	0.0124	0
	(0.0274)	(0.0339)	(0.0574)	(0.0215)	(0.0293)	(0.0325)	(0.0155)	(0.0211)	(0.0223)	(0.0102)	(0.0146)	(0)
Conscientiousness	-0.0281	-0.0415	0.00575	-0.0627**	-0.0670**	-0.0420	-0.0815***	-0.111***	-0.0239	-0.00910	-0.0136	0
	(0.0278)	(0.0289)	(0.0600)	(0.0260)	(0.0333)	(0.0427)	(0.0217)	(0.0297)	(0.0222)	(0.00874)	(0.0128)	(0)
Extraversion	0.0481*	0.0395	0.0506	0.0201	0.0327	-0.00665	-0.00311	0.00960	-0.0176	0.00195	0.00729	0
	(0.0264)	(0.0308)	(0.0592)	(0.0219)	(0.0272)	(0.0222)	(0.0152)	(0.0200)	(0.0169)	(0.00731)	(0.0121)	(0)
Neuroticism	-0.0701***	-0.0802**	-0.0775	0.00945	0.0170	-0.0295	0.00712	0.0180	-0.00278	-0.0125	-0.0157	0
	(0.0270)	(0.0334)	(0.0573)	(0.0191)	(0.0264)	(0.0231)	(0.0133)	(0.0211)	(0.00635)	(0.0110)	(0.0161)	(0)
Locus of Control	-0.0273	-0.0539	0.0281	0.0350	0.0607**	-0.0451	-0.0109	-0.0172	0.0143	0.00781	0.0104	0
	(0.0273)	(0.0337)	(0.0511)	(0.0239)	(0.0308)	(0.0319)	(0.0136)	(0.0200)	(0.0137)	(0.00763)	(0.00951)	(0)
Openness to Experiences	0.0139	0.0107	0.0198	-0.00156	-0.0164	0.0465	0.0341*	0.0500*	-0.00282	0.00443	0.00429	0
	(0.0251)	(0.0305)	(0.0562)	(0.0253)	(0.0325)	(0.0311)	(0.0190)	(0.0268)	(0.00555)	(0.0104)	(0.0152)	(0)
Male	0.158***			0.0768**			0.0536***			0.0258**		
	(0.0498)			(0.0365)			(0.0203)			(0.0110)		
Age	-0.00114	-0.00102	-0.00189	0.00351	-0.00150	0.0131*	-0.00178	-0.00200	-0.000608	0.00186*	0.00282*	0
	(0.00364)	(0.00386)	(0.00847)	(0.00358)	(0.00376)	(0.00692)	(0.00158)	(0.00219)	(0.00112)	(0.00103)	(0.00147)	(0)
White	0.000628	0.0515	-0.134	0.00398	0.0420	-0.103	-0.0330	-0.0316	-0.0368	-0.0275	-0.0380	0
	(0.0457)	(0.0507)	(0.108)	(0.0392)	(0.0476)	(0.0772)	(0.0294)	(0.0392)	(0.0343)	(0.0184)	(0.0246)	(0)
Constant	0.696***	0.834***	0.783***	-0.0589	0.122	-0.214	0.106**	0.158**	0.0727	-0.0293	-0.0216	0
	(0.112)	(0.112)	(0.249)	(0.0959)	(0.103)	(0.202)	(0.0535)	(0.0684)	(0.0712)	(0.0233)	(0.0271)	(0)
Observations	365	258	107	365	258	107	365	258	107	365	258	107
R-squared	0.089	0.082	0.115	0.075	0.085	0.235	0.120	0.130	0.167	0.055	0.064	
Parental Education Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 35 – OLS estimation for correlation between risky behavior outcomes and non-cognitive skills

	Probability to Binge Drinking			Probability to Drink and Drive			Probability to Use Alcohol More than Twice a Week		
VARIABLES	Full Sample	Men	Women	Full Sample	Men	Women	Full Sample	Men	Women
Agreeableness	-0.00344	0.00529	-0.0243	0.0151	0.0444	-0.00852	-0.00364	0.00738	-0.0301
	(0.0252)	(0.0331)	(0.0443)	(0.0266)	(0.0379)	(0.0277)	(0.0118)	(0.0152)	(0.0273)
Conscientiousness	0.0162	0.0184	0.0324	-0.0251	-0.0325	-0.0357	-0.00178	-0.000924	-0.0115
	(0.0291)	(0.0374)	(0.0433)	(0.0314)	(0.0415)	(0.0455)	(0.00740)	(0.0106)	(0.0113)
Extraversion	0.0616**	0.0586*	0.0699*	0.0802***	0.101***	-0.00137	0.0118	0.0177	-0.00730
	(0.0280)	(0.0351)	(0.0384)	(0.0285)	(0.0367)	(0.0382)	(0.0110)	(0.0146)	(0.00802)
Neuroticism	-0.114***	-0.156***	-0.0734	-0.107***	-0.170***	-0.0232	-0.0159*	-0.0261*	0.00234
	(0.0303)	(0.0392)	(0.0487)	(0.0293)	(0.0422)	(0.0329)	(0.00907)	(0.0140)	(0.00532)
Locus of Control	-0.0184	-0.0244	-0.0432	-0.0126	-0.0491	0.0316	-0.0116	-0.00981	-0.0296
	(0.0296)	(0.0368)	(0.0423)	(0.0288)	(0.0400)	(0.0348)	(0.0113)	(0.00988)	(0.0269)
Openness to Experiences	0.00299	-0.00994	0.0272	-0.0302	-0.0671*	0.0404	-0.0249*	-0.0375**	0.00108
	(0.0290)	(0.0382)	(0.0390)	(0.0284)	(0.0374)	(0.0369)	(0.0130)	(0.0189)	(0.00506)
Male	0.170***			0.336***			0.0317*		
	(0.0497)			(0.0438)			(0.0166)		
Age	-0.00212	-0.00708*	0.00222	0.00879**	0.00672	0.0102**	-0.00118	-0.00204	-0.000172
	(0.00322)	(0.00413)	(0.00601)	(0.00369)	(0.00491)	(0.00471)	(0.000869)	(0.00127)	(0.000701)
White	-0.0180	0.0616	-0.243**	-0.0402	-0.0209	-0.107	0.0143	0.0121	0.0123
	(0.0485)	(0.0560)	(0.0966)	(0.0509)	(0.0660)	(0.0715)	(0.0136)	(0.0192)	(0.0142)
Constant	0.101	0.352***	0.216	-0.157	0.232	-0.0733	0.00110	0.0562	0.00133
	(0.0938)	(0.123)	(0.183)	(0.102)	(0.142)	(0.116)	(0.0276)	(0.0410)	(0.0197)
Observations	365	258	107	365	258	107	365	258	107
R-squared	0.096	0.121	0.193	0.164	0.121	0.171	0.057	0.070	0.176
Parental Education Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

7 Crime and Risky Behavior

The next subsection briefly review the related literature about the association between education, crime and risky behavior. It is followed by the subsection presenting our results for crime and risky behavior outcomes.

7.1 Related Literature

Crime reducing is an important issue on public policy, since it brings large economic and social benefits. A relatively large body of work on social science, specially the quantitative ones, focus on finding characteristics which are good predictors to higher criminal participation. One important characteristic that has received special attention is education [1]. There are also well documented works by economists establishing a connection between crime and education (Machin et al, 2011, Lochner and Moretti, 2004, Lochner, 2004, Grogger, 1998, Anderson, 2010, Berthelon and Krueger, 2011, and Deming, 2011, Chioda et al, 2016). However, the literature argues that it is difficult to identify a causal effect of education on crime. The main challenge in estimating the effect of education on crime is that there are unobserved characteristics affecting schooling decisions that probably are correlated with unobservables which also affect the decision to engage in crime.

Thus, it is difficult to determine what is the direction of the causal effect, since there are also evidences of the effect of crime on education (Aizer, 2009, Rodrígues and Sanchez, 2012, Monteiro and Rocha, 2013). Crime is a negative externality with huge social costs, thus, estimating the effect of education on crime may shed some light on the impacts of education beyond economic dimension, and not only in private returns, but also in the social ones. Given the large social costs of crime, even small reductions in crime associated with education may be economically important (Lochner and Moretti, 2004).

Lochner and Moretti (2004) explored changes in state compulsory schooling laws over time in the United States to estimate the effect of education on participating of criminal activity. They deal with the endogeneity of schooling decision by using variation in state compulsory schooling laws as an instrumental variable for years of education. They found that schooling significantly reduces the probability of incarceration and arrest. One extra year of schooling reduces the probability of imprisonment by about 0.1 percentage points for whites and 0.3 to 0.5 percentage points for blacks, both estimations are for male. They use

[1] Examples from the education literature are Sabates (2008), and Sabates and Feinstein (2008).

the National Longitudinal Survey of Youth (NLSY) data to study the relationship between education and self-reported crime, concluding that the results are caused by changes in criminal behavior, and not by differences in the probability of arrest or incarceration conditional on crime.

Machin et al. (2011) also explored a change in the compulsory school leaving age law in England and Wales to estimate a causal effect of education on crime. They used Instrumental Variable and Regression-Discontinuity Design approaches to estimate the impact of the raise in the school leaving age, from 15 to 16 years, which generates a discontinuity in education measures at the time. IV estimates indicate a 4.7% point fall in the conviction rate in the years after the education reform. The discontinuity estimation is qualitatively the same as the IV one, although the coefficient is bigger in magnitude (in absolute terms): 27% point fall.

The four main channels through which schooling might affect criminal participation, according to the existing socio-economic literature, are: (i) income effects, (ii) time availability (iii) patience or risk aversion and (iv) social networks or peers of individuals (Lochner, 2011).

First channel is the income effect, through which education increases individual wage rates and, as a consequence, his opportunity cost of illegal behavior: incarceration is more costly for individuals with higher wages (Lochner, 2004; Lochner and Moretti, 2004; Hjalmarsson, 2008). Grogger (1998) established a negative correlation between crime and wages using data from the 1970s and 1980s. Machin and Meghir (2004) found a negative relation between crime and wage, specially in the bottom 25th percentile of income distribution. On the other hand, there is also some evidences pointing towards a positive relation between education and white-collar arrest rates, despite the effect has not been statistically significant (Lochner, 2004).

