Universidade Federal de Pernambuco
Centro de Ciências Sociais Aplicadas
Programa de Pós-graduação em Economia

Robson Douglas Tigre Santos

The impact of commuting duration on youth's academic performance: could mobility compromise school achievement?

Recife,
2014

Robson Douglas Tigre Santos

The impact of commuting duration on youth's academic performance: could mobility compromise school achievement?

Dissertação apresentada ao Programa de Pós-graduação em Economia da Universidade Federal de Pernambuco como requisito para obtenção do título de mestre em Ciências Econômicas, sob a orientação da Professora Tatiane Almeida de Menezes e coorientação do Professor Breno Ramos Sampaio.

Recife,
2014

Catalogação na Fonte
Bibliotecária Ângela de Fátima Correia Simões, CRB4-773

S237i Santos, Robson Douglas Tigre
The impact of commuting duration on youth’s academic performance: could mobility compromise school achievement? / Robson Douglas Tigre Santos . - Recife : O Autor, 2014.
28 folhas : il. 30 cm.

Orientadora: Profª. Dra. Tatiane Almeida de Menezes.
Dissertação (Mestrado em Economia) – Universidade Federal de Pernambuco, CCSA, 2014.
Inclui referências e apêndices.

1. Rendimento escolar. 2. Desempenho. 3. Mobilidade social. I. Menezes, Tatiane Almeida (Orientador). II. Título.

331 CDD (22.ed.) UFPE (CSA 2015 –18)

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE ECONOMIA
PIMES/PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE DISSERTAÇÃO DO MESTRADO EM ECONOMIA DE:

ROBSON DOUGLAS TIGRE SANTOS

A Comissão Examinadora composta pelos professores abaixo, sob a presidência do primeiro, considera o Candidato Robson Douglas Tigre Santos **APROVADO**.

Recife, 15 de dezembro de 2014.

Prof^a^. Dr^a^. Tatiane Almeida de Menezes
Orientadora

Prof. Dr. Breno Ramos Sampaio
Co-orientador e Examinador Interno

Profª. Drª. Gisleia Benini Duarte
Examinador Externo/UFRPE

Acknowledgements

I would like to thank the Joaquim Nabuco Institute, especially Mrs. Isabel Raposo, for the several clarifications on the database. I am also indebted to the Lemann Institute for Brazilian Studies, Werner Baer, Geoffrey Hewings, Gisleia Benini, and the seminar participants at the Regional Economics Applications Laboratory (R.E.A.L), University of Illinois at Urbana-Champaign, and at the Federal University of Pernambuco for the support, helpful comments and suggestions. Additionally, I must say that this research would not be possible without the grant provided by the National Council for Scientific and Technological Development (CNPq).

“The disciple is learning when he does not know that he is learning, and as a result he may well chafe. In winter, (...) a tree is collecting nutrient. People may think that it is idle, because they do not see anything happening. But in spring they see the buds. Now, they think, it is working. There is a time for collecting, and a time for releasing. This brings the subject back to the teaching: ’Enlightenment must come little by little – otherwise it would overwhelm’.”

Idries Shah - The Sufis, 1964.

Resumo

Inúmeros trabalhos acadêmicos tem analisado variáveis que supostamente afetam o desempenho escolar, especialmente aquelas mais suscetíveis a políticas públicas. A maioria desses estudos, entretanto, tende a focar em fenômenos restritos ao ambiente escolar por si, enquanto menos atenção é dada a fatores que antecedem o horário de início das aulas. Aproveitando a abundância de informações relevantes disponíveis em uma recente base de dados da Fundação Joaquim Nabuco, nós estimamos o efeito do tempo de viagem da casa até a escola sobre a performance dos alunos, utilizando técnicas de Propensity Score Matching (PSM). Adicionalmente, são aplicados testes de sensibilidade baseados em simulação, com o intuito de verificar a robustez das nossas estimativas ao viés de seleção. Os resultados sugerem que um tempo de viagem mais longo tem efeito negativo significante sobre a performance dos alunos. Em média, estudantes com tempo de viagem maior que uma hora tiveram sua nota diminuída em aproximadamente 0.2 desvio padrão, quando comparados com seus pares cujo tempo de viagem é mais curto. Ademais, os resultados se mostram insensíveis tanto à inclusão de variáveis adicionais quanto à falha da Hipótese de Independência Condicional.

Palavras-chave: *Commuting, Tempo de Viagem, Rendimento Escolar, Brasil.*

Abstract

Countless academic researches have been analyzing variables which supposedly affect academic achievement, specially those sensitive to public policy. Most of these studies however tend to focus on phenomena restricted to the school environment, while less attention has been paid to factors present prior to class starting time. Relying on the richness of information of a unique database provided by the Joaquim Nabuco Institute, we estimate the impact of a longer commuting duration on students' performance using Propensity Score Matching (PSM) techniques. We also apply a simulation-based sensitivity analysis in order to check the robustness of our estimates under selection bias. Our results suggest that longer commuting duration has a significant negative impact on academic performance. On average, students with a travel time greater than one hour had their test score decreased approximately by 0.2 standard deviation, when compared to their peers who had shorter commuting duration. Moreover, our results seem to be insensitive to both the inclusion of additional variables and the failure of the Conditional Independence Assumption (CIA).

