

EFFICACY OF VOCATIONAL EDUCATION: A LONGITUDINAL ANALYSIS

RONALDO DE ALBUQUERQUE E ARRAES *
FRANCISCA ZILANIA MARIANO †

Resumo

Este trabalho visa ampliar o debate na literatura da Economia da Educação no Brasil sobre o ensino público, através da avaliação do desempenho dos alunos de Escolas Estaduais de Ensino Profissionalizante (EEEP) do Ceará nos exames de ensino médio (ENEM). Algumas contribuições podem ser consideradas, especialmente a construção de uma base de dados única no país em nível longitudinal, a qual é aplicada em inferência estatística por técnicas de pareamento. Conclui-se que o efeito médio do tratamento das EEEPs se mostrou positivo e significativo em todas as áreas de conhecimento do exame ENEM, destaque para Redação, cujo efeito de performance dos estudantes das EEEPs excedeu em 40% os de escolas regulares.

Palavras-chave: eficácia do ensino profissionalizante, escola pública, dados longitudinais, Ceará.

Abstract

The impact of vocational schools on student's performance is controversial in the world literature. Thus, this paper contributes to such debate by evaluating the performance of Vocational Education State School (VESS) students on secondary education exams in a case study for the state of Ceará (Brazil). This is done through matching techniques applied to a unique longitudinal database. Among the results, it is highlighted that the mean effect of the VESS treatment is positive and significant in all knowledge areas of the exams, especially in Writing, for which the effect of VESSs on student performance exceeded that of students from regular schools by 40%.

Keywords: efficacy of vocational education, public school, longitudinal data, State of Ceará (Brazil).

JEL classification: C81, I21, I28.

DOI: <http://dx.doi.org/10.11606/1980-5330/ea156691>

* Federal University of Ceará. E-mail: ronald@ufc.br

† Federal University of Ceará (Sobral). E-mail: zilania@ufc.br

1 Introduction

Because education is considered one of the main driving forces for economic growth and development, improving its quality has been kept on the public policy agenda as a priority action, especially in countries whose performance relative to others as measured by international institutions is considered low, such as Brazil. The economic magnitude of a country follows its educational greatness, although regions such as Latin America exhibit lower educational levels compared to others than their income levels predict (Ganimian & Solano 2011, Hanushek & Woessmann 2012, Levy & Schady 2013). Hence, it is recognized that differences between countries with high and low incomes essentially depend on how fast they grow in the long term. Therefore, in a comparative analysis of countries, emphasis should be given to the need to measure educational levels in qualitative terms through scores on international exams (Barro 2013).

According to the guidelines of the Organização para a Cooperação e Desenvolvimento Econômico - OECD (2016) about low educational performance, policy makers should prioritize education and translate this priority into additional resources while recognizing that addressing this problem requires a multidimensional approach adapted to national and local circumstances, i.e., organizing schools and education systems such that they can offer opportunities for childhood education to all and identify low-performing students and schools and intervene with appropriate policies. Therefore, reducing the number of underachieving students is not only a goal in itself but also an effective means to improve the overall performance of an education system; thus, understanding the factors that contribute to this problem is only the first step because it must be followed immediately by implementation of policies to improve education.

Brazil has made progress in basic education according to the National Report of the Programme for International Student Assessment (PISA 2013). Although this has occurred with greater absolute progress in terms of proficiency in Mathematics when comparing exams between 2003 and 2012, especially at the secondary education level, education in Brazil is still far from the level desired by society, which points to education as the most stable foundation of economic competitiveness and for overcoming social and regional inequalities. To understand the reasons for this apparent puzzle, it is necessary to investigate how to measure the effect of public schools and what is the impact of different types of school on student performance. This would be very useful to identify practices that can be replicated elsewhere and how schools with different qualities affect student performance.

One of the policies aimed at improving public secondary education is the concept of vocational education. Although vocational education was created to meet the needs of young people living at the margins of society (Tavares 2015), after the passing of the first National Educational Bases and Guidelines Law, it became part of secondary education, with a focus on higher education as one of its purposes. Thus, students from vocational schools would not be disadvantaged because in addition to being prepared for the job market, they could also enter university. In Brazil, Araújo (2014) found that participating in technical and vocational education is positively related with school performance on the ENEM and facilitates entry into the labor market. However, Polidano (2016) observed that in Australia, combining academic and vocational

education reduces the chances of a student entering university.

By integrating the two educational networks - vocational and general - the structural duality between training for work and preparing for university can be broken (Castro 2005, Martins 2012, Moehlecke 2012). Vocational education may be an option for developing countries seeking to improve youngsters' opportunities and earnings in the labor market, by preparing them with specific job skills and trainings. However, the debate about the benefits of this type of school is inconclusive in the international literature (Mahirda & Wahyuni 2016). In some countries, vocational students get higher incomes when compared to general school students. Neuman & Ziderman (1989) found that in Israel vocational school graduates earn 10% more than traditional school ones. In Egypt, this type of education results in a prize of 29.3% for men and 27.2% for women (El-Hamidi 2006). Attanasio et al. (2011) assess the impact of a randomly selected training program on disadvantaged youth in Colombia in 2005 and found that participation in this program increases the chances of acquiring formal employment and increases earnings by 19.6% for women. Newhouse & Suryadarma (2011) used a rich database containing longitudinal information from 1993 to 2007 for youths in Indonesia and found that the estimated salary premium is higher for female vocational graduates than for male.

On the other hand, studies in Tanzania (Kahyarara & Teal 2008) and Suriname (Horowitz & Schenzler 1999) found that students from general schools have higher incomes than those from professional ones, while in Germany (Lechner 2000), South Korea (Chae & Chung 2009), Romania (Malamud & Pop-Eleches 2008) and Indonesia (Bellei 2009, Mahirda & Wahyuni 2016) the results between schools did not differ significantly.

In Brazil, professional education is divided into three levels: basic or professional, technical and technological qualification. The basic courses are aimed at anyone interested in it, regardless of previous education; Technical courses are offered simultaneously with the high school and have their own curriculum organization; Technical qualifications are higher education courses. There are three types of technical courses: Integrated - the student takes the high school and the technical course simultaneously; Concomitant - the technical course complements high school; Subsequent - it is aimed at students who have already completed high school. This article will concentrate its analysis on the technical courses, which belong to those offered by the VESS of Ceará.

Some studies focused on Brazil have investigated the effects of this type of education on labor market insertion. Severnini & Orellano (2010) investigate the graduates of professional courses until 1996 through a national survey denominated Research on Living Standards (PPV). This survey furnishes separate information of courses taken at technical schools from those taken at private schools, in addition to others. The results reveal that the marginal impact of vocational education on the probability of inactivity is negative and statistically significant for those graduating from technical level courses (courses linked to high school). On the other hand, the effect of this type of course on earnings was not significant.

Assunção & Gonzaga (2010), based on the microdata of the special supplement on vocational education of PNAD/2007, observed that the occurrence of individuals who choose vocational education, and whose families have per capita income below two minimum wages, is much lower than in other income

ranges, in addition to providing greater productivity for workers.

Angrist & Pischke (2008) evaluated the effects of mid-level vocational and technological education (VTE) on students' labor market insertion, and found an increase between 1.2 and 1.3 percentage points in the probability of being employed. On the other hand, the probability of working in the activity for which they have graduated is approximately 17 percentage points higher than the students who did not perform VTE.

Amoroso Neto et al. (2017) compared individuals with similar levels of formal education by study modality (technical high school). Using cross-sectional data from the PNAD/2014 supplement, they applied Propensity Score Matching, although without proper control for the students' educational background or corrections for sample selectivity. Nevertheless, they concluded that trainees in technical high school earn on average 15% higher salaries than individuals with similar characteristics who graduated from formal education.

It is worth mentioning that there are few studies in Brazil that assess the impacts of vocational education on public education. One of these few is the study by Angrist & Pischke (2008), who used data from ENEM 2009 to assess whether students who were taught the Professional and Technological Education (PTE) curriculum had increased proficiency in basic subjects. However, that study used cross-sectional data, which does not allow following students during the maturation of vocational education, a gap to be filled with the present case study through the use of longitudinal data.

Vocational Education State Schools (VESSs) in Brazil provide, in addition to the basic disciplines of the traditional secondary education curriculum, technical and vocational courses focused on several fields (e.g., health, agriculture, computing, and finance). This structure allows students to not only complete secondary education and become trained for the labor market but also enter universities. The present study therefore focuses on these educational policies, with an emphasis on vocational training. It should be noted, however, that the choice of the state of Ceará as the object of study stems from its national representativeness¹ and pertinence in generating samples for longitudinal empirical testing. In addition, this choice was based mainly on access to a rich longitudinal database with information about the personal and educational characteristics of students, their family background, and school infrastructure, which would allow following the students throughout secondary education and verifying their performance on the ENEM by crossing the following databases: Permanent Assessment System of Basic Education in Ceará (SPAECE), School Census and ENEM.

The VESS strategy is to integrate high school education to the professional training at a technical level, offering full-time education to youth students. In this way, the state professional education program is similar to the national Federal Institute of Education, Science and Technology (IFCE), both displaying on the diploma "Technical Course in High School", as well as the same guidelines of studying a common core along with the technical base. However, in the state vocational schools the internship is mandatory and a scholarship is paid by the state government. Also, schedules are more flexible in

¹This state came to prominence at the national level by combining full-time academic education with vocational education and increasing the participation of young people in the ENEM.

the IFCE, so that students have more time at their disposal, while in the VESS, education is full time, which requires complete dedication from the students.