Second channel is specially noted in teenagers during their school years: time spent in education avoid them to spend time participating of criminal activities. Tauchen et. al (1994) found a negative relation between time spent at school and the probability of been arrested. Jacob and Lefgren (2003), in order to measure the causal effect of education on crime, used exogenous teacher training days as an instrument to days off school and Luallen (2006), also establishing a causal relation, used unexpected teacher strikes as an instrument for days out of school. Both papers found incapacitation effects of education on criminal participation. Anderson (2010) also found evidence for US, using the minimum high school dropout ages variation across states, and a difference-in-difference-in-difference-type empirical strategy. He also found that keeping youth in school decreases arrest rates.

Third channel through which education may influence crime is its effect on patience

and risk aversion. "More patient and more risk-averse individuals would place more weight on the possibility of future punishments" (Lochner and Moretti, 2004, p.4). For one individual, future returns from a task are discounted according to his patience rate. According to the survey of psychological and neurological literature, carried out by Oreopoulus (2007), young people are probably myopic and focusing on immediate costs from schooling when deciding to drop out of school. This literature also suggests that adolescents "lack abstract reasoning skills and are more predisposed to risky behavior" (Vujic, 2009, p.107). Education can increase patience, reducing the discount rate of future earnings. Moreover, it can also increase risk aversion, which potentially reduces the probability of engaging in criminal activities and of being involved in risky behavior.

Finally, the social networks or peers of individuals channel is also specially important for young people: which are highly influenced by their peers and by their environment (Chowdry et al, 2009, Center for Mental Health in Schools at UCLA, 2007). School may provide an environment through which teenagers enhance their networks and peers, who could influence their probability to engage (or to not engage) in crime and risky behavior.

Spear (2000) made a review of evidences from age-related behaviors, and argues that adolescents may be particularly likely to initiate use of alcohol and other drugs, in comparison to people at other ages. He exposed that, according to the National Institute of Drug Abuse 1996 Monitoring the Future Study, at senior year in hight school, around 50 % of the teenagers have used marijuana/hashish, 65% have smoked cigarettes, and 82% have tried alcohol. Besides that, some of them reported to use this substances in an excessive way. For example, 31 % of 12th graders in the survey reported being drunk one or more times during the past month. According to Crome (1999), some adolescents can develop excessive drug use patterns. They are not immune to substance dependence, such as alcohol, cigarettes and other drugs. For those using substantial amounts of alcohol, they can present some dependence symptoms, such as alcohol tolerance, escalating pattern of alcohol use, and difficulty in reduce the amount of use or quitting (Pollock, 1999). Bachman et al.(1996) used data from the same Monitoring the Future survey to study the cigarette use patterns over years. They found evidences for nicotine addiction even for teenagers. After one year smoking cigarettes, most adolescents reported trying to quit, but they felt a variety of abstinence's adverse effects; 97% of them reported to be still smoking 2 years later. Moreover, most of those adolescents reported to be cigarette dependent. Substance addiction is not a pattern exclusive from young people, but evidences suggest that, at this age, their rate of progression to alcohol/drug dependence may be quicker when comparing to individuals who initiated drug use in adulthood (Spear, 2000, Clark, 1998). Moreover, there are evidences in the literature about a negative relation between school engagement and harmful risk behavior

(Center for Mental Health in Schools at UCLA, 2007, Porter and Lindberg, 2000. Ozer, 2005). According to Bond et al (2007) "young people who are not engaged with learning or who have poor relationships with peers and teachers are more likely to use drugs and engage in socially disruptive behaviors" (Bond et al., 2007, p. 14).

As vocational education is targeted to foster skills demanded by the labor market, it could be closely linked to influencing the development of skills related to personality traits. For example, it can help to develop maturity in young students, as well as their patience and risk averse, reducing risky behavior. Moreover, vocational education could avoid individuals to engage in illegal activities and risky behavior through some of those other channels listed above, reinforcing the effect of education on crime and risky behavior. On net, we expect that most of these channels presented above will lead to a negative relationship between vocational education, crime and risky behavior. In this study we do not attempt to empirically differentiate the many channels through which vocational education may affect criminal activity and risky behavior. The aim of this section of the study is to evaluate the overall impact of PRONATEC on crime and risky behavior participation. Furthermore, we take advantage of our unique dataset to explore some correlations largely documented in the literature, such as the relationship between academic achievement and the use of substances (for those in concomitant modality). In order to investigate these correlations, we constructed a dummy variable which is equal to one for those concomitant students who reported to be less than 18 years old [2] when using substances for the first time. Despite we are not able to establish a causal effect in these last exercises, we contribute to the literature providing evidences of their correlations.

7.2 Results

First, we investigate the causal effect of PRONATEC on crime and risky behavior outcomes. We estimated the Intention to Treat and LATE effects. Results for this estimations are reported in the next two subsections. After that, we carry out some OLS estimations in order to provide evidences of important correlations highlighted in the literature, such as academic achievement and use of substances among young people, and the correlation between individuals' use of substances and their peers use.

[2] According to the definition of adolescent from the Brazilian Statute of the Child and Adolescent. Spear (2000) also argues that the 12 - 18 age span is commonly considered adolescence in humans by the neurological literature

7.2.1 Intention to Treat

We ran Intention to Treat estimations on eleven crime and risky behavior outcomes: participation in an argument or fight, sale of pirate goods, sale of drugs, sale of stolen goods, drink and drive, use of alcohol, cigarette, marijuana, and other drugs, use of alcohol more than twice a week, and binge drinking. Our specification includes control variables for sex, age, color and parental education. It also contains lottery dummies, which identifies the class which individual i was enrolled in to participate of the lottery, and inverse probability of $Z_i = 1$ reweighting. Robust standard errors are reported in parentheses. We also split the sample by course modality and sex to analyze some possible heterogeneities on the effect of the program. First column of Table describes the dependent variable that is being analyzed in each estimation, second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women.

Tables 36 and 37 display our results for Intention to Treat estimations. It is possible to see in Table 36 that there is a positive and statistically significant effect of PRONATEC on the probability to be involved in an argument or fight for the subsample of women in subsequent modality. The probability is 14.2 percentage points higher for women in treatment group, when compared to those in control group, significant at the 5% level. In a first analyses, we would expect a negative effect of PRONATEC on the probability to be involved in an argument or in a fight. However, linking those results to the ones found in the labor market and socio-emotional sections, we could speculate that maybe this result indicates a potential transmission mechanism through women empowerment. For the subsample of women in subsequent modality, we find positive and statistically significant effects of PRONATEC on employment probability, formal work probability and on wage. This increase in their labor market outcomes could have an impact on changing women's perspective: increasing their self-confidence and pride, as well as their notion of taking more control over their lives. Moreover, we find a positive impact of PRONATEC on extraversion for women. Considering that self-esteem is a facet of extraversion, this increase is likely to be linked to the enhancement of women's empowerment. Maybe this greater empowerment is being reflected in this increase on the probability of being involved in an argument. Unfortunately, we had no data about women's household condition, or even about their marriage status, before the program, in order to try to make an in-depth analysis of this empowerment channel - verifying if married women are having more discussions, possibly with their husbands, for example.

We also find a positive and statistically significant effect of PRONATEC on the probability to sell pirate goods. We expected to find a negative effect of PRONATEC on

this kind of activity, but results point towards a positive direction. The probability of being involved selling pirate goods is 2.03 percentage points higher to individuals in treatment group than for the ones in control group, significant at the 5% level. When analyzing this outcome by gender, we note that the effect is statistically significant only for the subsample of men. Then, probably this effect is coming from the subsample of men. We find no significant impact of PRONATEC on other outcomes reported on Table 36.

Table 37 displays results for the remaining crime and risky behavior outcomes investigated in our estimations. There is a positive and significant impact on the probability to drink and drive in the subsample of women. We also expected to find a negative impact on this outcome, but results indicated a positive effect of PRONATEC on the probability of drink and drive for the subsample of women. There is an increase of 14.2 percentage points, significant at the 5% level, on the probability to drink and drive for beneficiaries women, when comparing to women in control group. This effect is specially noted in the subsample of women in subsequent modality (18.1 percentage points, significant at the 5 % level).

Results are pointing in the opposite direction, comparing to our expected results discussed above, specially in the subsample of women in subsequent modality. One possible transmission mechanism is the income one, since women experienced positive effects on employment probability, formal work probability and wage. We show some descriptive statistics for the subsample of women in subsequent modality in the end of next subsection: Intrumental Variable.