Keywords: *Commuting, Travel Time, School Achievement, Brazil.*

Contents

1 Introduction

Higher levels of formal education have been consistently linked to economic development and improvement in several well-being outcomes (TEMPLE, 2001; HANUSHEK; WÖßMANN, 2007, for an extensive review), with recent studies persistently showing evidence of externalities from education to crime reduction (LOCHNER; MORETTI, 2004), improvement in the health of children (CURRIE; MORETTI, 2003), and greater civic participation (DEE, 2004; MILLIGAN et al., 2004), just to cite few examples. Such connections, of course, cause education to be a topic of utmost concern to committed policymakers, especially in developing countries such as Brazil, where one of the major sources of regional inequality is the the availability of human capital (RANDS, 2011).

Nevertheless, school supply *per se* is no "silver bullet", and the role of improved schooling has become controversial once noticed that the merely expansion of school attainment has not guaranteed substantially improved economic conditions (HANUSHEK; WÖßMANN, 2007). Formal education then gradually came to be treated as a cumulative function of family, community, and school experiences, each varying in quantity and quality (RIVKIN et al., 2005). Aware of this nuances, policy makers' attention has been shifted to other factors on the education supply side that might help promoting education more efficiently, instead of just providing more schools. As consequence, much research effort is being allocated on better understanding the complex optimization problem faced by the institutionalized educational system, which comprises a virtually infinite number of production inputs, each with different marginal rate of return[1].

Unsurprisingly, in the last 15 years solely, countless works in the education economics literature have investigated very specific variables that are both susceptible to public intervention and supposedly exert impact on academic achievement: teacher's experience and education (RIVKIN et al., 2005), classroom size (ANGRIST; LAVY, 1999; KRUEGER, 2003), teacher absenteeism (BANERJEE; DUFLO, 2006) and class start time (EDWARDS, 2012), just to cite few examples. Some were proven too costly to yield only modest results, as is the case of classroom size, while others were shown to be cost-effective, as the daily class start time, which allegedly has a positive effect on academic performance.

[1] The educational system can be seen as an entity that produces education from scarce and to some degree substitutable inputs. For a brief discussion on the economic theory of the production of education, see Schwartz and Zabel (2005).

Curiously, most of the well-founded investigations focus on factors restricted to the school time and environment, while little attention has been paid to variables that are not part of the school shift and surroundings themselves but could significantly affect academic achievement. A relevant example, though almost neglected by the literature, is the duration of commuting from home to school[2] (i.e, the travel time). In the context of the theory of choice, parents choose the school that maximizes their utility as a function of both school and student characteristics, such as the expected academic achievement a given school can offer to the kid and time and travel costs related to the commuting from home to school (HASTINGS et al., 2005, 2006), respectively. Following this rationale, parents inevitably have to trade-off utility from lower commuting costs in order to gain utility from expected academic outcomes, when there are better schools further from home[3].

Nevertheless, time of students itself is one of the inputs into the educational process (BECKER, 1965), and the additional time spent traveling from home to school could be otherwise allocated on studying, practicing physical activities and sleeping, activities positively related with academic achievement (AHRBERG et al., 2012; PERKINSON-GLOOR et al., 2013). Moreover, outside the context of a sophisticated school planning program (for an example, see HANUSHEK; RAYMOND, 2005), in which information available to parents reflects well the academic standards of a given a school, sometimes it can be very difficult for parents to optimally choose between expected academic gains and present commuting costs in a manner that generates net gain. Thus, we strongly believe public policies intended to facilitate school commuting can play a striking role in promoting school attainment and decrease the differences in the acquisition of human capital, especially in developing communities.

In accordance with previous literature, our data suggest that, in contrast with parents of shorter commuting students, parents of longer commuting students tend to place more weight on expected academic gains than on travel costs when choosing schools, since they predominantly choose where to enroll their kids based on information about academic standards, instead of proximity or ease of enrolling their children (HASTINGS et al., 2005). Moreover, the longer commuters tended to

[2] When the outcome of interest is academic achievement, little is known about the potential effects of commuting, except that these two variables tend to be negatively related (RAUDENBUSH; BHUMIRAT, 1992). A reservation must be made for the case of mild levels of active commuting - from the perspective of physical exercise, which might boost cognitive performance (MARTÍNEZ-GÓMEZ et al., 2011).

[3] Field interviews revealed parents either seemed to choose their neighborhood school without much knowledge of other schools in the district, or seemed to know a lot about academic outcomes for many schools in the district, and based on that, tended to select a preferred school that was typically not their neighborhood school (HASTINGS et al., 2005).

show a lower performance in the test. With this information in mind, we carried out a thorough investigation on whether commuting duration alone could significantly affects students' performance in a scenario of very limited (and sometimes very costly to acquire) information about academic standards of public schools. To the best of our knowledge, the present paper is the only to analyze this potential causal link while encompassing the use of a sample as large as rich in information on schools, teachers, families and students, and to place under close review the possibility of omitted variable be driving our results.

The unique database we use was provided by The Joaquim Nabuco Institute for Social Research, a local agency linked to the Brazilian Ministry of Education and Culture. It is the result of a multilevel study on education conducted during the year of 2013 in Recife, one of the major cities in Brazil, and consists of two standardized math tests applied at the beginning and at the end of the year, respectively, and an *in locus* survey. 118 principals of municipal schools, its sixth-grade student and their respective parents and mathematics teacher responded to questionnaires about their individual characteristics and about objective and subjective characteristics of their schools.

In order to isolate the effect of commuting duration on academic performance from other confounding factors, we start by assuming the conditional independence assumption (CIA) holds (see RUBIN, 1974; HECKMAN; ROBB JR., 1985), and then follow the well known paradigm for the implementation of matching suggested by Caliendo and Kopeinig (2008). Because the relative performance of these estimators hinges so powerfully on features of the data generating process (ZHAO, 2006; BUSSO et al., 2009), we present the estimated average treatment effects obtained via two different matching algorithms and an inverse probability weighted estimator (HIRANO; IMBENS, 2001), for comparisons. We find that children with travel times greater than one hour had their average test score decreased by -0.2 standard deviation, a substantial effect when compared to other results of impact evaluation in education (for an example, see RIVKIN et al., 2005).