According to the Institute of Research and Economic Strategies of Ceará (Silva 2013), Ceará presented the fourth largest participation of state public school students in ENEM in 2011. The participation rate of these students rose from 2.78% in 2009 to 8.89% in 2011, making it the third largest variation (219.78%) among the Brazilian states. In addition, in 2011, the average grade of Ceará vocational schools in ENEM (480.8) was higher than the national average (476.6), while the average of non-vocational public schools in the state was 442.5. In this sense, if only the average of these schools to represent Ceará were considered, the state would occupy the eighth position in the general classification (See Table A.3 in the Appendix Appendix A). It is also noteworthy that, although the analysis conducted here deals with only one state of the Federation, it can be seen from figures A.5 A.1 to A.9 A5 in Appendix Appendix A that the performance of Ceará public school students in the overall average in the ENEM 2014 has similar distribution to students from other states of different regions. Thus, there are indications that the results found here may be representative to the other states.

It is hoped that the approach adopted in the present study will amplify the debate about vocational education that exists in the literature on education economics in Brazil, with emphasis on the effects generated on the performance of students on the ENEM. The aim is to provide new evidence about vocational education in public state schools.

One of the contributions of this article is the database used to measure the effect of vocational education on the ENEM exam through the construction of identifiers. After that, it was possible to generate a longitudinal sample by crossing different databases—SPAECE 2011 and 2012, School Census 2011 and ENEM 2014—a procedure that has not been performed in other Brazilian studies². Comparing the databases made it possible to observe students from the 9th grade through the end of secondary education, thus verifying their trajectory in the schools and allowing the creation of treatment groups of students who entered secondary education in vocational schools in 2012 and were attending the third and last year of secondary education (referred to as 3rd grade in Brazil) in 2014 in these school, and control groups, similarly defined but with students from regular schools. With data about the students in 2011, a period prior to secondary education, this study is able to observe the factors that possibly determined their entry into vocational schools in 2012, such as personal characteristics, educational background, parents' education and school infrastructure characteristics. Additionally, to isolate the student effect³, the coarsened exact matching (CEM) method is applied to the Mathematics and Writing scores in SPAECE 2011. Another contribution of this study to the literature about Brazil is the choice of the analysis method, which is based on a model recently developed by Belloni & Fernandez-Val (2015) and has thus far not been applied to Brazilian data. The method applies Lasso and post-Lasso selection to a set of explanatory variables, selecting those considered important in determining the treatment. After these procedures, the average treatment effect on the treated (ATT) is found.

²The difficulties and effort spent generating the final sample made it unfeasible to expand it to include other Brazilian states.

³Knowledge and skills acquired by students in the period prior to entry into vocational schools.

This article is organized into the following sections: theoretical- methodological aspects of the literature about vocational schools, methodological approach and results, followed by the concluding remarks.

2 Vocational School

There is consensus in the literature regarding the crucial importance of education for the development of a country. With this in mind, many public policies are implemented to increase educational quality. The advantages and disadvantages of the type of secondary education have been the focus of many debates in the literature. There are countries that group students based on academic skills or the curriculum and subsequently determine whether the student will pursue an academic path that is geared toward entry into universities or a professional path for entry into the labor market. According to Leme et al. (2016), the secondary education system differs among countries, with countries such as Portugal, Spain and Sweden requiring students to specialize in a certain area, whereas others, such as Finland, Denmark and the United States, follow a more general curriculum. To estimate the effect of this heterogeneity on student performance on the PISA and PIAAC (Program for the International Assessment of Adult Competencies) tests, the authors used a difference-in-differences model for 20 countries, and the results suggested that the type of secondary education (specialized versus general) does not explain the difference in the students average performance among the countries.

The vocational education is also a form of cognitive formal knowledge improvement because it prepares students to enter the labor market while attending high school. It may be an option for developing countries that seek to improve the opportunities and gains of young people in the labor market by preparing them with job-specific techniques and skills. In addition to that, it is also considered a means of promoting students' personal skills, experiences with other individuals, entry into the labor market and better performance in the regular disciplines (Angrist & Pischke 2008). According to Martins (2012), the structural duality between job training and preparation for university can be broken by integrating the vocational and general education networks.

2.1 The Case of Ceará

The Government of the State of Ceará initiated in 2008 a project to create State Vocational Education Schools, providing, in addition to the basic education of the high school curriculum, technical and vocational courses aimed at various areas and disciplines that seek personal training. This action allows students to complete high school and professionalize in areas targeted to the job market.

This type of education is aimed not only at the insertion of youngsters in the job market, but also in the university. According to the implementation of the first Law of Guidelines and Bases of National Education, one of the focuses of the vocational school deals with the student's entrance in the university (Castro 2005, Martins 2012, Moehlecke 2012).

A Management Report of the Government of Ceará of 2014 states that there was an expansion of approximately 10% of access to public schools by students from EEEP who were approved at statewide universities. This result

motivates further investigation, which is proposed here by assessing this type of educational on ENEM performance.

3 Methodology

In the analysis of vocational schools vis a vis student performance, it should not be neglected that there is a source of data bias associated with the variable that identifies VESS students, since the demand to attend high school is constrained by the limited supply of such schools. As a result, the Education Secretary of Ceará (SEDUC) establishes a criterion for the selective process of entry into each school, which is based on the highest grade averages obtained by applicants in elementary school until filling vacancies. The tie-break criteria, if necessary, are the age and proximity of the residence to the school. Thus, there is no pre-established value of the minimum average grade to guarantee that a student will be registered.

It is worth mentioning that there is a possibility of arising biases from unobservable and observable characteristics associated with the selection process. First, it could emerge from not taking into account the characteristic related to the student's motivation. In this case, one way of figuring it out would be to consider in the analysis all the candidates in each school, distinguished between those who were and were not classified. However, this database is not available. The observable characteristic is related to the unavailability of information on the performance of students enrolled in elementary school; otherwise, it would be possible to verify the grade limit of entry in each school and, depending on the sample size, it could be applied a discontinuous regression with several points of discontinuity.

Considering the unavailability of these informations, it seems that the second best procedure is to minimize eventual remaining bias through matching the students by their grades before entering high school, whose adequacy is justified by the supposition that there is a direct correlation with the middle school grade average. Besides, since candidates have prior knowledge that the selection process is based on performance, it is plausible to assume that this variable is one of the factors that affect the students' motivation to apply for a place in the school. Thus, the aim is not to try to solve these problems, but to minimize them by providing more homogeneity in the sample. The procedure chosen for this sample homogenization is based on the CEM algorithm to the following variables: Mathematical and Portuguese proficiency of the students of the 9th grade of elementary school in 2011. After that, the effect of vocational school will be estimated through the application of propensity score matching via selection of the Lasso method, followed by the Rosenbaum (2002) in order to verify the existence of omitted variables in the model that can influence the student's choice and performance.

3.1 CEM Algorithm

Given that there is a selection process for students to be admitted to VESSs, there may be arguments that relate the good performance of these schools to the skills of the selected students. Thus, to isolate the effects of students who already performed well before entering the vocational schools, the CEM algorithm is applied, and matching is performed ex-ante based on the students' 9th grade proficiency scores in Mathematics and Writing on the SPAECE 2011,

such that the resulting sample has no differences/imbalance in terms of previous performance between the treatment and control groups.

According to Blackwell et al. (2009), CEM is a matching method that reduces the imbalances in the empirical distribution between treatment and control, allowing the balance to be pre-selected by the researcher. This algorithm does not require any assumptions about the data generation process, except ignorability, and ensures that the imbalances between the groups after matching are no greater than a pre-selected threshold. The authors showed that CEM allows greater robustness of analysis than other methods such as PSM, which is the reason why it is used in this study⁴.

Iacus et al. (2008) developed a global imbalance measure, given by the statistic, that is based on the difference between multidimensional histograms of all pre-treatment covariates in the treatment and control groups and can be represented by

$$\mathcal{L}_1(f, g) = \frac{1}{2} \sum_{l_1, \dots, l_k} |f_{l_1, \dots, l_k} - g_{l_1, \dots, l_k}| \quad (1)$$

Perfect global balance is indicated by $\mathcal{L}_1 = 0$ and large values of \mathcal{L}_1 indicate major imbalances between the groups, with a maximum value of $\mathcal{L}_1 = 1$, which indicates complete separation. Denoting the relative frequencies of the matched data by f^m and g^m , a balanced matching will produce a reduction in the \mathcal{L}_1 statistic, i.e., $\mathcal{L}_1(f^m, g^m) \leq \mathcal{L}_1(f, g)$.

3.2 Propensity Score Matching (PSM) combined with Lasso and post-Lasso methods

For these propensity score-based estimators to be considered without bias, it is necessary that the treatment conditions are exogenous. In addition, another important point to be analyzed regards the inclusion of variables in the model as controls. Some authors suggest that they are chosen both for relevance in designation and participation in treatment, which consider the economic, social and political factors of the program, in addition to the outcome variables, whose impacts will be measured. Thus, the PSM method depends on the degree to which the observed characteristics explain the participation in the program.

Generally, economic intuition is used to choose the covariates; however, this may not be accurate in determining the exact manner in which control variables should be used, and the functional form could be subject to misconceptions, also. Therefore, if the choice of covariates does not include determinant variables of the treatment, the presence of these non-observable characteristics will make the chosen method unable to reproduce the results of a random experiment. In the case of incorrect choice of these variables and the functional form, the model can generate biased estimates if the treatment variable does not behave exogenously when conditioned on the set of controls.