Table 36 – ITT estimation on Crime and Risky Behavior Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Probability to be involved in	0.0123	-0.0288	0.0497	-0.0251	-0.0368	0.00534	0.0779*	-0.0702	0.142**
argument/fight	(0.0317)	(0.0852)	(0.0313)	(0.0415)	(0.143)	(0.0395)	(0.0463)	(0.0728)	(0.0598)
Control mean	0.0891	0.1759	0.0487	0.1295	0.2407	0.0756	0.0136	0.0451	0.0000
Probability to be involved in selling	0.0203**	0.0199	0.0242*	0.0165*	0.000179	0.0298	0.0186	0.0440	0.00552
Pirate Goods	(0.00891)	(0.0197)	(0.0128)	(0.00973)	(0.00771)	(0.0181)	(0.0167)	(0.0469)	(0.00989)
Control mean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Probability to be involved in selling	0.00981	-0.00357	0.0163	0.0136	-0.00460	0.0241	0.00308	0.00712	0
stolen goods	(0.00645)	(0.00571)	(0.0102)	(0.00972)	(0.00749)	(0.0147)	(0.00397)	(0.0217)	(0)
Control mean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Probability to Use Alcohol	0.0446	-0.0434	0.108	0.00434	0.00296	0.0448	0.0466	-0.177	0.180
	(0.0574)	(0.0966)	(0.0708)	(0.0564)	(0.0879)	(0.0684)	(0.0984)	(0.185)	(0.113)
Control mean	0.7475	0.8465	0.7014	0.8057	0.8686	0.7752	0.6389	0.8019	0.5683
Probability to Use Cigarette	-0.0117	-0.0312	0.0128	0.0369	0.0172	0.0834	-0.0597	0.0106	-0.0726
	(0.0402)	(0.0910)	(0.0473)	(0.0521)	(0.148)	(0.0596)	(0.0514)	(0.0282)	(0.0759)
Control mean	0.1658	0.1728	0.1625	0.1755	0.2585	0.1353	0.1477	0.0000	0.2116
Observations	718	318	400	487	202	285	231	116	115
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 37 – ITT estimation on Crime and Risky Behavior Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Probability to Use Marijuana	0.00510	0.0197	0.0223	0.00830	0.0397	0.0181	0.00529	-0.0279	0.00973
	(0.0224)	(0.0691)	(0.0237)	(0.0327)	(0.120)	(0.0305)	(0.0251)	(0.0295)	(0.0317)
Control mean	0.0448	0.0818	0.0276	0.0615	0.1223	0.0321	0.0136	0.0000	0.0195
Probability to Use Other Drugs	0.00565	0.0234	-0.00531	0.0108	0.0424	-0.00172	-0.00431	-0.00519	-0.0153
	(0.00996)	(0.0234)	(0.0115)	(0.0130)	(0.0377)	(0.0128)	(0.0199)	(0.00721)	(0.0246)
Control mean	0.0085	0.0000	0.0124	0.0057	0.0000	0.0085	0.0136	0.0000	0.0195
Probability to Use Alcohol More than	0.0215	0.0610	0.0198	0.0430*	0.118	0.0278	-0.00238	-0.00824	0
Twice a Week	(0.0149)	(0.0483)	(0.0182)	(0.0244)	(0.0893)	(0.0266)	(0.00466)	(0.0126)	(0)
Control mean	0.0141	0.0208	0.0109	0.0216	0.0312	0.0170	0.0000	0.0000	0.0000
Probability to Binge Drinking	0.0273	-0.0313	0.0668	-0.00634	0.0339	0.0398	0.0678	-0.167	0.0869
	(0.0463)	(0.115)	(0.0481)	(0.0639)	(0.150)	(0.0702)	(0.0779)	(0.214)	(0.0686)
Control mean	0.1906	0.3012	0.1391	0.2526	0.3502	0.2053	0.0747	0.2023	0.0195
Probability to Drink and Drive	0.0682	-0.0458	0.134**	0.0401	-0.101	0.144	0.142**	0.0352	0.181**
	(0.0535)	(0.0970)	(0.0649)	(0.0761)	(0.156)	(0.0877)	(0.0577)	(0.0596)	(0.0787)
Control mean	0.2914	0.3011	0.2868	0.4326	0.4502	0.4241	0.0273	0.0000	0.0391
Observations	718	318	400	487	202	285	231	116	115
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Besides that, robustness of our ITT estimations for crime and risky behavior outcomes is corroborated through randomization tests. Table 38 summarizes our randomization P-values (considering our full sample, the subsample of women and the subsample of women in subsequent modality), which is the probability we would see such an extreme test-statistic if the null hypothesis is true. Under the null hypothesis of no effect for all i, we note that, for those coefficients that we find significant treatment effect, results show that they are also able to reject the null hypotheses of no treatment effect anywhere. Particularly, we see no significance for any coefficient (except for the coeficient of probability to be involved in selling Pirate Goods), considering our full sample, and we note that the coefficient of probability to be involved in an argument is statistically significant for the subsample of women in subsequent modality.

Table 38 – Randomization Test

	Randomization P-value		
	Full sample	Women	Women (subsq.)
Prob. to be involved in argument/fight	0.757	0.234	0.088
Prob. to be involved in selling Pirate Goods	0.029	0.246	0.658
Probability to Use Alcohol	0.413	0.819	0.182
Probability to Use Cigarette	0.949	0.155	0.144
Probability to Use Marijuana	0.674	0.957	0.844
Probability to Use Other Drugs	0.685	0.815	0.575
Probability to Binge Drinking	0.606	0.387	0.419
Probability to Drink and Drive	0.212	0.082	0.063
Prob. to Use Alcohol More than Twice a Week	0.407	0.248	1.000

Source: author's elaboration.

7.2.2 Instrumental Variable

We also ran instrumental variable estimations on our eleven crime and risky behavior outcomes. As well as our OLS estimations, specification include control variables like sex, age, color and parental education. It also contains lottery dummies, which identifies the class which individual i was enrolled in and inverse probability of Z_i=1 reweighting. Robust standard errors are reported in parentheses and the F-test of excluded instrument is reported in the penultimate row. We also split the sample by course modality and sex to analyze some possible heterogeneities on the effect of the program. First column of Table lists our dependent variables. Second column shows estimation results for the whole sample, third and fourth show results for concomitant and subsequent modalities, respectively. Columns 5, 6 and 7 follow the same logic, but show results for the subsample of men, and columns 8, 9, and 10 show our findings for the subsample of women.

Tables 39 and 40 show the results of our IV estimations of the impact of PRONATEC on crime and risky behavior outcomes. For some variables, such as probability to be involved in selling drugs and probability to be involved in selling stolen goods, there was no variation in the dependent variable, in some subsamples, and this is the reason why we had no results reported in some rows of Table 39. Results for IV estimation are quite similar to ITT ones, but the coefficients have greater magnitude. Table 39 displays that there is a positive and statistically significant impact of PRONATEC on probability to be involved in an argument or in a fight for the subsample of women in subsequent modality. The probability is 19.6 percentage points higher for women from treatment group, in comparison to the ones from control group, significant at the 1% level. As we discussed in the ITT subsection, this result could indicate a channel of women's empowerment, but unfortunately, we do not have enough information to deeply investigate this channel and to properly speculate about this issue.

Another result of the IV estimation is the positive effect of PRONATEC on the probability to be involved in selling pirate goods. The probability is 2.68 percentage points higher to the treatment group, significant at the 5% level. And, spliting by gender, results indicate that the effect is stronger on men. Table 40 shows results from IV estimation for the remaining crime and risky behavior outcomes. The only impact statistically significant was on the probability to drink and drive outcome. Considering the full sample, there is a positive and significant effect in the subsample of subsequent modality. The probability is 18.2 percentage points higher (significant at the 5% level) for individuals from treatment group, in comparison to the ones in control group. For the subsample of women, the probability is 19 percentage points higher (significant at the 1% level) for women from treatment group, when compared to control group ones. This result is quite worrying specially because, according to the World Health Organization (2015), drinking and driving increases the risk of a crash and

the probability of death or serious injury resulting from the car crash. Each year, around 1.2 million people die and other millions suffer injures or disabilities as a consequence of road crashes, mostly in low-income and middle-income countries. These numbers are even more frightening considering the group of young people: young adults among 20 and 29 years have around three times the risk of a car crash, in comparison to drivers aged 30 years and above, at any level of alcohol consumption. According to this report, young people are the most vulnerable population, considering the drink-driving issue. We expected to find a negative effect of PRONATEC in this outcome, supposing that PRONATEC would at least reinforce the education channel through which it enhances people's risk aversion.

Table 41 shows descriptive statistics of the subsample of women in subsequent modality, which presented the most number of statistically significant effects of PRONATEC. Maybe they are reflecting the impact of the intervention in the subsample of women as a whole, but, since this subsample of women has a better balance between the treatment and control groups. Indeed, for some outcomes, the effects are statistically significant exclusively for the subsample of women in subsequent modality. Descriptive statistics are split in control and treatment groups. It is possible to note that women on treatment group are on average younger than the control group ones. They are on average 27 years old, whereas the control group are on average 30 years old. Most of the women in subsequent modality have secondary school as their highest degree of education. In both groups, they account for almost 60% of the highest education degree.

On the other hand, the number of women reporting to have tertiary school or above as their highest degree of education is not negligible: almost 40% of them reported to have at least accessed tertiary education, in both groups. Considering their family background, most of them, in both groups, reported to have both parents whose highest education degree is incomplete primary school, or complete primary school. Only about 20% of them reported to have parents that completed secondary school. As educational degree is commonly used as a predictor for income, we can suppose that on average women in subsequent modality come from disadvantage families, or at most middle-income families.

Furthermore, Table 41 displays that 47% of women in control group have at least one child, whereas in the treatment group this percentage is of about 38%. Considering their labor market outcomes, 88% of the women in subsequent modality reported to be employed, whereas in control group the percentage was of 63%. In treatment group, 75% of the women reported to be employed having a formal work, whereas in the control group, 50% of them reported to have a formal work. The wage of women in treatment group are on average R$ 1494,00, while for those in control group the wage is on average R$ 1336,00. Most of women, in both groups, have declared to be white. Considering their household condition, most of

them, in both groups, declared to be spouse (around 40%) or head of the household(around 37% for control group and around 22% for treatment group). The percentage of women declared as daughter is higher in treatment group(27%) than in control group(17%).

Finally, we ran OLS regression in order to investigate some correlations documented in the literature about the relationship between use of substances and academic achievement; and about the correlation between the use of substances and the fact that the peers of the individuals also use the same substance. These regressions are reported on Tables 42 to 49.