Finally, aware of the possibility of our results being driven by a strong endogenous factor, we applied the sensitivity analysis proposed by Ichino et al. (2008). Given that theory does not suggest a broadly accepted functional form for studying this problem, we chose this method as it is the only to assesses the robustness of point estimates (instead of interval estimates) without relying on any parametric model for the outcome equation. First we show that any unobserved factor correlated with each of the covariates considered in this study would not be sufficient to drive our estimated average treatment effect to zero. Subsequently, we perform a set of empirical exercises in order to check the strength of our results. We believe these aforementioned results indicate that duration of commuting has a significant causal negative effect on academic performance, even when considering the possibility of strong endogeneity.

The remainder of this article is organized in 4 additional sections. In section 2 we present the data set we are using. Section 3 displays the description of the methodological approaches adopted for this analysis. Section 4 discusses the results, which we separate in two subsections presenting, respectively, results using least squares methods and propensity score matching methods. Finally, conclusions are presented in section 5.

2 Data

Our data come from "Determinants of School Performance in the Primary Education System of Recife", a multilevel survey undertaken in the city of Recife[4] by The Joaquim Nabuco Institute for Social Research, a local agency linked to the Brazilian Ministry of Education and Culture. The survey was conducted during the year of 2013, and comprehended the application of two standardized mathematics tests and an *in locus* survey, in which principals of 118 public schools[5], its sixth-grade students and their respective parents and math teachers responded to questionnaires about their individual characteristics and other objective and subjective characteristics of the schools. The information was collected using identical instruments (tests and questionnaires) and within the same time period for every individual, which were drawn from the same local education system. These two requisites are of utmost importance for the reduction of bias when applying matching estimators (HECKMAN et al., 1997; MICHALOPOULOS et al., 2004).

Our dependent variable is test scores in a standardized math test, we transform test results into standardized scores, with mean zero and variance approximately equal to one, as in Rivkin et al. (2005), while our variable of interest is the duration of commuting from home to school, which on

[4] Recife is the 9th major city in Brazil and the 3rd largest northeastern Brazilian city, with approximately 1,6 million inhabitants (source: Brazilian Institute of Geography and Statistics (IBGE) - http://www.cidades.ibge.gov.br).

[5] In fact, 120 schools composed the sample. Two of them (*Colégio de Aplicação do Recife* and *Escola da Polícia Militar de Pernambuco*), however, are independently managed by the University of Pernambuco and the Militar Police of Pernambuco, respectively, differing in several dimensions from the others. As it compromises comparability, we decided to remove those two schools from the sample.

the Determinants of School Performance data is presented in hours and minutes. We believe time - instead of distance - comprises more relevant information of the real opportunity cost of commuting for two reasons: (a) it reflects foregone earnings due to time employed in a non-productive activity in the educational process (BECKER, 1965); and (b) distance does not embody costs related with congestion and environmental conditions, moreover, time and monetary commuting costs that are both proportional to distance (GUTIÉRREZ-I-PUIGARNAU; OMMEREN, 2010). Hence, throughout the whole work we use travel time between home and school, namely commuting duration.

In contrast with some studies that rely on devices such as GPS in order to obtain generally adequate measures of commuting (for example, STOPHER et al., 2007), our measure was self reported. Retrospective assessment of time duration, however, is mostly based on memory processes and depends on both the number and the complexity of events that occur during the period to be timed (BISSON et al., 2012). Recent experiments revealed that the actual characteristics of someone's travel are related in a systematic way to how they misreport the "true" measure of travel intensity in surveys (BRICKA et al., 2012). Following this rationale, it could be, for example, that students with a short commuting duration tended to underestimate (under-report) the their travel time, whereas students with a long commuting duration tended to overestimate the duration of their travel. If this hypothetical though very plausible scenario holds, our self-reported continuous variable would incorporate a non-random measurement error, which would lead to inflated estimates of the causal effect[6]. In order to attenuate the endogeneity that may arise from systematic measurement error, we decided to use a binary transformation of the duration variable[7], denoted by C. C equals 1 for individuals with one hour or more of commuting (hereafter long commuters) and 0 for individuals with less than one hour of commuting (hereafter short commuters). We believe this binary transformation is likely to decrease the magnitude of measurement error, since a full hour is a far less subtle unit of time[8].

Table 1 presents descriptive statistics of the variables used throughout our main regression analysis. It is clear that long commuters tend to get significantly lower scores at the end of the school

[6] Measurement error generally describes a situation in which we intend to measure the partial effect of a variable, say T^*, but we can observe only an imperfect measure of it, say T. Without any loss of generality, suppose that $T = T^* + e$, e being a random variable with unknown distribution. Depending on how e and T^* are related with each other, which we can't really observe in the sample, the model may lead to biased coefficient estimates (WOOLDRIDGE, 2010).

[7] As will be discussed later, the use of the binary transformation does not compromise the generality of our results. Further, we also provide additional results for other definitions of *commuting duration*, shown in Table 10, in the Appendix.

[8] Very similar results were produced when we tested other duration variables with cutoffs around one hour

year, when compared to short commuters. The slight difference we can observe along the other dimensions, however, could produce misleading policy implications if inference is made by direct comparison between the two groups. This is one of the concerns that guide our empirical strategy described in the next section.