The Lasso method (Tibsharani 1996) consists of estimating parameters by imposing constraints on the absolute values of the estimated coefficients. Belloni & Fernandez-Val (2015) developed methods to evaluate the inference of models whose estimation is performed via selection methods such as Lasso.

⁴Some studies have reported having had success in recent applications of the CEM: Datta (2015), Aroca et al. (2014), and Schurer et al. (2016).

These results apply to ATT models, with or without endogenous variables for the treatment variable.

Assuming that the propensity score is calculated using the logit link function, then,

$$E(Y_i | X) = \frac{\exp(g(X))}{1 + \exp(g(X))} + \epsilon_i \quad (2)$$

Traditional methods to estimate (2) use economic intuition to determine the variables and the functional form of $g(X)$. Let $\beta = (\beta_1, \dots, \beta_p)$ be the parameter vector of X . It is assumed that X is formed by original control variables and their transformations⁵ and that $p \ll \dim(X)$. The Lasso method selects s ($s \ll p$) parameters after application of the following minimization problem:

$$\beta_L = \operatorname{argmin} \hat{Q}(\beta) + \frac{\lambda}{n} \|\beta\|_1 \quad (3)$$

where $\hat{Q}(\beta) = E[(g_i - X\beta)^2]$, $g_i = \{0, 1\}$ being 1 if individual i participated in the program and 0 otherwise, $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$. That is, the method minimizes the sum of the squares of the residuals conditioned on the parameter constraints.

The term λ (tuning point) corresponds to the weight that will be given to the restriction; the greater its value, the greater the effect of the constraint on the estimated coefficients. Equation (3) can be rewritten as follows:

$$\beta_L = \operatorname{argmin} \sum_{i=1}^n (g_i - X\beta)^2 \quad (4)$$

$$\text{subject to } \sum_{j=1}^p |\beta_j| \leq t \quad (5)$$

This specification makes clear the manner in which the constrained optimization is applied by the Lasso method. The estimation of the parameters will occur provided that the sum does not exceed the turning point. The choice of λ is a crucial element in selecting the variables. Belloni et al. (2010) proposed a feasible λ that is applicable in the presence of heteroscedasticity; their procedure is employed in this work to estimate equation (2).

It was decided to apply the Lasso variable selection method to the propensity score. This enables obtaining a score through the estimation of $g(X)$ composed of robustly selected variables. Although the method does not require selecting the variables based on economic theory, in certain situations, there may be interest in some variables not being subject to constraints because some of them may have sufficient theoretical relevance such that keeping them in the model is required. Belloni et al. (2010), for example, estimate yield equations for the United States using the Lasso method to select relevant instrumental variables but do not subject variables such as sex, experience, and squared experience to constraints.

The Lasso method is a particular case of a set of variable selection methods known as shrinkage methods and has two advantages over other variable

⁵These transformations may include non-transformed variables and variables subject to some transformation, such as interactions between variables, polynomial variables, and B-splines, among others.

selection techniques. First, it selects variables when forcing certain irrelevant parameters to obtain a value of zero, in the sense of not reducing the mean square error by including the variable. Second, the functional form of equation (3) is convex. Therefore, the optimization problem is computationally simpler, and the optimal values obtained will be global. These two properties make the Lasso method appropriate to select control variables relevant to explain the behavior of the binary variable or the treatment variable, as in Belloni et al. (2014).

Therefore, this paper adopts the Lasso estimation technique, in which an *ex ante* selection is performed. It is important to note that not necessarily the selected variables need to have economic significance. In fact, these will only have strong explanatory power over the binary or treatment variable.

To compose the set of variables to be selected by the Lasso method—i.e., starts and stays in VESSs during secondary education from 2012 to 2014, variables representing personal and educational characteristics of students, parents' education and school infrastructure in 2011⁶, the period prior to the entry of these students into vocational schools—which will determine the probability that the student belongs to the treatment group, were included. These variables were transformed because some transformations may also have explanatory power on the treated.

3.3 Sensitivity Analysis

If there are unobserved variables that affect the selection process, they will also affect the results; therefore, estimators based on the propensity score are not consistent with the treatment effect. Thus, omission of these variables can generate biased results for the estimated mean ATT. The method known as Rosenbaum bounds (Rosenbaum 2002, DiPrete & Gangl 2004) allows determining how “strong” should be the influence of an omitted variable on the selection to participate. The probability of participation of an individual i is given by

$$\pi_i = Pr(D_i = 1 | x_i) = F(\beta x_i + \gamma u_i) \quad (6)$$

As already explained, D_i is equal to 1 if the individual receives the treatment and 0 otherwise; x_i are the observed characteristics of individual i ; u_i corresponds to the unobserved variable and represents the effect of u_i on the decision to participate in the program. If there is no selection bias, then will be zero, and the probability of participation will be exclusively determined by the observable characteristics. However, in the presence of selection bias, two individuals with the same observed covariates x will have different chances of receiving the treatment, $\frac{\theta_i}{1-\theta_i}$ and $\frac{\theta_j}{1-\theta_j}$ the odds ratio is given by

$$\frac{\frac{\theta_i}{1-\theta_i}}{\frac{\theta_j}{1-\theta_j}} = \frac{\theta_i(1-\theta_j)}{\theta_j(1-\theta_i)} = \exp[\gamma(u_i - u_j)] \quad (7)$$

If individuals have the same observable characteristics, then vector x is cancelled out. Thus, if there are no differences in unobserved variables ($u_i = u_j$) and if these variables do not influence the probability of participation ($\gamma =$

⁶All variables may be found in Table A.1 and Table A.1 in Appendix Appendix A.

0), the odds ratio is equal to 1, implying the non-existence of selection bias. It follows then that their odds of participation differ, i.e., if the odds ratio is other than 1, it can only be due to the presence of non-observable variables. The sensitivity analysis evaluates how much of the program's effect is altered by changes in the values of γ and $u_i - u_j$. This means examining the limits of the odds ratio of participation. Rosenbaum (2002) shows that (7) implies the following limits for the odds ratio:

$$\frac{1}{e^\gamma} \leq \frac{\theta_i(1 - \theta_j)}{\theta_j(1 - \theta_i)} \leq e^\gamma \quad (8)$$

3.4 Variables and Databases

To assess the impact of the vocational education in secondary education policy of the VESSs of Ceará on the performance of students on the ENEM, the students' scores in five areas and the overall mean score on the ENEM were used as indicators of impact, namely, Natural Sciences (NS), Humanities (HM), Language and Codes (LC), Mathematics (MT) Writing (WT), and Overall Mean (OM).

Data about the performance in the five areas of knowledge of the ENEM 2014⁷ and school infrastructure in 2011 were taken from the INEP, ENEM Microdata and the School Census. The personal and educational characteristics and parents' education of students in public state schools in Ceará and the identification of vocational schools were obtained from the Department of Education of Ceará (Secretaria de Educação do Ceará, SEDUC) through SPAECE 2011. Thus, to construct the sample to be used for this study, it was necessary to join three different databases, SPAECE, ENEM and School Census, which was only possible through the provision of identifiers created in the SEDUC to follow students from the 9th grade until the 3rd and last grade of secondary education.

To fulfill the objective established for this study, two groups were formed: one belonging to the program (treatment group) and another that did not undergo the program (control group). The treatment group is composed of students who entered secondary education in vocational schools in 2012 and were attending the 3rd and last grade of this level in 2014 in these schools; the same definition is used for the control group, but referring to regular schools. To form the treatment group, it is necessary to follow these students throughout secondary education, i.e., identify whether they stayed in vocational schools and whether they repeated a grade during the period from 2012 to 2014. The databases SPAECE/2012 and ENEM 2014 were used for this identification.

It is worth noting that it was not possible to observe these students in 2013 because SPAECE/2013 is based on just a sample. However, it is believed that this does not compromise the results because following the students in 2012, 2013 and 2014 only serves to identify whether they stayed in vocational schools during secondary education. Given that it was possible to observe the students who attended the 1st year of this level in a VESS in 2012 and the 3rd year in 2014 at the same school, it is expected that these students also

⁷Last year with available data about individual performance on ENEM.

attended the 2nd year in this school, as it is not probable that a student would leave the vocational school for one year to return the following year.

First, to be included the sample, students must have taken the SPAECE exam and have completed the 9th grade in 2011 such that the variables that will determine the entry into vocational schools in 2012 can be extracted. Students must also have taken the ENEM in 2014 to obtain the outcome variables. A total of 40,435 students were included in the sample at this point. Of these, 35,680 students could also be observed in SPAECE 2012, where 35,312 started secondary education that year⁸, and 5,491 students enrolled in vocational schools. In 2014, 5,092 students remained in VESSs, 24 repeated the 2nd year, and 5,068 were attending the 3rd year; of these, 4,977 were present on the day of the ENEM 2014, but only 4,635 had data regarding personal characteristics (treatment group). Regarding regular schools, 29,658 students stayed in these schools in 2014, of which 6 students were in the 1st year, 1,937 were in the 2nd year and 27,715 were in the 3rd year. Of these, 22,201 took the exam that year and only 19,811 answered the questionnaire about personal characteristics (control group)⁹. For the model estimations, different covariates obtained from the originals ones were considered, such as described in Table A.1 and Table A.1 in Appendix Appendix A.

4 Results

4.1 Descriptive Analysis

Descriptive statistics regarding the dependent variables and some control variables¹⁰ related to students from VESSs and regular schools in Ceará, extracted from the SPAECE 2011, School Census 2011 and ENEM 2014 databases, are presented in Table 1.