Table 39 – IV estimation on Crime and Risky Behavior Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Probability to be involved in	0.0161	-0.0381	0.0678	-0.0316	-0.0469	0.00673	0.104*	-0.0943	0.196***
argument/fight	(0.0408)	(0.105)	(0.0419)	(0.0501)	(0.163)	(0.0469)	(0.0587)	(0.0823)	(0.0747)
Control mean	0.0891	0.1759	0.0487	0.1295	0.2407	0.0756	0.0136	0.0451	0.0000
Probability to be involved in selling	0.0268**	0.0264	0.0330**	0.0208*	0.000228	0.0376*	0.0249	0.0590	0.00764
Pirate Goods	(0.0113)	(0.0243)	(0.0167)	(0.0117)	(0.00888)	(0.0213)	(0.0210)	(0.0578)	(0.0123)
Control mean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Probability to be involved in selling	0.0129	-0.00473	0.0222	0.0171	-0.00586	0.0303*	0.00413	0.00956	-
Stolen Goods	(0.00831)	(0.00711)	(0.0135)	(0.0118)	(0.00878)	(0.0176)	(0.00501)	(0.0263)	-
Control mean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Probability to Use Alcohol	0.0588	-0.0575	0.148	0.00547	0.00376	0.0565	0.0625	-0.237	0.249*
	(0.0749)	(0.121)	(0.0985)	(0.0683)	(0.101)	(0.0822)	(0.122)	(0.230)	(0.138)
Control mean	0.7475	0.8465	0.7014	0.8057	0.8686	0.7752	0.6389	0.8019	0.5683
Probability to Use Cigarette	-0.0154	-0.0414	0.0175	0.0465	0.0219	0.105	-0.0800	0.0142	-0.100
	(0.0514)	(0.113)	(0.0617)	(0.0634)	(0.171)	(0.0714)	(0.0644)	(0.0323)	(0.0940)
Control mean	0.1658	0.1728	0.1625	0.1755	0.2585	0.1353	0.1477	0.0000	0.2116
F-test of excluded instrument	301.01***	117.94***	154.08***	413.07***	65.38***	295.42***	178.05***	25.76***	102.69***
Observations	718	318	400	487	202	285	231	116	115
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 40 – IV estimation on Crime and Risky Behavior Outcomes, by course modality and sex

DEPENDENT VARIABLES	Full Sample			Men			Women		
	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent	All Subscribers	Concomitant	Subsequent
Probability to Use Marijuana	0.00671	0.0261	0.0304	0.0105	0.0505	0.0228	0.00708	-0.0374	0.0135
	(0.0287)	(0.0855)	(0.0309)	(0.0396)	(0.139)	(0.0362)	(0.0314)	(0.0352)	(0.0393)
Control mean	0.0448	0.0818	0.0276	0.0615	0.1223	0.0321	0.0136	0.0000	0.0195
Probability to Use Other Drugs	0.00745	0.0310	-0.00724	0.0136	0.0540	-0.00216	-0.00577	-0.00697	-0.0212
	(0.0128)	(0.0289)	(0.0148)	(0.0158)	(0.0433)	(0.0151)	(0.0248)	(0.00832)	(0.0304)
Control mean	0.0085	0.0000	0.0124	0.0057	0.0000	0.0085	0.0136	0.0000	0.0195
Probability to Use Alcohol More than	0.0283	0.0808	0.0270	0.0542*	0.150	0.0350	-0.00319	-0.0111	-
Twice a Week	(0.0193)	(0.0615)	(0.0241)	(0.0301)	(0.109)	(0.0318)	(0.00586)	(0.0147)	-
Control mean	0.0141	0.0208	0.0109	0.0216	0.0312	0.0170	0.0000	0.0000	0.0000
Probability of Binge Drinking	0.0360	-0.0415	0.0912	-0.00799	0.0432	0.0502	0.0907	-0.224	0.120
	(0.0596)	(0.143)	(0.0649)	(0.0772)	(0.172)	(0.0840)	(0.0957)	(0.277)	(0.0842)
Control mean	0.1906	0.3012	0.1391	0.2526	0.3502	0.2053	0.0747	0.2023	0.0195
Probability to Drink and Drive	0.0898	-0.0606	0.182**	0.0505	-0.128	0.182*	0.190***	0.0472	0.250***
	(0.0698)	(0.119)	(0.0897)	(0.0925)	(0.176)	(0.106)	(0.0701)	(0.0689)	(0.0960)
Control mean	0.2914	0.3011	0.2868	0.4326	0.4502	0.4241	0.0273	0.0000	0.0391
F-test of excluded instrument	301.01***	117.94***	154.08***	413.07***	65.38***	295.42***	178.05***	25.76***	102.69***
Observations	718	318	400	487	202	285	231	116	115
Lottery Dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
Note: Specification included control variables for sex, age, color and parental education. To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 41 – Descriptive Statistics for Women in Subsequent Modality

	Control		Treatment	
Variables	Obs.	Mean	Obs.	Mean
Age	30	30.2	85	27.4
Highest Education				
Secondary School (inc.)	30	0.033	85	0.012
Secondary School (comp.)	30	0.567	85	0.588
Tertiary School (inc.)	30	0.100	85	0.082
Tertiary School (ongoing.)	30	0.133	85	0.200
Tertiary School (comp.)	30	0.100	85	0.059
Graduate	30	0.067	85	0.059
Mother's Highest Grade				
Primary School (inc.)	30	0.567	85	0.471
Primary School	30	0.233	85	0.224
Secondary School	30	0.200	85	0.235
Tertiary School	30	0.000	85	0.047
Not Informed	30	0.000	85	0.024
Father's Highest Grade				
Primary School (inc.)	30	0.500	85	0.518
Primary School	30	0.300	85	0.153
Secondary School	30	0.200	85	0.224
Tertiary School	30	0.000	85	0.035
Not Informed	30	0.000	85	0.071
Child	30	0.467	85	0.376
Employment Status	30	0.633	85	0.882
Formal Work Status	30	0.500	85	0.753
Wage	19	1336	74	1494
White	30	0.833	85	0.706
Other Color	30	0.167	85	0.294
Household Condition				
Head of Household	30	0.367	85	0.224
Spouse	30	0.400	85	0.435
Son/daughter	30	0.167	85	0.271
Other	30	0.067	85	0.071

Source: author's elaboration using survey data.

7.2.3 The Use of Substances and Academic Achievement

In this subsection, we aim to investigate the correlation between use of substances and academic achievement. According to the existing literature, young people who are more connected with the school have better academic achievement and lower rates of risky behavior. Thus, they claim that there is a negative relation between academic achievement and use of substances (Chowdry et al, 2009, Chowdry, 2009, Potter and Lindberg, 2000, Oze, 2005, Bond et al, 2007, Fredricks et al, 2004). We ran OLS estimations in order to analyze this relationship in our dataset. For those who we had access to their general school grades, we analyzed the relationship between individual's grade with his use of different substances. We made estimations for Math and Portuguese. We constructed a dummy variable which is equal to one for those concomitant students who reported to be less than 18 years old when using such substance for the first time and zero otherwise. We also included control variables such as gender, age, color and parental education. Robust standard errors are reported in parentheses. Grades are standardized by the grade distribution of the sample, and all estimations include school fixed effect. Table 42 displays results for the OLS regression investigating the correlation between alcohol use and academic achievement. We note no statistically significant correlation for any subject.

Table 42 – OLS estimation for correlation between alcohol use and academic achievement

VARIABLES	Math			Portuguese		
	All Subscribers	Men	Women	All Subscribers	Men	Women
Alcohol Use	-0.200	-0.228	0.00649	-0.124	-0.264	0.0129
	(0.155)	(0.245)	(0.245)	(0.167)	(0.282)	(0.243)
Male	-0.0561			-0.274*		
	(0.158)			(0.163)		
Age	0.0236	0.0368	0.0175	0.0131	0.158	-0.0758
	(0.0727)	(0.0977)	(0.145)	(0.0834)	(0.0961)	(0.191)
White	0.0445	-0.191	0.602*	-0.0492	-0.0422	0.165
	(0.149)	(0.201)	(0.310)	(0.139)	(0.197)	(0.255)
Constant	-0.691	-0.800	-0.612	0.229	-2.821	2.220
	(1.479)	(2.021)	(2.562)	(1.728)	(2.062)	(3.347)
Observations	290	183	107	290	183	107
R-squared	0.307	0.393	0.449	0.346	0.416	0.480
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 43 shows results for the OLS regression investigating the correlation between the cigarette use and academic achievement. As well as for the alcohol use, we observe no statistically significant correlation between the cigarette use and academic achievement neither for math, neither for portuguese.

Table 43 – OLS estimation for correlation between the cigarette use and academic achievement

VARIABLES	Math			Portuguese		
	All Subscribers	Men	Women	All Subscribers	Men	Women
Cigarette Use	-0.206	-0.162	0.120	0.0243	0.0149	0.135
	(0.251)	(0.302)	(0.611)	(0.155)	(0.188)	(0.367)
Male	-0.0554			-0.294*		
	(0.157)			(0.160)		
Age	0.0271	0.0425	0.0180	0.0160	0.166*	-0.0752
	(0.0734)	(0.0964)	(0.144)	(0.0839)	(0.0966)	(0.190)
White	0.0362	-0.206	0.627**	-0.0420	-0.0544	0.192
	(0.149)	(0.202)	(0.295)	(0.139)	(0.197)	(0.254)
Constant	-0.937	-1.091	-0.637	0.0780	-3.171	2.195
	(1.491)	(1.949)	(2.593)	(1.738)	(2.068)	(3.417)
Observations	290	183	107	290	183	107
R-squared	0.306	0.391	0.449	0.344	0.408	0.481
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 44 shows results for the OLS regression investigating the correlation between the marijuana use and academic achievement. For this substance, we find a negative and statistically significant correlation between the marijuana use and academic achievement for math grades. And it seems that this negative correlation is stronger for men.

Table 44 – OLS estimation for correlation between the marijuana use and academic achievement

VARIABLES	Math			Portuguese		
	All Subscribers	Men	Women	All Subscribers	Men	Women
Marijuana Use	-0.674*	-0.734*	-0.221	-0.0720	-0.105	0.206
	(0.345)	(0.395)	(0.848)	(0.179)	(0.206)	(0.440)
Male	-0.0523			-0.294*		
	(0.155)			(0.164)		
Age	0.0277	0.0413	0.0218	0.0172	0.163*	-0.0833
	(0.0734)	(0.0966)	(0.146)	(0.0838)	(0.0952)	(0.193)
White	0.0415	-0.201	0.593**	-0.0580	-0.0595	0.190
	(0.147)	(0.201)	(0.280)	(0.139)	(0.192)	(0.245)
Constant	-0.903	-0.994	-0.621	0.0653	-3.139	2.325
	(1.491)	(1.967)	(2.645)	(1.742)	(2.056)	(3.488)
Observations	290	183	107	290	183	107
R-squared	0.303	0.385	0.450	0.320	0.393	0.463
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Finally, Table 45 displays results for the OLS regression investigating the correla-

tion between the other drugs use and academic achievement. For this variable, we note no statistically significant correlation between other drugs use and academic achievement.