Since some of the the nomenclature we adopted in Table 1 may be not so subtle to understand at a first glance, that is to say: *School quality* denotes if the parent enrolled the child in a given school based on information about the quality of this school; *Bolsa Família* indicates whether the family attends an homonym government assistance program directed to parents that maintain their kids enrolled and regularly attending school; *Temporary shutdown* denotes if the school experienced any temporary shutdown along the year in which the survey was undertaken, what possibly could compromise the schooling of its students and lead to lower grades on the standardized test; *Offers math book* indicates whether the public school lends gratuitous math books to students in the beginning of the school year, thus providing more subsidy for students to score higher. A detailed description of all variables relevant to this work is presented in Table 9.

As we will discuss later, if commuting/studying choices and academic performance are jointly determined by any personal/school characteristics not controlled for in the model, our results would be potentially biased (selection bias). Two immediate sources of selection into treatment can stem from household choice and enrollment decision. For example, parents who attribute higher value to education may either decide to live closer to good schools, thus, decreasing the potential commuting duration, or to enroll their kid in a better school but farther from home, increasing the potential commuting duration (HASTINGS et al., 2005). Considering that we are studying low-income families living in Brazilian urban areas, we have strong reasons for believing that these two aforementioned selection issues are despicable, when compared to the context faced in the United States of America, where school characteristics may play an important role on neighborhood choices.

Indeed, 26.63% of the parents enrolled their kids in a given school for having information about the quality of that school, a characteristic for which we control, while 45,8% chose the closest school to their homes. Moreover, on the locational choice of home, only 1,63% of the parents chose to live in a given neighborhood based on proximity to schools. Thus, we believe that the urbanization dynamics in Recife, particularly for the group of study, provides a favorable scenario for our analysis.

Table 1: Summary Statistics

	Long commuters		Short commuters	
	mean	stand. dev.	mean	stand. dev.
Student				
2nd test score	-0.184	0.939	0.036	1.000
Gender (male=1)	0.413	0.493	0.504	0.500
Age	11.323	0.910	11.311	0.958
Race (black=1)	0.104	0.307	0.127	0.333
Has already failed	0.274	0.447	0.261	0.439
Works	0.050	0.218	0.061	0.238
Family				
Mother is the responsible	0.856	0.352	0.8330	0.373
School quality*	0.318	0.467	0.259	0.438
Responsible's education	8.463	3.365	8.749	3.324
Per capita income (ln)	5.198	1.254	5.233	1.128
Bolsa Família	0.597	0.491	0.582	0.493
Responsible works	0.497	0.501	0.512	0.499
School				
Temporary shutdown	0.408	0.493	0.441	0.496
Teacher's experience	0.587	0.493	0.647	0.477
Roll call	0.985	0.121	0.977	0.148
Provides math book	0.756	0.430	0.749	0.433
Mandatory tests	0.861	0.347	0.892	0.310
Observations	201		2,617	

* One might suggest also to use information of the school census National Assessment of Educational Achievement (ANRESC), also called "Prova Brasil", as a proxy for school quality. However schools must have at least 20 students enrolled in grades evaluated to be eligible, which is not the case for 34 of the 118 schools in our sample.

In Table 2, above, we display this information by commuting status. Unfortunately, for our purpose, since the question which comprised this information was not of multiple choice, parents could not report more than one major factor on which they based their school choice. One may notice that both type of parents appear place more weight on both proximity and academic information when choosing in which school to enroll their kids, in accordance with the previous literature (HASTINGS et al., 2005). As previously said, parents of long commuters tend to place more weight on information they have about academic standards than parents of short commuters when choosing in which school to enroll their kids. Furthermore, the residential location decision seems to be predominantly driven by previous factors not that are not related to school choice, since on average only 1.6% of the parents take into account proximity to schools when choosing a neighborhood in which to live.

Table 2: School and Neighborhood choice of parents

	Long commuters		Short commuters	
	mean	stand. dev.	mean	stand. dev.
School choice				
Nearest school	0.378	0.486	0.458	0.498
Couldn't afford a private school	0.069	0.255	0.061	0.240
Had info on school quality	0.318	0.467	0.259	0.438
Ease of enrollment	0.139	0.347	0.124	0330
School bus available	0.000	0.000	0.006	0.075
Other	0.094	0.293	0.091	0.288
Neighborhood choice				
Proximity to workplace	0.020	0.140	0.032	0.176
Proximity to schools	0.015	0.121	0.016	0.124
Cheaper neighborhood	0.114	0.319	0.121	0.326
Security	0.004	0.070	0.012	0.109
Close to family and friends	0.159	0.367	0.197	0.397
Was born in this neighborhood	0.264	0.442	0.281	0.450
Has acquired home ownership	0.358	0.481	0.292	0.455
Other	0.064	0.246	0.048	0.213
Observations	201		2617	

3 Methodology

In this section, we define our notation, methods and assumptions. We begin by specifying a generic form of the equation of interest. To estimate the effect of commuting on academic performance, consider the following model

$$y = \alpha + \beta C + X^{'}\gamma + \varepsilon \tag{1}$$

where y is an outcome of interest, C is a dummy variable that takes a value equal to 1 when the individual is a long commuter[9] and 0 otherwise, X is a vector of controls, and ε is an error term. The parameter of interest, β, represents the effect of commuting on a specific outcome y.