The mean scores on the ENEM 2014 of students from VESSs are higher than those of students from regular schools, and the coefficient of variation for VESSs is less than or equal to those for regular schools for all exams. In Natural Sciences, the mean difference in scores is 40.12 points in favor of vocational schools, and the relative variation is the same. There are differences in terms of the scores in Humanities and Language and Codes of 52.32 and 49.39 points, respectively, with dispersion relative to the mean 16% and 36% higher, respectively, for the students from VESSs.

The differences in the Mathematics and Writing scores were higher and less homogeneous, as the students from VESSs scored 57.51 and 143.77 points higher, respectively (relative variation of 4.7% and 56.6%), in these subjects than the students from regular schools. The Overall Mean showed that the students from VESSs outperformed the students from regular schools by approximately 68 points (15.4%). However, the variation relative to the mean was approximately 21% higher for the students from regular schools.

Based on this descriptive analysis of the dependent variables, it is expected that the effect of vocational education on the performance of students on the ENEM is significant, but this cannot be verified only by comparing the two groups. This is a “naïve” technique that is commonly used by non-experts

⁸368 students repeated the 9th grade in 2012.

⁹For details, see the Appendix Appendix A.

¹⁰Total of 95 control variables.

because not having undergone the intervention does not mean that the untreated group is a good representation of what would have happened to the treatment group had it not been treated, i.e., absence of treatment for some does not automatically generate the counterfactual of no treatment for others. Thus, appropriate methods that seek to isolate the effect of the programs from the effects of other factors affecting the variables of interest are needed.

The variables related to the characteristics of the students observed in 2011, the period prior to the entry of the students into secondary education, are presented in Table 1. The students from vocational schools had approximately 10% higher scores in Mathematics and Writing on the SPAECE exam than the students from regular schools. These variables were included to represent the educational background and skills of students in primary education and may have been important in the selection of these students by the vocational schools.

Both VESSs and regular schools have similar characteristics regarding the students in 2011: mean age of 15.7 years, more than 50% are female and approximately 20% consider themselves white. However, the students who entered vocational schools have parents with better educational levels, as 21% and 31% of the students have fathers and mothers, respectively, with at least incomplete secondary education, whereas this proportion for students from regular schools is 14% and 20%. It is important to note that the proportion of the students from VESSs who could not provide an answer about the educational levels of their fathers and mothers were 29%, and 17%, respectively, and the proportions were 32% and 22%, respectively, for the students from regular schools.

The variables representing the socioeconomic conditions of students in 2011 presented in Table 1 indicate that the students who entered vocational schools have better economic status than students who went to regular schools; for instance, the proportion of VESS students whose family are enrolled in the Cash Transfer Program is lower and the proportions that own a car, employ a domestic worker and have a washing machine are greater.

The characteristics related to the infrastructure of high schools (vocational and regular) in 2011 were used for this study because it is expected that these influenced the decision of the student in choosing the school after primary education. It is believed that the students observe these characteristics before choosing the type of school they plan to study in. The choice certainly will depend upon school infrastructure. Thus, this justifies adding these variables lagged one period prior to entry school. Based on Table 1, vocational schools have some better characteristics than regular schools, such as 99% have a teachers' lounge, 88% have a library, and they have 59% more computers than regular schools.

4.2 Results of Pre-Matching

Before estimating the effect of this type of school on students' performance in ENEM, we chose to balance through CEM algorithm, using as covariates the Mathematics and Portuguese proficiency in SPAECE, based on the assumption that there is a direct relation of these variables with the schools' selective process. This will allow comparing similar individuals in terms of school performance. Thus, after balancing the groups, students who entered vocational and regular schools in 2011 will not have significant differences in terms of

Table 1: Descriptive statistics regarding the sample variables - 2014

	Variables	V ESSs			Regular Schools		
		Mean	Standart deviation	Variation coefficient	Mean	Standart deviation	Variation coefficient
Var. Results	Score_NS	489.82	65.29	0.13	449.70	58.92	0.13
	Score_HM	550.34	66.15	0.12	498.02	69.57	0.14
	Score_MT	479.07	102.10	0.21	421.56	90.85	0.22
	Score_LC	512.98	71.55	0.14	463.59	88.06	0.19
	Score_WT	533.46	158.54	0.30	389.69	184.61	0.47
	Score_OM	512.98	71.55	0.14	444.51	74.90	0.17
Algorithm CEM	Spaace_MT_2011	268.67	42.35	0.16	244.49	41.70	0.17
	Spaace_PT_2011	263.37	40.76	0.15	237.74	42.92	0.18
Covariates*							
Student characteristics	Age	15.7	7.82	0.50	15.7	7.27	0.46
	Sex	0.42	0.49	1.16	0.43	0.49	1.14
	Race	0.21	0.41	1.95	0.19	0.39	2.05
Educational level Father	E_father1	0.07	0.25	3.57	0.11	0.31	2.82
	E_father2	0.43	0.49	1.14	0.43	0.49	1.14
	E_father3	0.17	0.38	2.24	0.11	0.31	2.82
	E_father4	0.04	0.21	5.25	0.03	0.18	6.00
Educational level Mother	E_mother1	0.03	0.17	5.67	0.05	0.23	4.60
	E_mother2	0.49	0.50	1.02	0.53	0.49	0.92
	E_mother3	0.23	0.42	1.83	0.14	0.35	2.50
	E_mother4	0.08	0.28	3.50	0.06	0.24	4.00
Socio-economic condition	Cash transfer	0.72	0.44	0.61	0.80	0.39	0.49
	Car	0.51	0.49	0.96	0.47	0.49	1.04
	Domestic_worker	0.09	0.28	3.11	0.07	0.27	3.86
	Computer	0.33	0.47	1.42	0.20	0.40	2.00
School characteristics	Washing_machine	0.31	0.46	1.48	0.24	0.42	1.75
	Waste_collection	0.99	0.06	0.06	0.99	0.08	0.08
	Principal's_office	0.98	0.12	0.12	0.98	0.12	0.12
	Teachers'_lounge	0.99	0.04	0.04	0.96	0.18	0.19
	Computer_lab	0.99	0.12	0.12	0.99	0.06	0.06
	Library	0.88	0.32	0.36	0.83	0.36	0.43
	No. Computers	68.31	22.13	0.32	42.90	19.35	0.45

Source: Own calculations based on SPAECE 2011 Microdata, School Census 2011 and ENEM 2014.

Note: Education level – (1) did not study, (2) primary education, (3) secondary education, (4) higher education. (*) Since there were 95 covariates to estimate the Lasso method, we chose to expose only a few ones related to student characteristics, family background and characteristics of high schools in 2011.

observed scholar skills. Therefore, the binary variable of treatment was considered for the students who entered the vocational schools (treaties) in 2011 and also graduated in these schools; the control refers to the students of other schools.

The student effect can harm the estimation results when finding the effect of the vocational school policy on student academic performance on the ENEM is desired. Thus, as a strategy to isolate the effects of students considered as good before entering the vocational schools, the coarsened exact matching (CEM) algorithm was applied to the variables proficiency in Mathematics and in Writing on the SPAECE 2011 exam of 9th graders.

Table 2 presents the measures of imbalances before and after matching. It is observed that the multivariate of each variable was reduced to values closer to zero. Moreover, the difference between the means of the treated and control groups was reduced in the two proficiencies: 24.17 to 0.05 in Mathematics and 25.62 to 0.0833 in Writing. This algorithm allowed identifying students in the two groups (control and treatment) who can be compared. It was possible to observe that 19,660 and 4,634 students belonging to regular and vocational schools, respectively, can be matched based on the established variables to isolate the student effect. Thus, the application of PSM will be restricted only to students contained in the matching and identified by the CEM algorithm.

Table 2: Results of the CEM algorithm

Variables	Pre-Matching		Post-Matching	
	\mathcal{L}	Difference of means	\mathcal{L}	Difference of means
MT_SPAECE_2011	0.2424	24.17	0.0522	0.2836
PT_SPAECE_2011	0.2526	25.62	0.0479	0.0833
Multivariate \mathcal{L}	0.3458		0.2253	
No. Treated	4.635		4.634	
No. Control	19.811		19.660	

Source: Own calculations.

4.3 Lasso method and PSM

After restricting the sample by considering only students who were selected by the CEM algorithm, in order to minimize the selection bias by the school, the estimation of the results will consist of the following principle: Application of the Lasso method, which will select from 449 variables (95 level variables exposed in Table A.2 in the Appendix Appendix A and 354 created from their power and interaction), those that are important in determining the choice of the type of school to be attended in high school by the student.

It is noteworthy that this method selects the variables not by economic intuition, but by the process of minimizing the distance between the two treated groups and control, in order to find the variables that will have only strong explanatory power about the treatment variable. Therefore, the variables selected by the Lasso method are not interpreted. They are used in PSM to find the average treatment effect on those treated with the lowest bias between the two groups.

Once the variables have been selected, the traditional propensity score matching procedures are applied: Logit is estimated, the common support region is identified, the weights are chosen for balancing using a specific method (radius, nearest neighbor or kernel), and finally, the average effect of treatment on treaties is found.