Table 45 – OLS estimation for correlation between the other drugs use and academic achievement

VARIABLES	Math			Portuguese		
	All Subscribers	Men	Women	All Subscribers	Men	Women
Other Drugs Use	0.174	-0.366	1.022*	0.418	0.0996	0.581
	(0.265)	(0.310)	(0.601)	(0.293)	(0.336)	(0.568)
Male	-0.108			-0.304*		
	(0.158)			(0.164)		
Age	0.0405	0.0574	0.0179	0.0241	0.170*	-0.0859
	(0.0741)	(0.0951)	(0.145)	(0.0844)	(0.0950)	(0.193)
White	0.0749	-0.156	0.620**	-0.0538	-0.0501	0.176
	(0.146)	(0.197)	(0.301)	(0.138)	(0.191)	(0.239)
Constant	-1.141	-1.369	-0.517	-0.0695	-3.281	2.402
	(1.517)	(1.949)	(2.641)	(1.757)	(2.048)	(3.475)
Observations	290	183	107	290	183	107
R-squared	0.278	0.351	0.463	0.324	0.393	0.466
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Overall, our estimations indicate that there is evidence, in our sample, for a negative and statistically significant correlation between academic achievement in math and the use of marijuana, in line with the evidences reported in the literature. However, we observe no evidences of negative and statistically significant correlation between academic achievement and other substances, such as alcohol, cigarette and other drugs.

7.2.4 The Role of Peers in the Use of Substances

According to Chowdry (2009), one's environment and peers play a key role on behavioral outcomes (Duncan et al, 1968, Kandel, 1980, Crane, 1991, Mayer, 1991). Case and Katz (1991), investigated the relationship between risky behavior and peer effects, and found that participation in activities such as drug abuse and criminality, for example, were strongly influenced by the participation of family and neighborhood peers in those activities. In this subsection, we do not attempt to estimate the effect of peers on use of substances, we only aim to provide evidences that there is a correlation between one's use of substances and the use by their peers.

In our survey, we asked to our respondents if their neighbors and friends have used a variety of substances. For each substance, we constructed a dummy variable that is equal to one if one answered positively for the use of that substance by his peers, and zero otherwise. We ran OLS estimations in order to analyze this relationship in our dataset. We also include in these estimations control variables such as gender, age, color and parental education. Robust standard errors are reported in parentheses. We make estimations considering our full sample and considering only the subsample of young people, that is, individuals under 25 years old, giving the importance of environment and peers in this age span, according to the literature. We also made estimations splitting the sample by gender. Table 46 displays results for the OLS regression investigating the correlation between the alcohol use and the peer's use of alcohol. We find a positive and statistically significant correlation between the alcohol use and the peer's use of alcohol for all subsamples.

Table 46 – OLS estimation for correlation between the alcohol use and the peers' use of alcohol

VARIABLES	Full Sample	Full Sample < 25	Men	Men < 25	Women	Women < 25
Peers' Alcohol Use	0.206***	0.182***	0.168***	0.132***	0.278***	0.255***
	(0.0355)	(0.0430)	(0.0416)	(0.0494)	(0.0656)	(0.0790)
Male	0.0858***	0.0733**				
	(0.0323)	(0.0364)				
Age	-0.00352	0.0200**	-0.00150	0.0131	-0.00630	0.0364**
	(0.00252)	(0.00857)	(0.00275)	(0.0102)	(0.00576)	(0.0169)
White	0.0217	0.0136	0.0598	0.0703	-0.0525	-0.120*
	(0.0322)	(0.0360)	(0.0381)	(0.0433)	(0.0587)	(0.0647)
Constant	0.672***	0.222	0.696***	0.431*	0.763***	-0.0404
	(0.0830)	(0.196)	(0.0950)	(0.235)	(0.172)	(0.360)
Observations	718	453	487	298	231	155
R-squared	0.081	0.084	0.063	0.067	0.130	0.156
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 47 shows results for the OLS regression investigating the correlation between the cigarette use and the peer's use of cigarette. Also for cigarette, we see a positive and statistically significant correlation between one's cigarette use and his peer's use of cigarette for all subsamples.

Table 47 – OLS estimation for correlation between the cigarette use and the peers' use of cigarette

VARIABLES	Full Sample	Full Sample < 25	Men	Men < 25	Women	Women < 25
Peers' Cigarette Use	0.113***	0.128***	0.133***	0.147***	0.0728**	0.0853**
	(0.0243)	(0.0296)	(0.0326)	(0.0412)	(0.0319)	(0.0343)
Male	0.0810***	0.0898***				
	(0.0252)	(0.0298)				
Age	0.00291	0.0101	0.00119	0.00643	0.00730*	0.0183
	(0.00229)	(0.00878)	(0.00266)	(0.0111)	(0.00431)	(0.0123)
White	-0.0202	-0.0267	-0.00130	0.00545	-0.0571	-0.103*
	(0.0297)	(0.0374)	(0.0383)	(0.0474)	(0.0473)	(0.0594)
Constant	-0.0522	-0.194	0.0463	-0.0706	-0.101	-0.256
	(0.0644)	(0.179)	(0.0778)	(0.231)	(0.121)	(0.262)
Observations	718	453	487	298	231	155
R-squared	0.052	0.073	0.042	0.051	0.076	0.148
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 48 presents our results for the OLS regression investigating the correlation between one's use of marijuana and his peer's use of marijuana. As well as the two first substances, we find a positive and statistically significant correlation between one's marijuana use and his peer's use of marijuana for all subsamples.

Table 48 – OLS estimation for correlation between the marijuana use and the peers' use of marijuana

VARIABLES	Full Sample	Full Sample < 25	Men	Men < 25	Women	Women < 25
Peers' Marijuana Use	0.237***	0.264***	0.264***	0.280***	0.211**	0.245***
	(0.0460)	(0.0546)	(0.0561)	(0.0684)	(0.0817)	(0.0915)
Male	0.0349**	0.0312				
	(0.0161)	(0.0217)				
Age	-0.000670	-0.000783	0.000200	-0.00530	-0.00126	0.00995
	(0.00101)	(0.00610)	(0.00135)	(0.00713)	(0.00114)	(0.00940)
White	-0.0141	-0.0349	-0.0238	-0.0403	0.0113	-0.0256
	(0.0195)	(0.0262)	(0.0253)	(0.0326)	(0.0283)	(0.0423)
Constant	0.0265	0.0437	0.0286	0.146	0.0511	-0.136
	(0.0338)	(0.129)	(0.0414)	(0.160)	(0.0536)	(0.200)
Observations	718	453	487	298	231	155
R-squared	0.129	0.159	0.135	0.162	0.162	0.226
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Finally, Table 49 displays our results for the OLS regression investigating the correlation between one's use of other drugs and his peer's use of other drugs. Differently from our previous results in this subsection, we find none statistically significant correlation between one's use of other drugs and his peer's use of other drugs for any subsample. Maybe because we have too few variation among the responses for these two variables: that is, the variable identifying one's use of other drugs and the variable identifying his peers' use of other drugs.

Table 49 – OLS estimation for correlation between other drugs use and peers' use of other drugs

VARIABLES	Full Sample	Full Sample < 25	Men	Men < 25	Women	Women < 25
Peers' Other Drugs Use	0.0624	0.0302	0.0807	0.0435	-0.0379	-0.0458
	(0.0415)	(0.0348)	(0.0497)	(0.0445)	(0.0236)	(0.0320)
Male	-0.00166	-0.00961				
	(0.00885)	(0.0101)				
Age	0.000924	0.00248	0.00160*	0.000244	-0.000439	0.00767
	(0.000668)	(0.00293)	(0.000897)	(0.00150)	(0.000857)	(0.00672)
White	-0.0158	0.00960*	-0.0310*	0.00578	0.0216*	0.0133
	(0.0122)	(0.00526)	(0.0163)	(0.00463)	(0.0127)	(0.0115)
Constant	-0.000155	-0.0408	-0.0171	-0.0117	0.0224	-0.127
	(0.0186)	(0.0561)	(0.0209)	(0.0341)	(0.0276)	(0.129)
Observations	718	453	487	298	231	155
R-squared	0.034	0.052	0.048	0.053	0.085	0.137
Parental Education Dummies	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Overall, our estimations indicate that there is evidence in our sample for a positive and statistically significant correlation between one's use of substances and the use of substances by his peers'. These results are in line with the existing evidences in the literature. These kind of evidences, provided in the last two subsections, are gaining importance in recent years, specially because of the increasing interest by economists in understanding the importance of personality in economic outcomes[3].

[3] Heckman et al. (2006), Borghans et al.(2008), Brunello and Schlotter (2011), Carneiro et al. (2007), Carneiro (2010)

8 Conclusion

There is no consensus, in academic literature or public debate, about returns of investments in Vocational Education and Training. Evidence suggests that the effect of VET courses varies across countries, depending on their institutional arrangements to provide this modality of education and on the design of the program. For Brazil, this discussion is especially relevant, since there has been a substantial increase of public resources invested in Vocational Education and Training programs in recent years.

The main challenge of evaluating the effect of VET courses on labor market outcomes is to isolate the impacts exclusively assigned to the abilities developed by the program from those that are inherent to the individual. There are unobservable characteristics that can influence individuals' decisions and generate selection bias on impact estimation. The contribution of this study was to provide, using primary data collected through a survey, rigorous evidence of the impact of PRONATEC on its beneficiaries. For this, we took advantage of the selection criteria for the program implemented by S-System in the Brazilian state of Santa Catarina. These providers selected their beneficiaries through a lottery when there was excess of demand for technical courses.

Moreover, through our unique dataset, we were able to evaluate the effect of PRONATEC not only on economic dimension, but also on human capital, socio-emotional, crime and risky behavior ones. This is a valuable contribution since there is a lack of evidence in the literature about the impact of VET programs on these outcomes.

Our estimations were based on an ITT and LATE strategies. We assigned to treatment group individuals who received an offer to enroll in technical education courses at SENAI and SENAC. Those who were randomly assigned to the waiting list and did not receive an offer to enroll in technical courses composed our control group. Data on our outcomes were collected through a survey at least 6 months after graduating technical education. The same questions were carried out for treatment and control groups individuals.

Our main results show that PRONATEC had an effect on increasing beneficiaries' probability to work in the course area (15 p.p.), probability to graduate technical education at SENAI (60 p.p.) and probability to graduate any technical education (19 p.p.). We also find a negative impact on the probability to graduate other technical education (22 p.p.). Considering the concomitant modality, results pointed towards a negative impact of PRONATEC on the probability to graduate regular school at correct age (15 p.p.) and on academic achievement in Portuguese in regular education (0.1). Moreover, we find a positive impact of PRONATEC on the probability to sell pirate goods (3 p.p.). Considering our full

sample, we found no impact of PRONATEC on other labor market, socio-emotional, crime and risky behavior outcomes.