As is well known, consistently estimating β via equation (1) requires the error term to be uncorrelated with the variable of interest (i.e., $COV(C;\varepsilon) = 0$) or, in other words, that individuals are randomly assigned to long commuting or are assigned on the basis of variables observed by the econometrician - *i.e*, selection on the observed characteristics. If this assumption fails to hold and

[9] This is a broad definition, but in our case we will consider individuals with over a hour of commuting time as a long commuter.

selection into treatment is based on variables unobserved to the researcher but correlated with the outcome of interest (C) - *i.e*, selection on the unobserved characteristics, then the researcher is left with the task of, for example, finding a valid instrumental variable (IV) to correctly estimate the causal effect of commuting. However, finding a convincing IV is not always viable, and one must rely on different identification strategies to make inferences regarding the parameter of interest[10]

To deal with this caveat, two strategy has been jointly used in some empirical works (CALIENDO; KOPEINIG, 2008). The first step is to balance covariates between treatment and control samples, to obtain more precise estimates, which is usually done using propensity score matching estimators[11] (hereafter PSM). Then the researcher can test the sensitivity of the estimate to selection on unobserved variables. Given that theory does not unanimously agree on the specific functional form for outcome equations, and given the evidence on the importance of non-linearity in returns to schooling (BLACK; SMITH, 2004), PSM seems to be the best method available for our setup.

To briefly present the first step of our strategy, let there be two potential outcome variables for individual i, along the lines of Rubin (1974), such that

$$y_i = \begin{cases} y_{1i}, \text{ if } C_i = 1 \\ y_{0i}, \text{ if } C_i = 0 \end{cases} \tag{2}$$

where y_{1i} is the outcome given long commuting and y_{0i} is the outcome without long commuting. The causal effect of the treatment ($C_i = 1$) relative to the control ($C_i = 0$) is defined as the difference between the corresponding potential outcomes $\beta_i = y_{1i} - y_{0i}$. Many population parameters might be of interest. Here, we focus on the average treatment effect on the treated (ATT) which is defined as

$$\beta_{ATT} \quad = \quad E[\beta_i|C = 1] = E[y_{1i} - y_{0i}|C = 1] \tag{3}$$

[10] As a matter of fact, there are empirical works that obtained similar results using both methods, IV estimators and some other that rely on the CIA. Arpino and Aassve (2013), for example, use the strong preference for sons in Vietnam as IV to estimate the impact of fertility on poverty, obtaining estimates that are not dramatically different from those based on the CIA.

[11] The popularity of propensity score matching caused it to be described as “the estimator *du jour* in the evaluation literature” by Smith and Todd (2005).

The problem the researcher faces when estimating equation (3) arises from the fact that comparisons of two outcomes for the same individual, when exposed and when not exposed to the treatment, is an unfeasible task, as the same worker can either be treated or not in the same time period (IMBENS; WOOLDRIDGE, 2009). That is, we only observe one of the two potential outcomes given treatment status, $y_i = y_{0i} + (y_{1i} - y_{0i})C_i$.

Therefore, one must find different individuals (some treated and some not) such that after adjusting for differences in observed characteristics, or pretreatment variables, comparisons can be made (ANGRIST; PISCHKE, 2009). This is precisely the intuition behind matching estimators that, under the conditional independence (CIA) or unconfoundedness assumption (RUBIN, 1974; HECKMAN; ROBB JR., 1985), imply that treatment assignment is independent of potential outcomes conditional on a set of covariates X or, as shown by Rosenbaum and Rubin (1983), on the propensity score, $p(X)$, defined as the conditional probability of being treated, $Pr(C = 1|X)$. In this case, the ATT is obtained by

$$\beta_{ATT} = E[\beta_i|C = 1, p(X_i)] = E[y_{1i} - y_{0i}|C = 1, p(X_i)] \tag{4}$$

Conditioning on the propensity score essentially implies that the distribution of covariates for the untreated individuals are balanced in a way that it looks very similar to the distribution of covariates for the treated individuals, which makes comparisons between outcomes more reasonable than estimates obtained via equation (1). Thus, the matching procedure, under CIA, eliminates any bias due to the non-random selection to treatment, allowing the parameter to be interpreted in a causal way.

To construct the counterfactual for the treated individuals (long commuters), we start by using the nearest-neighbor matching estimator. For long commuter, nearest-neighbor matching selects one (or more in the case of ties) short commuter with the closest propensity score to that of a long commuter. Matching is done with replacement, and ties are equally weighted, as is commonly done in the literature. The estimator for the average treatment effect is given by

$$ATT = \frac{1}{N^T} \sum_{i \in T} \left[y_{1i} - \sum_{j \in C(i)} w_{ij} y_{0i} \right] \tag{5}$$

where T is the set of treated units (therefore, N^T is the number o treated units), $C(i)$ is the set of control units matched to the treated unit i with an estimated value of the propensity score of p_i and defined as $C(i) = \min_j ||p_i - p_j||$, and weights $w_{ij} = \frac{1}{N_i^C}$ if $j \in C(i)$ and $w_{ij} = 0$ otherwise.

As emphasized above, PSM should only be applied if the underlying identifying assumption (CIA) can be credibly invoked based on the informational richness of the data and a detailed understanding of the institutional set-up by which selection into treatment takes place (BLUNDELL et al., 2005), which we believe is our case, as discussed in section 2. Though all matching estimators contrast the outcome of a treated individual with outcomes of comparison group members, the relative performance of estimators also depends on specific features of the data generating process in question (ZHAO, 2006; BUSSO et al., 2009). For this reason we follow the empirical recommendation of presenting results from a variety of approaches (CALIENDO; KOPEINIG, 2008; BUSSO et al., 2009).