The selection of variables input in the PSM model was guided by the Lasso method. Based on the 95 control variables contained in the sample, 449 variables were created from their interactions, and the selection process was applied. Of these, the method proposed by Belloni & Fernandez-Val (2015) selected 39 variables considered as important in determining the treatment, of which 24 are original values (if the school has equipment for photo, satellite, VCR, slide projector, stereo, piped water, nursery, garbage collection, reading room, shower in the bathroom, cafeteria, utility room, generator power, toilet outside the building, special care room, buried trash, public sewer, if student has internet at home, washing machine at home, if student's mother has not studied, if student's mother has complete or incomplete elementary school, if the student's parent has not studied and if the student's parent has completed or incomplete high school), 4, are in the power form (number of computers squared, number of computers cubed, number of employees cubed, number of rooms cubed and 11 are the number of results (if the school has an auditorium and internet, green area and photo equipment, special care room and green area, buried trash and green area, computer number if math proficiency, number of computers and Portuguese proficiency, number of computers and race, number of computers and if the mother did not study, proficiency in Portuguese and number of existing rooms, reading and parabolic room, sound equipment and green area).

Following the procedures of Becker & Ichino (2002), first a logit binary model of a student belonging to a vocational school was estimated using, as observable characteristics, the variables selected by the Lasso method to describe the treatment and control groups. Subsequently, students with propensity scores outside of the support were discarded from the sample.

The results presented in Table A.2 in the Appendix Appendix A indicate that 30 of the 39 variables were significant in determining the treatment when only students who were matchable by the CEM algorithm were considered. It is noted that the common support region is the interval whose propensity score ranges from 0.001769 to 0.9999984. This region ensures that the observations in the treatment group are comparable to those of the control group in terms of the variables selected by the Lasso method.

Thus, the matching to obtain the ATT will occur within this range. However, other methods in addition to PSM are necessary to estimate the ATT, such as the nearest-neighbor, radius and kernel matching methods. An important procedure for constructing the propensity score and implementing matching is to verify the balancing conditions. Figures A.2 through A.4 in the Appendix Appendix A show the results of testing for density overlap between the pre- and post-matching groups using these methods. Tight and similar overlapping is observed between the kernel and radius methods. Thus, we chose matching using the kernel (0.01) method, which was used to find the ATT, whose results are presented in Table 3.

The differences in scores between the two types of school estimated by matching with the first kernel are displayed in Table 3. The table indicates that these differences are significant in favor of VESSs in all areas of knowl-

Table 3: Average treatment effect in the areas of knowledge of ENEM using the kernel (0.01) method

Variables	VESS	Regular	ATT	t-statistic
Score_NS	490.92	445.11	45.11 (10.1%)	13.57
Score_HM	553.05	498.19	54.86 (11%)	14.30
Score_MT	484.05	417.55	66.50 (15.92%)	13.03
Score_LC	517.04	458.85	58.19 (12.68%)	12.23
Score_WT	540.86	386.54	154.32 (39.9%)	15.36
Score_OM	517.18	441.39	75.79 (17.17%)	18.41

Source: Own calculations.

edge. On average, a student belonging to a vocational school performed approximately 12.7% and 11% better in Languages and Codes (LC) and Humanities (HM), respectively, than a student from a regular school in the control group. In Mathematics (MT), the difference is 66.5 points, 15.92% in favor of vocational education students. The lowest and highest effects were observed in the areas of Natural Sciences (NS) and Writing (WT), with VESS students scoring 45.11 points (10.13%) and 154.32 points (39.9%) more than students from regular schools, respectively. For the Overall Mean (OM), the impact of VESS was 75.79 points (17.17% better than regular education). In short, the results indicate that vocational education is more effective at enhancing cognitive abilities than regular education. The results presented in Table 4 show that the ATT is robust to the possible presence of selection bias in all areas of knowledge and in the overall mean.

Concluding Remarks

We believe that this article provides a reasonable contribution to the current debate about vocational education in the literature regarding education economics in Brazil, with emphasis on the effects on school performance. New evidence about the role of public vocational schools is provided by comparing students from these schools with students from regular education schools, isolating the student effect and testing the existence of significant differences between them.

One of the contributions of this study is the database used to measure the effect of vocational education because through the construction of identifiers in partnership with SEDUC, it was possible to assemble a longitudinal sample by combining different databases, namely, SPAECE 2011 and 2012, School Census 2011 and ENEM 2014, which, to the best of our knowledge, has not been performed in any study thus far. Thus, it was possible to observe the students from the 9th grade until the end of secondary education, thereby verifying their trajectory in vocational schools.

Data about performance on ENEM 2014 in five areas of knowledge and school infrastructure in 2011 were taken from the INEP via ENEM Microdata and the School Census, respectively. Data about the personal and educational characteristics and parents' education of students from public state schools in Ceará and the identification of vocational schools were obtained from SEDUC via SPAECE 2011 and 2012.

Table 4: Sensitivity analysis (Rosenbaum Bounds) for the areas of knowledge and the overall mean on the ENEM

Γ	HM	NS	MT	LC	WT	OM
	P^+	P^+	P^+	P^+	P^+	P^+
1	0.000	0.000	0.000	0.000	0.000	0.000
1.05	0.000	0.000	0.000	0.000	0.000	0.000
1.1	0.000	0.000	0.000	0.000	0.000	0.000
1.15	0.000	0.000	0.000	0.000	0.000	0.000
1.2	0.000	0.000	0.000	0.000	0.000	0.000
1.25	0.000	0.000	0.000	0.000	0.000	0.000
1.3	0.000	0.000	0.000	0.000	0.000	0.000
1.35	0.000	0.000	0.000	0.000	0.000	0.000
1.4	0.000	0.000	0.000	0.000	0.000	0.000
1.45	0.000	0.000	0.000	0.000	0.000	0.000
1.5	0.000	0.000	0.000	0.000	0.000	0.000
1.55	0.000	0.000	0.000	0.000	0.000	0.000
1.6	0.000	0.000	0.000	0.000	0.000	0.000
1.75	0.000	0.000	0.000	0.000	0.000	0.000
1.8	0.000	0.000	0.000	0.000	0.000	0.000
1.85	0.000	0.000	0.000	0.000	0.000	0.000
1.9	0.000	0.000	0.000	0.000	0.000	0.000
1.95	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000

Source: Own calculations.

To meet the objective set for this study, a group receiving vocational education and another group not receiving it were formed. The treatment group was composed of students who entered secondary education in vocational schools in 2012 and were attending the last year (3rd grade) in 2014 in these schools; the same definition was used for the control group except that it contained students from regular schools. Based on data about these students in 2011, it was possible to observe the factors that may have led to their entry into vocational schools in 2012.

As a strategy to isolate the effects of students considered good before entering the vocational schools, this study applied the CEM algorithm to the variables proficiency in Mathematics and Writing on SPAECE 2011 for 9th graders because this effect can harm the estimation results when the objective is to find the effect of the vocational school policy on performance on the ENEM. This method established that 19,660 and 4,634 students belonging to regular and vocational schools, respectively, could be matched based on the established variables to isolate the student effect.

Another contribution of this study to the literature about Brazil is the methodological approach chosen to conduct the assessment, as it was based on a model recently developed by Belloni & Fernandez-Val (2015), which applies Lasso and post-Lasso selection in the estimation of the propensity score. This method selected 39 of the 449 variables created from the 95 original variables.

To measure the ATT, matching was performed using the nearest-neighbor, radius and kernel methods. Although these methods yielded similar results,

the kernel (0.01) method was chosen because it was the most balanced for determining the effect. It was found that the students from vocational schools exhibited significant differences in all areas of knowledge; among those who attended VESSs and those who attended only regular school, the most significant effect was on Writing, as the students from vocational schools performed 39.9% better than their counterparts. On average, the students from vocational schools performed 12.7% and 11% better in Languages and Codes and Humanities, respectively, than the students from regular schools. In mathematics, the difference was 66.5 points, 15.9% in favor of vocational students. A lower effect was observed for Natural Sciences (45.11 points; 10.1%). In terms of Overall Mean, the impact of VESSs was 75.79 points, or 17.17%. Furthermore, through sensitivity testing, it was found that the results are robust to the presence of omitted variables.

Therefore, vocational schools were found to be effective at enhancing cognitive skills compared to regular education. By observing the educational background of the student in 2011 and considering it to obtain the matching, this study reduced the student effect and sought to isolate the school effect on school performance in 2014.

These results corroborate and complement Araújo (2014), since they point to a positive correlation between professional education and school performance. However, this study seeks to fill the gap left by the authors regarding selection bias, in which they admit, but do not deal with it. That is, there exists a positive bias of vocational education that results exclusively from selective processes and which may occur not only in vocational schools, but also in private and federal regular schools.

The performance of VESS students at ENEM is reflected in their entrance to the University. According to the Coordination of the Development of Professional Education Development of the State Department of Education, the percentage of approval of students from the VESS increased from 26.5% in 2012 to 49% in 2016. In addition, the presence of these schools in most municipalities in the state tends to reduce regional educational disparities, since the percentage of approval of these students in public universities was expanded in all groups – city capital, Metropolitan Region, medium and small municipalities – between the years of 2014 and 2015.

Acknowledgments

The authors thank the Ceará State Education Secretariat for providing part of the data applied in the paper.

Bibliography

Amoroso Neto, V., Menezes-Filho, N. & Komatsu, B. K. (2017). Os efeitos da educação profissional e do PRONATEC sobre os salários. *Policy Paper, São Paulo*, n. 25, p. 3–24.

Angrist, J. D. & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. 1. ed. Princeton university press.