Our results also indicated that women were likely to benefit more from PRONATEC. Considering the subsample of women, we found positive impacts of PRONATEC on formal work probability (43 p.p.), probability to work in course area (11 p.p.) and on wages (43%). Furthermore, for socio-emotional skills, we found a positive effect of PRONATEC on extraversion (0.96 σ). We also noticed a positive impact on the probability to be involved in an argument/fight (10 p.p.). Linking the results found for women in labor market and socio-emotional dimensions, we speculated that this positive impact on the probability to be involved in an argument could reflect an increase on women's empowerment. The increase on their extraversion and labor market outcomes could have an impact on changing women's perspective: enhancing their self-confidence and pride, as well as their notion of taking more control over their lives. Unfortunately, we had no data about women's household condition, or even about their marriage status before the program, which would have been useful to allow for a more in-depth analysis of this empowerment channel.

We also made robustness checks on the statistical significance of our ITT treatment coefficients through randomized tests. For those coefficients that we found significant treatment effect using conventional test, results showed that they were also able to reject the null hypotheses of no treatment effect anywhere. Therefore, results from randomized tests corroborated the statistical significance of our treatment coefficients. Furthermore, we had a high rate of attrition. Concerned with this problem, we carried out nonparametric extreme bounds (Lee, 2007). Unfortunately, our results generated wide bounds, which were not very informative.

Overall, our results indicated that, at least in the short-run, there was no effect of PRONATEC on conventional labor market outcomes, such as employment and formal work probability and wages. Also, we found no effect in socio-emotional, crime and risky behavior dimensions. Nonetheless, our results pointed towards a possible heterogeneity of the program, since we found positive and significant effects of PRONATEC on women, specially in labor market (formal work probability and wage) and socio-emotional (extraversion) outcomes.

Bibliography

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J. & Pathak, P. A. (2011), 'Accountability and flexibility in public schools: Evidence from boston's charters and pilots', *The Quarterly Journal of Economics* **126**(2), 699–748.

Adelman, S., Gilligan, D. & Lehrer, K. (2008), *How effective are food for education programs?: A critical assessment of the evidence from developing countries*, Vol. 9, Intl Food Policy Res Inst.

Agan, A. Y. (2011), 'Non-cognitive skills and crime', *University of Chicago, unpublished* pp. 972–1059.

Agnew, R., Brezina, T., Wright, J. P. & Cullen, F. T. (2002), 'Strain, personality traits, and delinquency: Extending general strain theory', *Criminology* **40**(1), 43–72.

Aizer, A. (2010), 'Poverty, violence, and health the impact of domestic violence during pregnancy on newborn health', *Journal of Human Resources* **46**(3), 518–538.

Almeida, R., Anazawa, L., Menezes Filho, N. & Vasconsellos, L. (2014), 'Retornos da educação profissional e técnica no brasil', *Brazil. Mimeographed document* .

Almlund, M., Duckworth, A. L., Heckman, J. J. & Kautz, T. D. (2011), Personality psychology and economics, Technical report, National Bureau of Economic Research.

Anderson, D. M. (2014), 'In school and out of trouble? the minimum dropout age and juvenile crime', *Review of Economics and Statistics* **96**(2), 318–331.

Angrist, D. & Pischke, J.-S. (2009), 'Mostly harmless econometrics princeton', *Princeton University Press.[13] Epifani, Paolo, and Gino Gancia (2011)."Trade, markup heterogeneity and misallocations." Journal of International Economics* **83**(1), 1–13.

Angrist, J. D. & Imbens, G. W. (1995), 'Two-stage least squares estimation of average causal effects in models with variable treatment intensity', *Journal of the American statistical Association* **90**(430), 431–442.

Assunção, J. & Gonzaga, G. (2010), 'Educação profissional no brasil: inserção e retorno', *SENAI. Série* .

Attanasio, O., Kugler, A. & 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. & Mount, M. K. (1991), 'The big five personality dimensions and job performance: a meta-analysis', *Personnel psychology* **44**(1), 1–26.

Barro, R. J. (1991), 'Economic growth in a cross section of countries', *The quarterly journal of economics* **106**(2), 407–443.

Barro, R. J. (2001), 'Human capital and growth', *The American Economic Review* **91**(2), 12–17.

Barros, R., Franco, S., Grosner, D., Mendonça, R. & Rosalém, A. (2011), 'Educação técnica e distribuição de renda no espírito santo', *Revista Brasileira de Monitoramento e Avaliação* (1), 104–135.

Behaghel, L., Crépon, B., Gurgand, M. & Le Barbanchon, T. (2015), 'Please call again: Correcting nonresponse bias in treatment effect models', *Review of Economics and Statistics* **97**(5), 1070–1080.

Bellisle, F. (2004), 'Effects of diet on behaviour and cognition in children', *British Journal of Nutrition* **92**(S2), S227–S232.

Belot, M. & James, J. (2011), 'Healthy school meals and educational outcomes', *Journal of health economics* **30**(3), 489–504.

Benton, D. (2007), 'The impact of diet on anti-social, violent and criminal behaviour', *Neuroscience & Biobehavioral Reviews* **31**(5), 752–774.

Berthelon, M. E. & Kruger, D. I. (2011), 'Risky behavior among youth: Incapacitation effects of school on adolescent motherhood and crime in chile', *Journal of public economics* **95**(1), 41–53.

Bishop, J. H. & Mane, F. (2005), 'Raising academic standards and vocational concentrators: Are they better off or worse off?', *Education Economics* **13**(2), 171–187.

Blattman, C., Jamison, J. C. & Sheridan, M. (2017), 'Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia', *The American Economic Review* **107**(4), 1165–1206.

Bond, L., Butler, H., Thomas, L., Carlin, J., Glover, S., Bowes, G. & Patton, G. (2007), 'Social and school connectedness in early secondary school as predictors of late teenage substance use, mental health, and academic outcomes', *Journal of Adolescent Health* **40**(4), 357–e9.

Bowles, S., Gintis, H. & Osborne, M. (2001), 'Incentive-enhancing preferences: Personality, behavior, and earnings', *The American Economic Review* **91**(2), 155–158.

Caldes, N. & Ahmed, A. (2004), 'Food for education: A review of program impacts', *International Food Policy Research Institute, Washington, DC* .

Card, D. (1999), 'The causal effect of education on earnings', *Handbook of labor economics* **3**, 1801–1863.

Carneiro, P. (2009), 'Lessons from the technology of skill formation: implications for education policy'.

Carneiro, P., Dearden, L. & Vignoles, A. (2010), The economics of vocational education and training.

Carneiro, P. M. & Heckman, J. J. (2003), 'Human capital policy'.

Cattan, S. (2010), 'Heterogeneity and selection in the labor market', *PhD thesis, University of Chicago* .

Chen, X., Flores, C. & Flores-Lagunes, A. (2012), Bounds on population average treatment effects with an instrumental variable, Technical report, mimeo, University of Miami, Dept. of Economics.

Chioda, L., De Mello, J. M. & Soares, R. R. (2016), 'Spillovers from conditional cash transfer programs: Bolsa família and crime in urban brazil', *Economics of Education Review* **54**, 306–320.

Chiteji, N. (2010), 'Time-preference, non-cognitive skills and well-being across the life course: Do non-cognitive skills encourage healthy behavior?', *The American economic review* **100**(2), 200.

Crome, I. B. (1999), 'Substance misuse and psychiatric comorbidity: Towards improved service provision', *Drugs: education, prevention and policy* **6**(2), 151–174.

Cunha, F. & Heckman, J. J. (2009), 'The economics and psychology of inequality and human development', *Journal of the European Economic Association* **7**(2-3), 320–364.

Curto, V. E. & Fryer Jr, R. G. (2014), 'The potential of urban boarding schools for the poor: Evidence from seed', *Journal of Labor Economics* **32**(1), 65–93.

de Chaisemartin, C. & Behaghel, L. (2015), 'Next please! a new definition of the treatment and control groups for randomizations with waiting lists', *arXiv preprint arXiv:1511.01453* .

Dearden, L., McIntosh, S., Myck, M. & Vignoles, A. (2002), 'The returns to academic and vocational qualifications in britain', *Bulletin of economic research* **54**(3), 249–274.

Deming, D. J. (2011), 'Better schools, less crime?', *The Quarterly Journal of Economics* p. qjr036.

Dohmen, T. & Falk, A. (2010), 'You get what you pay for: Incentives and selection in the education system', *The Economic Journal* **120**(546), F256–F271.

Duflo, E., Glennerster, R. & Kremer, M. (2007), 'Using randomization in development economics research: A toolkit', *Handbook of development economics* **4**, 3895–3962.

Eichhorst, W., Rodríguez-Planas, N., Schmidl, R. & Zimmermann, K. F. (2015), 'A road map to vocational education and training in industrialized countries', *ILR Review* p. 0019793914564963.

Eichhorst, W., Rodriguez Planas, N., Schmidl, R. & Zimmermann, K. F. (n.d.), 'A roadmap to vocational education and training systems around the world'.

Falch, T. & Sandgren Massih, S. (2012), 'The effect of education on cognitive ability', *Economic Inquiry* **49**(3), 838–856.

Feinstein, L., Duckworth, K. & Sabates, R. (2008), *Education and the family: Passing success across the generations*, Routledge.

Feinstein, L. & Sabates, R. (2008), 'Skills and social productivity', *Not Just the Economy. The Public value of adult learning, C. Flint and C. Hughes (eds.) Leicester: NIACE* pp. 50–66.

Figlio, D. N. & Winicki, J. (2005), 'Food for thought: the effects of school accountability plans on school nutrition', *Journal of public Economics* **89**(2), 381–394.

Fisher, R. (1935), 'The design of experiments. 252 pp', *Edinburgh and London* .

for Economic Co-operation, O. & Development (2015), 'Skills for social progress: the power of social and emotional skills'.

Friedlander, D., Greenberg, D. H. & Robins, P. K. (1997), 'Evaluating government training programs for the economically disadvantaged', *Journal of economic literature* **35**(4), 1809–1855.

Friedman, H. S. & Kern, M. L. (2014), 'Personality, well-being, and health', *Annual Review of Psychology* **65**, 719–742.