It should be clear that there is no 'winner' matching algorithm for all situations. For example, if there are only a few control observations, it makes no sense to match without replacement. On the other hand, if there are a lot of comparable untreated individuals it might be worth using more than one Nearest-Neighbor (either by oversampling or Kernel Matching) to gain more precision in estimates. Pragmatically, it seems sensible to try a number of approaches. Should they give similar results, the choice may be unimportant. Should results differ, further investigation may be needed in order to reveal more about the source of the disparity (CALIENDO; KOPEINIG, 2008). At last, we take advantage of the good distribution overlap to also estimate the ATT through the normalized inverse probability weighted (IPW) estimator (HIRANO; IMBENS, 2001).

4 Empirical Results

The main econometric specification adopted throughout the paper was chose based on both statistical criteria, economic theory, previous literature and information about the specific institutional and urban settings of Recife. We carefully considered an extensive set of variables that included anthropometric characteristics, parents' perception of the neighborhood in which they live, students' perception of their teacher and school, and so on. Some variables were subsequently excluded based on two conditions: lack of statistical improvement to the model[12] and insufficient empirical support on the previous literature.

In Table 3 we present the results obtained via Ordinary Least Squares (OLS) conditioning to different sets of controls. We use the second standardized math test score as outcome for four different specifications. In column (1), we estimated the unconditional relation between commuting duration and test score. As expected, given what was reported on Table 1, the correlation between commuting duration and test score is negative and statistically significant.

In columns (2), (3) and (4) we add controls for personal, family and school characteristics to account for heterogeneity that might affect enrollment decision and school achievement. The coefficient decreases to -0.212 in our main specification, however, still statistically significant. A slight variation was already expected, as we will show later in Table 5 that both groups have statistically similar means for several characteristics. Thus, we confirm the existence of a negative correlation between commuting duration and school achievement.

Table 3: OLS Specifications

	(1)	(2)	(3)	(4)
Commuting	-0.220***	-0.216***	-0.211***	-0.212***
	(0.069)	(0.067)	(0.066)	(0.067)
Student characteristics	NO	YES	YES	YES
Responsible characteristics	NO	NO	YES	YES
School characteristics	NO	NO	NO	YES
Observations	2,816	2,816	2,816	2,816
R^2	0.003	0.051	0.064	0.074
F statistic	10.21***	27.06***	16.76***	13.61***

Note: Standard errors in parentheses. ***, ** and * represent p<1%, p<5% and p<10% respectively.

Although the obtained estimates imply that greater commuting duration and academic performance are statistically related, the validity of the results presented above depends on important assumptions regarding omitted variable bias. That is, by directly comparing long to short commuters, as we have done so far, one would ignore any selection issue that stem from the fact that

[12] Although the inclusion of non-significant variables in the OLS or propensity score specification will not bias our estimates, it can significantly increase their variance.

living/studying choices and academic performance are jointly determined by observed and unobserved personal/school characteristics, hence, subject to bias from variables not accounted for in the model. Therefore, we use propensity score matching techniques to compare long commuters with an appropriate control group of short commuters, in order to quantify the effect of commuting on academic performance using a better balanced sample, which generally implies bias reduction, and then test for selection on unobserved variables.

4.1 PSM Estimations

Before estimating the ATT, the researcher has to specify the set of covariates used in the estimation of the propensity score (*p(X)*), i.e., what variables explain differences on the probability of being a long commuter. The vector of controls should only include variables that influence simultaneously both treatment and outcome (HECKMAN; ICHIMURA; TODD, 1997; DEHEJIA; WAHBA, 2002; SMITH; TODD, 2005)), that is, the probability of having a longer commuting and the standardized test score, respectively. In principle any discrete choice model can be used for this purpose, whilst probit and logit are the most common choices. Here we use the logit model, since the logistic distribution accommodates more density mass in the bounds when compared to the probit model.

Our specification to estimate the propensity score involved regressing the treatment status on all the variables shown in Table 1 but ‘test score’, obviously. The estimated coefficients confirmed that school quality significantly affect the likelihood of being a long commuter (results omitted). The Hosmer-Lemeshow χ^2 test's p-value was 0.616, indicating that we fail to find a significant lack of fit[13].

As previously mentioned, we use various matching procedures to see whether they give similar results: Nearest Neighbor Matching, 4 Nearest Neighbors Matching and Kernel Matching. We present standard errors and, when appropriate, bootstrapped standard errors, to account for the fact that the propensity score is estimated. Our results are quite similar for all five procedures adopted, including the OLS, and imply that longer commuting duration affects significantly and negatively our measure of academic performance. Quantitatively speaking, longer commuting decreases test score by approximately -0.22 standard deviations.

[13] The null hypothesis of the Hosmer-Lemeshow χ^2 model performance test is that the model is correctly specified. Hence we fail to reject that the model is properly specified.

Table 4: PSM - Estimated Average Treatment on the Treated (ATT)

Algorithm	Treated	Control	Difference (ATT)	Standard error	Bootstrapped stand. error
Nearest neighbor	-0.184	0.030	-0.214**	0.098	
4 nearest neighbors	-0.184	0.053	-0.237**	0.077	
Kernel	-0.184	0.024	-0.208***	0.069	0.061
IPW	-0.184	0.026	-0.210***	0.068	

Note: Standard errors in parentheses. ***, ** and * represent p<1%, p<5% and p<10% respectively. As the theoretical validity of the standard bootstrap for nearest-neighbor matching estimators is questionable, we do not display bootstrapped standard errors for those matching estimators (ABADIE; IMBENS, 2008)

4.1.1 Matching Quality and Robustness Checks

In order to assess the matching quality and the robustness of our estimates, we: (a) compare the situation before and after matching in order to check if the matching procedure was able to balance the distribution of the relevant variables in both the control and treatment group; (b) provide results of a simulation-based sensitivity analysis presented by Ichino et al. (2008), to test the reliability of our findings under a potential failure of the Conditional Independence Assumption (CIA); and (c) add a set of intuitively desirable controls, of which some are likely to be "bad controls" (ANGRIST; PISCHKE, p. 64-68, 2009), in order to see whether our estimated ATT would statistically vanish or even change its magnitude in controversial ways.