- Araújo, A. J. N. (2014). Ensino Profissionalizante, Desempenho Escolar e Inserção Produtiva: uma análise com dados do Enem. *Pesquisa e Planejamento Econômico*, p. 131–160.
- Aroca, P., Brida, J. G., Pereyra, J. S. & Volo, S. (2014). *Tourism statistics: correcting data inadequacy using coarsened exact matching*. Bozen Economics & Management Paper Series. BEMPS22.
- Assunção, J. & Gonzaga, G. (2010). *Educação Profissional no Brasil: Inserção e Retorno*. Série Cenários, n. 3.
- 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*, p. 188–220.
- Barro, R. J. (2013). Education and Growth. *Annals of Economics and Finance*, v. 14, n. 2, p. 301–328.
- Becker, S. O. & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata journal*, v. 2, n. 4, p. 358–377.
- Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review*, v. 28, n. 5, p. 629–640.
- Belloni, A, C. V. & Fernandez-Val, I, H. C. (2015). *Program evaluation with high-dimensional data*. Institute for Fiscal Studies.
- Belloni, A., Chernozhukov, V. & Hansen, C. (2010). *LASSO Methods for Gaussian Instrumental Variables Models*. SSRN Electronic Journal.
- Belloni, A., Chernozhukov, V. & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, v. 28, n. 2, p. 29–50.
- Blackwell, M., Iacus, S., King, G. & Porro, G. (2009). CEM: Coarsened Exact Matching in Stata. *The Stata Journal*, v. 9, n. 4, p. 524–546.
- Castro, M H G, T. S. (2005). Os Desafios da Educação no Brasil. In: *A Reforma do Ensino Médio e a Implantação do Enem no Brasil*. Nova Fronteira, p. 119–154.
- Chae, C. & Chung, J. (2009). *Pre-employment vocational education and training in Korea*. World Bank.
- Datta, N. (2015). Evaluating impacts of watershed development program on agricultural productivity, income, and livelihood in Bhalki watershed of Bardhaman District, West Bengal. *World Development*, v. 66, n. 1, p. 443–456.
- DiPrete, T. A. & Gangl, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological methodology*, v. 34, n. 1, p. 271–310.
- El-Hamidi, F. (2006). General or vocational schooling? Evidence on school choice, returns, and 'sheepskin' effects from Egypt 1998. *Journal of Economic Policy Reform*, v. 9, n. 2, p. 157–176.

- Ganimian, A. J. & Solano, R. A. (2011), 'Measuring up? How did Latin America and the Caribbean perform on the 2009 Programme for International Student Assessment (PISA)?', Working Paper. n.38410.
- Hanushek, E. A. & Woessmann, L. (2012). Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics*, v. 99, n. 2, p. 497–512.
- Horowitz, A. W. & Schenzler, C. (1999). Returns to general, technical and vocational education in developing countries: recent evidence from Suriname. *Education Economics*, v. 7, n. 1, p. 5–20.
- Iacus, S. M., King, G. & Porro, G. (2008). *Matching for Causal Inference Without Balance Checking*. SSRN Electronic Journal, Elsevier BV.
- Kahyarara, G. & Teal, F. (2008). The returns to vocational training and academic education: Evidence from Tanzania. *World Development*, v. 36, n. 11, p. 2223–2242.
- Lechner, M. (2000). An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs In East Germany. *The Journal of Human Resources*, v. 35, p. 347–375, abr.
- Leme, A. C., Escardíbul, J. O., Nunes, L. C., Reis, A. R. & Seabra, C. (2016). *The effect of a specialized versus a general upper-secondary school curriculum on students' performance and inequality. A difference-in-differences cross-country analysis*. Documents De Treball De L'ieb 16.
- Levy, S. & Schady, N. (2013). Latin America's social policy challenge: Education, social insurance, redistribution. *Journal of Economic Perspectives*, v. 27, n. 2, p. 193–218.
- Mahirda, K. & Wahyuni, H. (2016). Returning to general and vocational high-schools in Indonesia. *Review of Economic and Business Studies*, v. 9, n. 2, p. 9–28.
- Malamud, O. & Pop-Eleches, C. (2008). *General education vs. vocational training: evidence from an economy in transition*. NBER, Working Paper.
- Martins, A. P. (2012). Pressupostos de Gramsci na Educação Profissional e Tecnológica de Nível Médio. *Revista de Educação, Ciência e Tecnologia*, v. 1, n. 2.
- Moehlecke, S. (2012). O ensino médio e as novas diretrizes curriculares nacionais: entre recorrências e novas inquietações. *Revista Brasileira de Educação*, v. 17, n. 49, p. 39–58.
- Neuman, S. & Ziderman, A. (1989). Vocational Secondary Schools can be more Cost-effective than Academic Schools: the case of Israel. *Comparative Education*, v. 25, n. 2, p. 151–163.
- Newhouse, D. & Suryadarma, D. (2011). The value of vocational education: High school type and labor market outcomes in Indonesia. *The World Bank Economic Review*, v. 25, n. 2, p. 296–322.

Organização para a Cooperação e Desenvolvimento Econômico - OECD (2016). *Who Are The Low-Performing Students?*. OECD Publishing. Pisa in Focus, 60.

PISA (2013). *Relatório Nacional PISA 2012 – Resultados brasileiros*. Fundação Santillana.

Polidano, C Tabasso, D. (2016). *Fully Integrating Upper-Secondary Vocational and Academic Courses: A Flexible New Way?*. Working Paper, n. 9694.

Rosenbaum, P. (2002). *Observational Studies*. 2. ed. Springer.

Schurer, S., Alspach, M., MacRae, J. & Martin, G. (2016). *The Medical Care Costs of Mood Disorders: A Coarsened Exact Matching Approach*. Iza Discussion Papers.

Severnini, E. R. & Orellano, V. I. F. (2010). *O efeito do ensino profissionalizante sobre a probabilidade de inserção no mercado de trabalho e sobre a renda no período pré-Planfor*. Fundação Getulio Vargas, Escola de Economia de São Paulo.

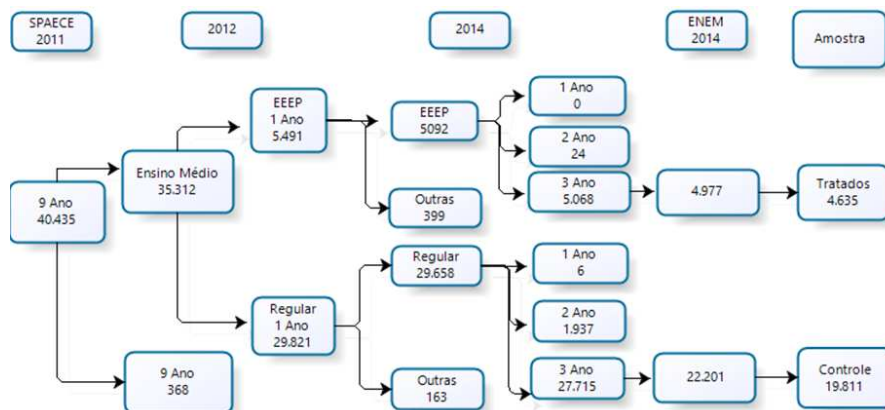
Silva, V. H. O. (2013). *Análise da Participação das Escolas Públicas Estaduais Cearenses no Exame Nacional do Ensino Médio (Enem): 2009-2011*. Ipece Informe, 54.

Tavares, P. A. (2015), PhD thesis. *Três ensaios em economia da educação*. Tese (Doutorado) - FGV EESP - Escola de Economia de São Paulo.

Tibsharani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 58, n. 1, p. 267–288.

Appendix A

Figure A.1: Construction of groups: treatment and control



Source: Own Elaboration based on Microdata of SPAECE 2011, 2012; Escolar Census 2011 e ENEM 2014

Table A.1: Description of Variables

Variables	Description	
Areas of knowledge		
NS	Proficiency in Nature Science	ENEM 2014
HM	Proficiency in Humanities	ENEM 2014
MT	Proficiency in Mathematics	ENEM 2014
LC	Proficiency in Language and Codes	ENEM 2014
WT	Proficiency in Writing	ENEM 2014
OM	Proficiency in Overall Mean	ENEM 2014
Treatment		
EEEE	1, if vocational school; 0, otherwise	SPAECE
Algorithm CEM		
MT_2011	Proficiency in Mathematics in 2011	SPAECE 2011
Nota_WT_2011	Proficiency in Writing in 2011	SPAECE 2011
Covariates		
Age	Years of age	SPAECE 2011
Gender	1, if male; 0, otherwise	SPAECE 2011
Race	1, if white; 0, otherwise	SPAECE 2011
Father's schooling		
E_pai01	1, if father is illiterate; 0, otherwise	SPAECE 2011
E_pai02	1, if father concluded fundamental school; 0, otherwise	SPAECE 2011
E_pai03	1, if father concluded high school; 0, otherwise	SPAECE 2011
E_pai04	1, if father concluded university; 0, otherwise	SPAECE 2011

Source: Own calculations.