Frölich, M. (2007), 'Propensity score matching without conditional independence assumption—with an application to the gender wage gap in the united kingdom', *The Econometrics Journal* **10**(2), 359–407.

Gensowski, M., Heckman, J. & Savelyev, P. (2014), 'The effects of education, personality, and iq on earnings of high-ability individuals', *Unpublished manuscript, University of Chicago, Department of Economics* .

Grogger, J. (2000), 'An economic model of recent trends in violence', *The crime drop in America* pp. 266–287.

Hanushek, E. A. & Kimko, D. D. (2000), 'Schooling, labor-force quality, and the growth of nations', *American economic review* pp. 1184–1208.

Hanushek, E. A., Woessmann, L. & Zhang, L. (2011), General education, vocational education, and labor-market outcomes over the life-cycle, Technical report, National Bureau of Economic Research.

Hanushek, E. A. et al. (2006), 'Does educational tracking affect performance and inequality? differences-in-differences evidence across countries', *The Economic Journal* **116**(510), C63–C76.

Harhoff, D. & Kane, T. J. (1997), 'Is the german apprenticeship system a panacea for the us labor market?', *Journal of population economics* **10**(2), 171–196.

Hayes, A. F. (2000), 'Randomization tests and the equality of variance assumption when comparing group means'.

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica 47* pp. 153 – 161.

Heckman, J. J. (2007), 'The economics, technology, and neuroscience of human capability formation', *Proceedings of the national Academy of Sciences* **104**(33), 13250–13255.

Hill, P. L., Roberts, B. W., Grogger, J. T., Guryan, J. & Sixkiller, K. (2011), Decreasing delinquency, criminal behavior, and recidivism by intervening on psychological factors other than cognitive ability: A review of the intervention literature, Technical report, National Bureau of Economic Research.

Hjalmarsson, R. (2008), 'Criminal justice involvement and high school completion', *Journal of Urban Economics* **63**(2), 613–630.

Hogan, J. & Holland, B. (2003), 'Using theory to evaluate personality and job-performance relations: a socioanalytic perspective.'.

Ikesako, H. & Miyamoto, K. (2015), 'Fostering social and emotional skills through families, schools and communities'.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average'.

Indicators, O. (2007), 'Education at a glance 2007', *Table B1. 1b, www. oecd. org/-dataoecd/36/4/40701218. pdf* p. 187.

Indicators, O. (n.d.), 'Education at a glance 2016'.
URL: */content/book/eag-2016-en*

Jacob, B. A. & Lefgren, L. (2003), Are idle hands the devil's workshop? incapacitation, concentration and juvenile crime, Technical report, National Bureau of Economic Research.

Jacoby, E., Cueto, S. & Pollitt, E. (1996), 'Benefits of a school breakfast programme among andean children in huaraz, peru', *FOOD AND NUTRITION BULLETIN-UNITED NATIONS UNIVERSITY-* **17**, 54–64.

Jamieson, A., Sabates, R., Woodley, A. & Feinstein, L. (2009), 'The benefits of higher education study for part-time students', *Studies in Higher Education* **34**(3), 245–262.

Jencks, C. et al. (1979), 'Who gets ahead? the determinants of economic success in america.'.

Jenkins, A., Greenwood, C. & 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.

Jukes, M. C., Drake, L. J. & Bundy, D. A. (2007), *School health, nutrition and education for all: levelling the playing field*, CABI.

Kane, T. J. & 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* pp. 205–221.

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B. & Borghans, L. (2014), Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success, Technical report, National Bureau of Economic Research.

Kazianga, H., De Walque, D. & Alderman, H. (2009), 'Educational and health impacts of two school feeding schemes: Evidence from a randomized trial in rural burkina faso', *World Bank Policy Research Working Paper* (4976).

Kuhn, P. & Weinberger, C. (2005), 'Leadership skills and wages', *Journal of Labor Economics* **23**(3), 395–436.

Lambert, J., Agostoni, C., Elmadfa, I., Hulshof, K., Krause, E., Livingstone, B., Socha, P., Pannemans, D. & Samartín, S. (2004), 'Dietary intake and nutritional status of children and adolescents in europe', *British Journal of Nutrition* **92**(S2), S147–S211.

Lee, D. S. (2009), 'Training, wages, and sample selection: Estimating sharp bounds on treatment effects', *The Review of Economic Studies* **76**(3), 1071–1102.

Lerman, R. I. (2012), Can the united states expand apprenticeship? lessons from experience, Technical report, IZA Policy Paper.

Lindberg, L. D., Boggess, S., Porter, L. & Williams, S. (2000), 'Teen risk-taking: A statistical portrait.'.

Lochner, L. & Moretti, E. (2004), 'The effect of education on crime: Evidence from prison inmates, arrests, and self-reports', *The American Economic Review* **94**(1), 155–189.

Luallen, J. (2006), 'School's out... forever: A study of juvenile crime, at-risk youths and teacher strikes', *Journal of urban economics* **59**(1), 75–103.

Luepker, R. V., Perry, C. L., McKinlay, S. M., Nader, P. R., Parcel, G. S., Stone, E. J., Webber, L. S., Elder, J. P., Feldman, H. A., Johnson, C. C. et al. (1996), 'Outcomes of a field trial to improve children's dietary patterns and physical activity: the child and adolescent trial for cardiovascular health (catch)', *Jama* **275**(10), 768–776.

Machin, S., Marie, O. & Vujić, S. (2011), 'The crime reducing effect of education*', *The Economic Journal* **121**(552), 463–484.

Machin, S. & Meghir, C. (2004), 'Crime and economic incentives', *Journal of Human Resources* **39**(4), 958–979.

Machin, S. & Vignoles, A. (2005), *What's the Good of Education?: The Economics of Education in the UK*, Princeton University Press.

Malamud, O. & 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.

Mankiw, N. G., Romer, D. & Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *The quarterly journal of economics* **107**(2), 407–437.

McCann, D., Barrett, A., Cooper, A., Crumpler, D., Dalen, L., Grimshaw, K., Kitchin, E., Lok, K., Porteous, L., Prince, E. et al. (2007), 'Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: a randomised, double-blinded, placebo-controlled trial', *The Lancet* **370**(9598), 1560–1567.

McEwan, P. J. (2013), 'The impact of chile's school feeding program on education outcomes', *Economics of Education Review* **32**, 122–139.

Mendolia, S. & Walker, I. (2013), 'of laborthe effect of non-cognitive traits on health behaviours in adolescence'.

Monteiro, J. & Rocha, R. (2013), Tráfico de drogas e desempenho escolar no rio de janeiro, Technical report.

Nastari, R. L. B. (n.d.), Três ensaios em economia da educação, PhD thesis.

Neri, M. C. (2010), 'A educação profissional e você no mercado de trabalho', *Rio de Janeiro: FGV/CPS* .

Neuman, S. & Ziderman, A. (1999), 'Vocational education in israel: wage effects of the voced-occupation match', *Journal of human resources* pp. 407–420.

Ñopo, H., Saavedra-Chanduví, J. & Robles, M. (2007), 'Occupational training to reduce gender segregation: the impacts of projoven'.

Nyhus, E. K. & 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.

Oreopoulus, P., von Wachter, T. & Heisz, A. (2007), 'The short-and long-term career effects of graduating in a recession1'.

Ozer, E. M., Adams, S. H., Lustig, J. L., Gee, S., Garber, A. K., Gardner, L. R., Rehbein, M., Addison, L. & Irwin, C. E. (2005), 'Increasing the screening and counseling of adolescents for risky health behaviors: a primary care intervention', *Pediatrics* **115**(4), 960–968.

Piatek, R. & Pinger, P. (2010), 'Maintaining (locus of) control? assessing the impact of locus of control on education decisions and wages'.

Pollock, N. K. & Martin, C. S. (1999), 'Diagnostic orphans: adolescents with alcohol symptoms who do not qualify for dsm-iv abuse or dependence diagnoses', *American Journal of Psychiatry* .

Poropat, A. E. (2009), 'A meta-analysis of the five-factor model of personality and academic performance.', *Psychological bulletin* **135**(2), 322.

Powell, C. A., Walker, S. P., Chang, S. M. & Grantham-McGregor, S. M. (1998), 'Nutrition and education: a randomized trial of the effects of breakfast in rural primary school children.', *The American journal of clinical nutrition* **68**(4), 873–879.

Quintini, G., Martin, S. et al. (2006), Starting well or losing their way?: the position of youth in the labour market in oecd countries, Technical report, OECD Publishing.

Roberts, P., Rojas Ruiz, H. T., Dréze, J., Sen, A., Braum, J. v., Haen, H., Blanken, J., Sims, L., Kennedy, E., Cogill, B. et al. (2009), Building up the national policy and system for food and nutrition security: the brazilian experienceconstrucción del sistema y de la política de seguridad alimentaria y nutricional: la experiencia brasileña, Technical report, CONSEA.

Rodriguez, C. & Sanchez, F. (2012), 'Armed conflict exposure, human capital investments, and child labor: Evidence from colombia', *Defence and peace economics* **23**(2), 161–184.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* pp. 41–55.

Sabates, R. (2008), 'Educational attainment and juvenile crime area-level evidence using three cohorts of young people', *British Journal of Criminology* **48**(3), 395–409.

Salgado, J. F. (1997), 'The five factor model of personality and job performance in the european community.'.

Schanzenbach, D. W. (2009), 'Do school lunches contribute to childhood obesity?', *Journal of Human Resources* **44**(3), 684–709.

Schulenberg, J., O'Malley, P. M., Bachman, J. G., Wadsworth, K. N. & Johnston, L. D. (1996), 'Getting drunk and growing up: trajectories of frequent binge drinking during the transition to young adulthood.', *Journal of studies on alcohol* **57**(3), 289–304.

Schwartzman, S. (2014), Academic drift in brazilian education, *in* 'The Forefront of International Higher Education', Springer, pp. 61–72.

SENAI, S. & IEL (2014), 'Relatório anual'.

Shavit, Y. & Muller, W. (1998), *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations.*, ERIC.

Sidaner, E., Balaban, D. & Burlandy, L. (2013), 'The brazilian school feeding programme: an example of an integrated programme in support of food and nutrition security', *Public health nutrition* **16**(06), 989–994.