4.1.2 Balance Tests

To assess whether the matching was effective, we report in Table 5 the results of the t-tests for equal means before and after applying the matching procedure. It is clear that even before matching (unmatched sample), long commuters (treated) are already similar to short commuters (control) in several dimensions: out of the 16 variables included in the table, only 2 had mean differences that were statistically different at conventional levels. Nevertheless, we achieved substantial reduction in bias after matching individuals.

As emphasized by Dehejia and Wahba (2002), a matching procedure will be successful if members in the selected control group (matched sample) have similar observable characteristics (variables included in the vector of covariates) when compared to the treated group. As one can observe in Table 5, not even one of the mean differences were statistically significant after matching.

Table 5: Covariate imbalance testing

	Sample	Treated (mean)	Control (mean)	% bias	t-test p-value	% |bias| reduction
Gender (male=1)	Unmatched	0.412	0.504	-18.3	0.013	-
	Matched	0.412	0.426	-2.7	0.787	85.3
Age	Unmatched	11.323	11.311	1.4	0.856	-
	Matched	11.323	11.324	-0.1	0.993	93.7
Race (black=1)	Unmatched	0.104	0.127	-7.1	0.348	-
	Matched	0.104	0.104	0.1	0.994	98.9
Has already failed	Unmatched	0.273	0.261	2.8	0.703	-
	Matched	0.273	0.271	0.4	0.968	85.3
Works	Unmatched	0.049	0.061	-4.8	0.527	-
	Matched	0.049	0.051	-0.7	0.941	85.2
Mother is the responsible	Unmatched	0.855	0.833	6.3	0.404	-
	Matched	0.855	0.854	0.3	0.973	94.7
School quality	Unmatched	0.318	0.258	13.2	0.064	-
	Matched	0.318	0.311	1.6	0.873	87.5
Responsible's education	Unmatched	8.462	8.749	-8.6	0.240	-
	Matched	8.462	8.478	-0.5	0.963	94.6
Per capita income (ln)	Unmatched	5.199	5.233	-2.9	0.680	-
	Matched	5.199	5.197	0.1	0.992	96.5
Bolsa Família	Unmatched	0.597	0.582	3.0	0.684	-
	Matched	0.597	0.595	0.3	0.973	88.7
Responsible works	Unmatched	0.497	0.512	-3.0	0.684	-
	Matched	0.497	0.498	-0.2	0.982	92.4
Temporary shutdown	Unmatched	0.407	0.441	-6.7	0.358	-
	Matched	0.407	0.411	-0.6	0.948	90.4
Teacher's experience	Unmatched	0.348	0.390	-8.7	0.241	-
	Matched	0.348	0.354	-1.4	0.889	84.0
Roll call	Unmatched	0.985	0.977	5.6	0.478	-
	Matched	0.985	0.985	-0.2	0.984	96.7
Provides math book	Unmatched	0.756	0.749	1.5	0.837	-
	Matched	0.756	0.757	-0.3	0.974	78.8
Mandatory tests	Unmatched	0.860	0.892	-9.6	0.168	-
	Matched	0.860	0.864	-1.2	0.908	87.4

The difference in percentage of male students between long commuters and short commuters in the matched sample, for example, is only -1.34 percentage points (the average for long commuter is 41.29%, and the average for matched short commuters is 42.63%). This percentage difference between the two groups in the original unmatched sample, on the other hand, is approximately -9.1 percentage points (41.29%-50.4%). Ergo, the matching we implemented successfully ensured the balancing property required for the consistency of this procedure, i.e., it perfectly balanced individual characteristics between long and short commuters. This is also clear when looking at the reduction of bias attained for each covariate, presented in the last column of the table. For most covariates, the bias reduction was around 85%, therefore inducing a more precise estimate of the treatment effect.

4.1.3 Sensitivity Analysis

Even though the results displayed in the previous section endorse the validity of our propensity score matching implementation, we recall that the trustworthiness of results obtained via this strategy rely on the conditional independence assumption (CIA). Since this assumption is non-testable by its nature[14], one may still question the plausibility of CIA in our case, and argue that our results (or at least its statistical significance) are probably being driven by an omitted variable strongly correlated with commuting duration.

In an attempt to circumvent this suspicion, we apply the simulation-based sensitivity analysis presented by Ichino et al. (2008) as resource to check the robustness of our estimates. This analysis aims to assess the bias of our estimates when the CIA is assumed to fail in some meaningful ways. A failure in the CIA is equivalent to say that the assignment to treatment is not unconfounded given the set of observable variables X, i.e.,

$$Pr(C = 1|y_0, y_1, X) \neq Pr(C = 1|X) \tag{6}$$

The central assumption in the analysis presented by Ichino et al. (2008), which is common to many other sensitivity analyses (ROSENBAUM; RUBIN, 1983a; IMBENS, 2003; ALTONJI et al., 2005)[15], is that the CIA would hold given X and an unobserved binary factor U. Equation 7 represents the extended CIA in the scenario where we could observe U:

$$Pr(C = 1|y_0, y_1, X, U) = Pr(C = 1|X, U) \tag{7}$$

In turn, the distribution of this unobserved confounding factor can be characterized by simply specifying the parameters p_{ij}, presented in equation 8, which give the probability that $U = 1$ for each of the four groups defined by the treatment status and the outcome value.