Table A.1: Description of Variables (continued)

Variables	Description	
Mother's schooling		
E_mae01	1, if mother is illiterate; 0, otherwise	SPAECE 2011
E_mae02	1, if mother concluded fundamental school	SPAECE 2011
E_mae03	1, if mother concluded high school; 0, otherwise	SPAECE 2011
E_mae04	1, if mother concluded university; 0, otherwise	SPAECE 2011
Home Socioeconomic Characteristics		
Calç_ rua	1, if home locates at nondirty road; 0, otherwise	SPAECE 2011
E_ rua	1, if electric power; 0, otherwise	SPAECE 2011
Água_ rua	1, if piped water; 0, otherwise	SPAECE 2011
Coleta_ rua	1, if garbage collection, 0, otherwise	SPAECE 2011
Bolsa_ familia	1, if cash transfer program; 0, otherwise	SPAECE 2011
Empregada	1, if housekeeper; 0, otherwise	SPAECE 2011
Banheiro_ casa	1, if bathroom; 0, otherwise	SPAECE 2011
Radio_ casa	1, if radio; 0, otherwise	SPAECE 2011
Geladeira_ casa	1, if refrigerator; 0, otherwise	SPAECE 2011
TV_ casa	1, if TV; 0, otherwise	SPAECE 2011
Maq_ lavar_ casa	1, if laundry machine; 0, otherwise	SPAECE 2011
DVD_ casa	1, if DVD; 0, otherwise	SPAECE 2011
Carro/ moto	1, if automobile; 0, otherwise	SPAECE 2011
Computador	1, if computer; 0, otherwise	SPAECE 2011
Internet	1, if internet; 0, otherwise	SPAECE 2011
Jornais	1, if newspaper; 0, otherwise	SPAECE 2011
Revistas	1, if magazines; 0, otherwise	SPAECE 2011
Dicionários	1, if dictionary; 0, otherwise	SPAECE 2011

Source: Own calculations.

Table A.1: Description of School Variables (continued)

Variables	Description	
School Characteristics		
Localização	1, if urban; 0, otherwise	Escolar Census2011
Agua_filtrada	1, if filtered water; 0, otherwise	Escolar Census2011
Agua_rede_publica	1, if public water network; 0, otherwise	Escolar Census2011
AGUA_POCO	1, if water from deep well; 0, otherwise	Escolar Census2011
AGUA_CACIMBA	1, if water from shallow well; 0, otherwise	Escolar Census2011
AGUA_RIO	1, if water from river; 0, otherwise	Escolar Census2011
E_GERADOR	1, if electric power generator; 0, otherwise	Escolar Census2011
E_OUTROS	1, if other type energy supply; 0, otherwise	Escolar Census2011
ESGOTO_PUBLICA	1, if sewage public network; 0, otherwise	Escolar Census2011
ESGOTO_FOSSA	1, if septic tank; 0, otherwise	Escolar Census2011
ESGOTO_INEX	1, if sewage; 0, otherwise	Escolar Census2011
LIXO_COLETA	1, if garbage collection; 0, otherwise	Escolar Census2011
LIXO_QUEIMA	1, if garbage is burned; 0, otherwise	Escolar Census2011
LIXO_JOGA	1, if garbage is discarded; 0, otherwise	Escolar Census2011
LIXO_RECICLA	1, if garbage is recycled; 0, otherwise	Escolar Census2011
LIXO_ENTERRA	1, if garbage is buried; 0, otherwise	Escolar Census2011
LIXO_OUTROS	1, if other discard of garbage; 0, otherwise	Escolar Census2011
S_DIRETORIA	1, if pincipal's office; 0, otherwise	Escolar Census2011
S_PROFESSOR	1, if teacher's office; 0, otherwise	Escolar Census2011
L_INFORMATICA	1, if computer lab; 0, otherwise	Escolar Census2011
L_CIENCIAS	1, if science lab; 0, otherwise	Escolar Census2011
S_A_ESPECIAL	1, if special hearing office; 0, otherwise	Escolar Census2011
Q_ESPORTES_C	1, if indoor sports court; 0, otherwise	Escolar Census2011
Q_ESPORTES_D	1, if outdoor sports court; 0, otherwise	Escolar Census2011
COZINHA	1, if kitchen; 0, otherwise	Escolar Census2011
BIBLIOTECA	1, if library; 0, otherwise	Escolar Census2011
S_LEITURA	1, if reading room; 0, otherwise	Escolar Census2011
P_INFANTIL	1, if playground; 0, otherwise	Escolar Census2011
BERCARIO	1, if nursery room; 0, otherwise	Escolar Census2011
SANIT_FORA	1, if toilet outside building; 0, otherwise	Escolar Census2011
SANIT_EI	1, if proper toilet to children; 0, otherwise	Escolar Census2011

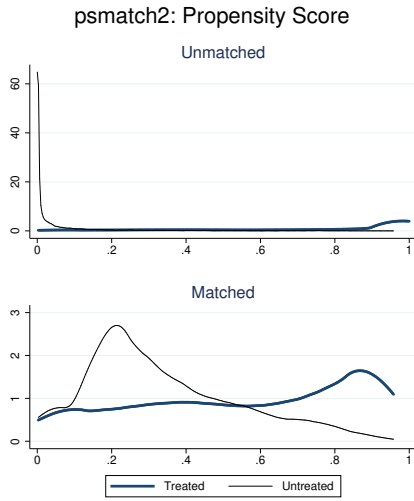
Source: Own calculations.

Table A.1: Description of School Variables (continued)

Variables	Description	
SANIT_PNE	1, if bathroom suitable for the disabled; 0 otherwise	Escolar Census2011
SECRETARIA	1, if there is secretary; 0, otherwise	Escolar Census2011
BANHEIRO	1, if bathroom; 0, otherwise	Escolar Census2011
REFEITORIO	1, if there is dining room; 0, otherwise	Escolar Census2011
DESPENSA	1, if there is food storage; 0, otherwise	Escolar Census2011
ALMOXARIFADO	1, if warehouse; 0, otherwise	Escolar Census2011
AUDITORIO	1. if auditorium; 0, otherwise	Escolar Census2011
PATIO_C	1, if indor courtyard; 0, otherwise	Escolar Census2011
ALOJAM_A	1, if accomodation for students; 0, otherwise	Escolar Census2011
ALOJAM_P	1, if accomodation for teachers; 0, otherwise	Escolar Census2011
AREA_VERDE	1, if green area; 0, otherwise	Escolar Census2011
LAVANDERIA	1, if laundry; 0, otherwise	Escolar Census2011
N_SALAS	Total number of classrooms	Escolar Census2011
N_SALAS_U	Number of classrooms used	Escolar Census2011
TV	1, if TV; 0, otherwise	Escolar Census2011
VIDEOCASSETTE	1, VCR; 0, otherwise	Escolar Census2011
DVD	1, if DVD; 0, otherwise	Escolar Census2011
PARABOLICA	1, if satellite antena ; 0, otherwise	Escolar Census2011
COPIADORA	1, if copy machine; 0, otherwise	Escolar Census2011
RETRO	1, if slide projector; 0, otherwise	Escolar Census2011
IMPRESSORA	1, if printer; 0, otherwise	Escolar Census2011
E_SOM	1, if sound equipment; 0, otherwise	Escolar Census2011
E_MULTIMIDIA	1, if multimedia; 0, otherwise	Escolar Census2011
E_FAX	1, if fax machine; 0, otherwise	Escolar Census2011
E_FOTO	1, if camera; 0, otherwise	Escolar Census2011
N_COMP	Total number of computers	Escolar Census2011
INTERNET	1, if internet; 0, otherwise	Escolar Census2011
N_FUNC	Total number of employees	Escolar Census2011
AEE	1, if special attendance; 0, otherwise	Escolar Census2011
M_ATIV_C	1, if any complement activities; 0, otherwise	Escolar Census2011

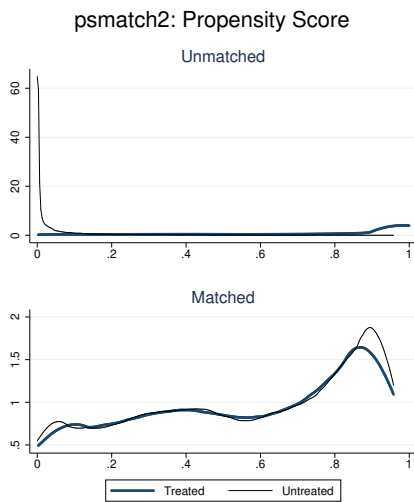
Source: Own calculations.

Figure A.2: Propensity score overlapping via the nearest-neighbor method



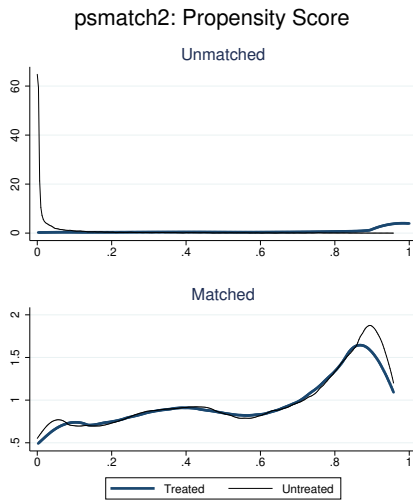
Source: Own elaboration.

Figure A.3: Propensity score overlapping via the kernel method



Source: Own elaboration.

Figure A.4: Propensity score overlapping via the radius method



Source: Own elaboration.