Silva, J., Gukovas, R. & Caruso, L. (2015), 'The wage returns and employability of vocational training in brazil: Evidence from matched provider-employer administrative data', *Research paper—background paper for this report, World Bank, Washington, DC* .

Sorhaindo, A. & Feinstein, L. (2006), *What is the relationship between child nutrition and school outcomes?[Wider Benefits of Learning Research Report No. 18]*, Centre for Research on the Wider Benefits of Learning, Institute of Education, University of London.

Spear, L. P. (2000), 'The adolescent brain and age-related behavioral manifestations', *Neuroscience & Biobehavioral Reviews* **24**(4), 417–463.

Tansel, A. (1998), General versus vocational high schools and labor market outcomes in turkey, *in* 'Economic Research Forum Working Paper', number 9905.

Teixeira, A. (2016), 'Sinopse estatística da educação básica', *Brasília, DF* .

Vasconcellos, L., Lima, F. C., Fernandes, J. G. & Menezes-Filho, N. (2010), 'Avaliação econômica do ensino médio profissional', *Fundação Itaú Social, Brasilia* .

Vermeersch, C. & Kremer, M. (2005), *School meals, educational achievement, and school competition: evidence from a randomized evaluation*, Vol. 3523, World Bank Publications.

Whaley, S. E., Sigman, M., Neumann, C., Bwibo, N., Guthrie, D., Weiss, R. E., Alber, S. & Murphy, S. P. (2003), 'The impact of dietary intervention on the cognitive development of kenyan school children', *The Journal of nutrition* **133**(11), 3965S–3971S.

White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D. J. & Stouthamer-Loeber, M. (1994), 'Measuring impulsivity and examining its relationship to delinquency.', *Journal of abnormal psychology* **103**(2), 192.

Witte, A. D. & Tauchen, H. (1994), Work and crime: An exploration using panel data, Technical report, National Bureau of Economic Research.

Woessmann, L. (2008), 'Efficiency and equity of european education and training policies', *International Tax and Public Finance* **15**(2), 199–230.

Young, A. (2015), 'Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results', *E, 0: 0–0* .

9 Apendix

Table 51 – Nonparametric Extreme Bounds

	Lower Bound	Upper Bound	Trimming Proportion	Obs.	S.E. Lower Bound	S.E. Upper Bound
Employment Probability	-0.15457	0.184508	0.428	718	0.056	0.038
Formal Work Probability	-0.18979	0.36231	0.428	718	0.075	0.047
Ln (Wage)	-0.23876	0.368632	0.418	556	0.068	0.075
Probability to Work in Course Area	-0.14084	0.34696	0.428	718	0.034	0.063
Probability to Access Tertiary Ed.	-0.42282	0.247511	0.428	718	0.048	0.084
Prob. to Grad. Tech. Ed. at SENAI	0.075101	0.824758	0.428	718	0.087	0.092
Prob. to Grad. Any Tech. Ed.	-0.13296	0.422792	0.428	718	0.076	0.048
Prob. to Grad. Other Tech. Ed.	-0.49805	0.021458	0.428	718	0.049	0.074
Prob. to be involved in argument/fight	-0.07775	0.112636	0.428	718	0.026	0.038
Prob. to be involved in selling Pirate Goods	0.00000	0.03009	0.428	718	0.000	0.009
Probability to Use Alcohol	-0.06647	0.24947	0.428	718	0.057	0.042
Probability to Use Cigarette	-0.15386	0.110874	0.428	718	0.035	0.049
Probability to Use Marijuana	-0.03954	0.053414	0.428	718	0.019	0.026
Probability to Use Other Drugs	-0.00835	0.016923	0.428	718	0.009	0.012
Probability to Binge Drinking	-0.18197	0.206548	0.428	718	0.038	0.059
Probability to Drink and Drive	-0.27185	0.293641	0.428	718	0.044	0.075
Prob. to Use Alcohol More than Twice a Week	-0.01328	0.054999	0.428	718	0.011	0.018
Agreeableness	-0.60767	0.382198	0.392	366	0.151	0.153
Conscientiousness	-0.55188	0.459035	0.392	366	0.166	0.171
Extraversion	-0.563	0.462931	0.391	366	0.159	0.168
Neuroticism	-0.70793	0.457307	0.392	366	0.193	0.176
Locus of Control	-0.63275	0.56495	0.392	366	0.182	0.180
Openness to Experiences	-0.68741	0.425067	0.392	366	0.167	0.173

Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 50 – Descriptive Statistics for Women

	All Subscribers				Concomitant				Subsequent			
	Control		Treatment		Control		Treatment		Control		Treatment	
Variables	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs.	Mean
Age	37	28.2	194	22.8	7	19.6	109	19.2	30	30.2	85	27.4
Highest Education												
Secondary School (inc.)	37	0.027	194	0.057	7	0.000	109	0.092	30	0.033	85	0.012
Secondary School (comp.)	37	0.514	194	0.448	7	0.286	109	0.339	30	0.567	85	0.588
Tertiary School (inc.)	37	0.081	194	0.093	7	0.000	109	0.101	30	0.100	85	0.082
Tertiary School (ongoing.)	37	0.243	194	0.351	7	0.714	109	0.468	30	0.133	85	0.200
Tertiary School (comp.)	37	0.081	194	0.026	7	0.000	109	0.000	30	0.100	85	0.059
Graduate	37	0.054	194	0.026	7	0.000	109	0.000	30	0.067	85	0.059
Mother's Highest Grade												
Illiterate	37	0.514	194	0.371	7	0.286	109	0.294	30	0.567	85	0.471
Primary School	37	0.216	194	0.237	7	0.143	109	0.248	30	0.233	85	0.224
Secondary School	37	0.189	194	0.309	7	0.143	109	0.367	30	0.200	85	0.235
Tertiary School	37	0.081	194	0.072	7	0.429	109	0.092	30	0.000	85	0.047
Not Informed	37	0.000	194	0.010	7	0.000	109	0.000	30	0.000	85	0.024
Father's Highest Grade												
Illiterate	37	0.459	194	0.423	7	0.286	109	0.349	30	0.500	85	0.518
Primary School	37	0.297	194	0.180	7	0.286	109	0.202	30	0.300	85	0.153
Secondary School	37	0.216	194	0.273	7	0.286	109	0.312	30	0.200	85	0.224
Tertiary School	37	0.027	194	0.036	7	0.143	109	0.037	30	0.000	85	0.035
Not Informed	37	0.000	194	0.088	7	0.000	109	0.101	30	0.000	85	0.071
Child	37	0.405	194	0.222	7	0.143	109	0.101	30	0.467	85	0.376
Employment Status	37	0.676	194	0.722	7	0.857	109	0.596	30	0.633	85	0.882
Formal Work Status	37	0.459	194	0.598	7	0.286	109	0.477	30	0.500	85	0.753
Wage	25	1227.000	139	1325.532	6	881.167	65	1133.338	19	1336.211	74	1494.351
White	37	0.811	194	0.737	7	0.714	109	0.761	30	0.833	85	0.706
Other Color	37	0.189	194	0.263	7	0.286	109	0.239	30	0.167	85	0.294
Household Condition												
Head of Household	37	0.297	194	0.139	7	0.000	109	0.073	30	0.367	85	0.224
Spouse	37	0.378	194	0.242	7	0.286	109	0.092	30	0.400	85	0.435
Son/daughter	37	0.270	194	0.557	7	0.714	109	0.780	30	0.167	85	0.271
Other	37	0.054	194	0.062	7	0.000	109	0.056	30	0.067	85	0.071

Table 52 – Linear probability model regressions for conditional mean difference test ("Administrative 2" data) for women, by course modality

VARIABLES	All Subscribers	Concomitant	Subsequent
White	-0.0497	-0.0276	-0.104
	(0.0464)	(0.0357)	(0.0757)
Color Not Declared	-0.253***	-0.207**	-0.241***
	(0.0526)	(0.0971)	(0.0635)
Age	-0.0103	-0.0399***	-0.00514
	(0.00628)	(0.00954)	(0.00687)
Secondary School (incomplete)	-0.0906		
	(0.0798)		
Tertiary School	0.289**		
	(0.116)		
Constant	1.402***	1.986***	1.349***
	(0.229)	(0.242)	(0.147)
Observations	459	290	169
R-squared	0.286	0.513	0.160
Lottery Dummies	yes	yes	yes

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 53 – Linear probability model regressions for conditional mean difference test (Sample data) for women, by course modality

	Women		
VARIABLES	All subscribers	Concomitant	Subsequent
Age	-0.0158***	-0.000374	-0.0162***
	(0.00538)	(0.0654)	(0.00585)
White	-0.0804*	0.0774	-0.134*
	(0.0468)	(0.0654)	(0.0694)
Mother's Highest Education			
Primary School	0.124	0.302	-0.0801
	(0.152)	(0.185)	(0.0926)
Secondary School	0.116	0.307*	-0.117
	(0.149)	(0.168)	(0.104)
Do not Know	0.352**		0.112
	(0.177)		(0.114)
Father's Highest Education			
Primary School	-0.0153	0.247	-0.205
	(0.215)	(0.322)	(0.125)
Secondary School	-0.0578	0.287	-0.306**
	(0.216)	(0.314)	(0.128)
Do not Know	0.0847	0.262	-0.000856
	(0.218)	(0.312)	(0.137)
Constant	1.381***	0.459	1.898***
	(0.113)	(1.393)	(0.219)
Observations	231	116	115
R-squared	0.262	0.362	0.296
Lottery Dummies	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.

Table 54 – OLS estimation for correlation between attrition and treatment status ("Administrative 2" data, women)

VARIABLES	All subscribers	Concomitant	Subsequent
Z_i=1	0.0554	0.0361	0.00850
	(0.112)	(0.0928)	(0.166)
White	0.0325	0.0485	0.00667
	(0.0993)	(0.128)	(0.157)
Color Not Declared	-0.0252	-0.0190	0.0362
	(0.103)	(0.136)	(0.157)
Age	6.37e-05	-0.0336***	0.00216
	(0.00715)	(0.0127)	(0.00812)
Constant	-0.0900	1.722***	-0.0901
	(0.294)	(0.368)	(0.314)
Observations	459	290	169
R-squared	0.158	0.196	0.128
Lottery Dummies	yes	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Note: To address the imbalance created by pooling lottery strata, all equations contain inverse probability reweighting.