Table A.2: Estimation of Propensity Score post Lasso

Variables	Coefficients	p-value	Variables	Coefficients	p-value
E_foto	-1.8178	0.000	despensa	-0.4902	0.000
Parabólica	-0.6771	0.000	E_gerador	-1.4019	0.000
Retro	-0.5597	0.000	S_leitura*parabolica	-1.0987	0.000
E_som	-1.3684	0.000	Sanit_fora	1.66349	0.000
(N_comp) ²	-0.2696	0.000	AEE	-1.5638	0.000
videocassete	0.0025	0.000	AEE*área_verde	-1.6719	0.000
(N_comp) ³	-0.0001	0.000	Lixo_enterra	3.6820	0.401
(N_func) ³	-5.49E(-06)	0.000	Lixo_enterra*área_verde	0.0480	0.825
(N_sala_u) ³	-0.0010	0.000	Esgoto_rede_pub	0.5221	0.000
Agua_poço	0.8079	0.000	N_comp*Prof_mt	0.0001	0.000
Agua_cacimba	-0.7612	0.000	N_comp*Prof_pt	-0.0001	0.000
bercario	-4.2957	0.000	N_comp*raça	0.0020	0.168
Lixo_coleta	0.5127	0.000	N_comp*e_mae1	-0.0017	0.302
Mq_lavar_casa	0.1093	0.130	E_mae1	-0.7750	0.000
S_leitura	0.1231	0.268	E_mae2	-0.3356	0.000
Internet_casa	0.1813	0.147	E_pai1	-0.2806	0.047
Auditório*internet	0.1101	0.180	E_pai3	0.6220	0.000
E_foto*Area_verde	0.1147	0.159	E_som*área_verde	-0.2412	0.001
banheiro	2.0626	0.000	Prof_port*n_sala_existente	0.0014	0.000
Refeitório	1.9734	0.000	constante	-4.6691	0.000

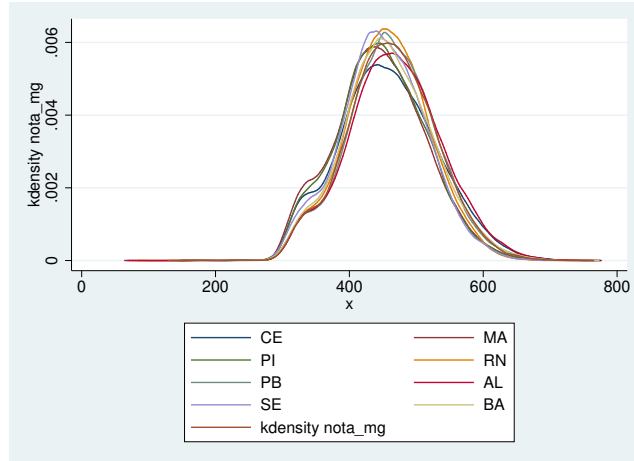
Source: Own calculations.

Table A.3: Overall Mean in ENEM , Brazilian States, 2009 and 2011

State	2009				2011				2011/2009			
	Grade		Participation		Grade		Participation		Grade		Participation	
	Value	Rank	%	Rank	Value	Rank	%	Rank	Value	Rank	Value	Rank
AC	438.6	26	0.16	26	443.7	26	1.07	22	1.17%	4	568.75%	1
PA	479.1	9	0.5	23	454.5	16	1.79	17	-5.13%	27	258.00%	2
CE	448.8	2	2.78	14	449.5	19	8.89	4	0.16%	8	219.78%	3
CE-EM	448	-	4.72	-	442.5	-	7.6	-	-	-	-	-
CE-EP	514.9	-	0.13	-	480.8	-	1.52	-	-	-	-	-
AL	451.3	18	0.11	27	456	15	0.32	27	1.04%	5	190.91%	4
RR	441.1	25	0.23	24	448.6	21	0.46	25	1.71%	1	100.00%	5
AP	457.8	16	0.2	25	450.3	18	0.38	26	-1.66%	22	90.00%	6
SP	509.9	1	9.81	3	491.5	4	18.58	1	-3.61%	26	89.40%	7
GO	473.8	11	2.01	17	468.6	11	3.72	7	-1.09%	20	85.07%	8
PB	462.8	14	1.04	22	453.9	17	1.63	20	-1.92%	23	56.73%	9
PR	481.4	8	6.43	5	481.3	7	6.36	5	-0.03%	11	-1.09%	10
PI	449.8	19	2.01	16	448.6	22	1.89	15	-0.25%	12	-5.97%	11
RO	472.4	12	1.84	19	466.7	12	1.66	18	-1.22%	21	-9.78%	12
PE	490.9	6	2.9	13	479	8	2.51	13	-2.44%	25	-13.45%	13
MS	477	10	3	11	473.7	9	2.59	12	-0.69%	18	-13.67%	14
RJ	497.9	3	3.88	9	494.4	2	3.25	9	-0.70%	19	-16.24%	15
RS	496.7	4	12.69	2	494.3	3	10.18	3	-0.49%	16	-19.78%	16
TO	446.2	23	1.95	18	444.5	24	1.56	21	-0.36%	14	-20.00%	17
MG	491.5	5	13.99	1	489.2	5	11.14	2	-0.47%	15	-20.37%	18
DF	506.7	2	2.72	15	495.3	1	2.15	14	-2.24%	24	-20.96%	19
AM	437.6	27	4.43	7	443.7	25	3.45	8	1.38%	2	-22.12%	20
BA	470.6	13	4.07	8	469.1	10	2.7	11	-0.34%	13	-33.66%	21
MA	444.4	24	4.58	6	441.5	27	2.84	10	-0.64%	17	-37.99%	22
MT	456.5	17	3.14	10	462.6	13	1.89	16	1.33%	3	-39.81%	23
ES	461.6	15	9.75	4	461.7	14	5.81	6	0.02%	10	-40.41%	24
SE	446.2	22	1.24	21	448.5	23	0.72	24	0.51%	7	-41.94%	25
SC	483.8	7	2.98	12	487.9	6	1.66	19	0.85%	6	-44.30%	26

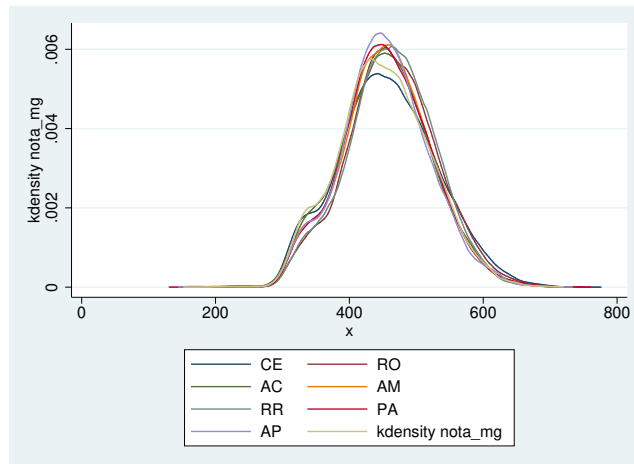
Source: Own calculations.

Figure A.5: Overall Mean Performance in ENEM, Ceará and Northeastern States - 2014



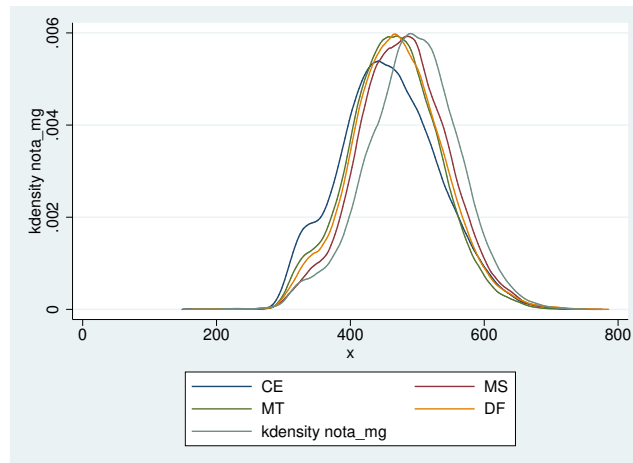
Source: Own elaboration.

Figure A.6: Overall Mean Performance in ENEM, Ceará and Northern States - 2014



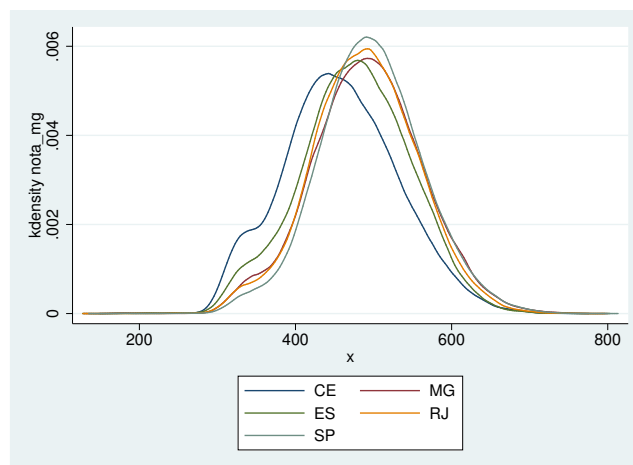
Source: Own elaboration.

Figure A.7: Overall Mean Performance in ENEM, Ceará and Mid-western States - 2014



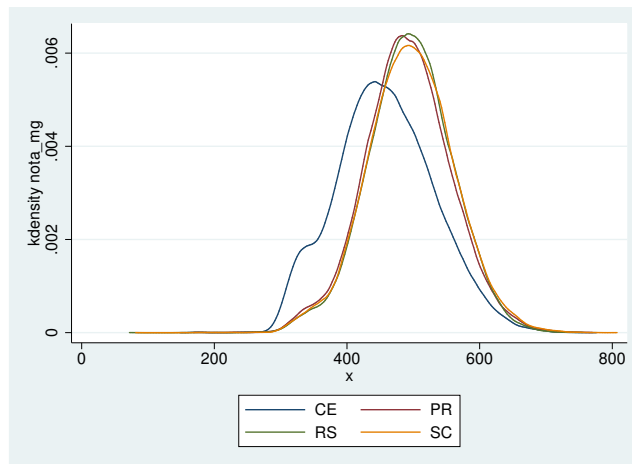
Source: Own elaboration.

Figure A.8: Overall Mean Performance in ENEM, Ceará and Southeastern States - 2014



Source: Own elaboration.

Figure A.9: Overall Mean Performance in ENEM, Ceará and Southern States - 2014



Source: Own elaboration.