[14] Because the data are completely uninformative about the distribution of the untreated outcome (y_{0i}) for treated units ($C_i = 1$).

[15] Despite employing common assumptions, the method proposed by Ichino et al. (2008) has several advantages, such as avoiding a possibly incorrect parametric specification of the distribution of $y|C, U, X$, which is the strategy adopted by competing types of sensitivity analysis (ALTONJI et al. 2005), and retrieving point estimates instead of only bounds (ROSENBAUM; RUBIN, 1983a).

$$p_{ij} \equiv Pr(U = 1|C = i, y = j, X) = Pr(U = 1|C = i, y = j); \quad i, j \in \{0, 1\} \tag{8}$$

Given empirically meaningful values of the parameters p_{ij}, the sensitivity analysis proceeds by attributing a value of U to each subject in the sample[16] according to its belonging to one of the four groups defined by the treatment status and the outcome value. U then is treated as any other observed covariate, and included in the set of matching variables used to estimate the propensity score and to compute the ATT. Employing a given set of values of the sensitivity parameters, we repeat the matching estimation m times to obtain an estimate of the ATT, which is an average of the ATTs over the distribution of the simulated U. Thus, for any given configuration of the parameters p_{ij}, we can retrieve a point estimate of the ATT which is robust to the specific failure of the CIA implied by that configuration.

Regarding the nature of the unobserved variable U, Ichino et al. (2006) presents two Monte Carlo exercises showing that: (a) the assumption about the relation between U and X, presented in equation 8, does not critically affect the results of the sensitivity analysis; (b) the assumption of a binary confounder, when the true one is continuous, tends to produce conservative ATT estimates. Hence, with respect to the modeling assumptions concerning U, this sensitivity analysis should not lead us to infer that the ATT estimates are robust to failures of the CIA when in fact they are not. This result, as said in their paper, is consistent with the finding of Wang and Krieger (2006) that causal conclusions are more sensitive to unobserved binary covariates than to (normal) continuous unobservables.

With this rationale in mind, one may pick the parameters p_{ij} to make the distribution of U mimic the empirical distribution of important covariates. This simulation exercise reveals the extent to which the baseline estimates are robust to deviations from the CIA induced by the impossibility of observing factors similar to those observed covariates. Ichino, Mealli and Nannicini (2008) emphasize that 'this is a different exercise from the simple removal of an observed variable from the matching set X, since in every sensitivity-analysis estimation we still control for all the relevant covariates observed by the econometrician'.

Following the above reasoning, we implement the sensitivity analysis calibrating U to mimic

[16] The imputation of U is considered a normal problem of missing data, which can be solved by repeatedly imputing the missing values of U according to p_{ij} (RUBIN, 1987).

each of the controls presented in Table 1. For the sake of our previous results, table 6 exhibits that any unobserved factor correlated with each of the covariates used in this study would not be sufficient to drive our estimated ATT to zero.

Table 6: Sensitivity analysis: effect of calibrated confounders

	$Pr(U=1\|C=i,y=j)$				Γ	Λ	ATT	SE
	p_{11}	p_{10}	p_{01}	p_{00}				
No confounder	0.00	0.00	0.00	0.00	-	-	-0.214	0.098
Neutral confounder	0.50	0.50	0.50	0.50	1.00	1.00	-0.212	0.121
Confounder-like								
Gender (male=1)	0.43	0.40	0.49	0.52	0.875	0.701	-0.209	0.123
Age								
$\leq$ 1st quartile	0.14	0.10	0.17	0.11	1.570	0.800	-0.208	0.121
$\geq$ 3rd quartile	0.23	0.39	0.25	0.39	0.518	1.047	-0.220	0.124
Race (black=1)	0.10	0.11	0.12	0.14	0.799	0.809	-0.220	0.119
Has already failed	0.21	0.32	0.20	0.33	0.505	1.073	-0.207	0.123
Works	0.01	0.08	0.05	0.07	0.695	0.814	-0.220	0.121
Mother is the responsible	0.80	0.90	0.84	0.83	1.058	1.242	-0.211	0.122
School quality	0.28	0.35	0.28	0.24	1.250	1.359	-0.210	0.121
Responsible's education								
Elementary	0.16	0.24	0.19	0.23	0.790	1.041	-0.216	0.119
Secondary	0.38	0.37	0.31	0.36	0.797	1.184	-0.195	0.122
High school	0.40	0.37	0.47	0.39	1.375	0.796	-0.206	0.118
Higher	0.07	0.02	0.04	0.02	1.674	1.260	-0.209	0.123
Per capita income (ln)								
$\leq$ 1st quartile	0.27	0.27	0.24	0.26	0.921	1.098	-0.220	0.121
$\geq$ 3rd quartile	0.26	0.20	0.26	0.25	1.082	0.836	-0.210	0.123
Bolsa Família	0.53	0.65	0.56	0.61	0.786	1.071	-0.209	0.122
Responsible works	0.48	0.51	0.52	0.50	1.068	0.952	-0.212	0.122
Temporary shutdown	0.37	0.44	0.43	0.45	0.915	0.884	-0.210	0.121
Teacher's experience	0.49	0.67	0.66	0.63	1.190	0.784	-0.202	0.120
Roll call	0.99	0.98	0.98	0.98	1.204	2.021	-0.220	0.121
Provides math book	0.80	0.72	0.78	0.71	1.476	1.061	-0.216	0.122
Mandatory tests	0.83	0.88	0.90	0.88	1.320	0.766	-0.205	0.121

Note: Let U be a binary confounding factor and denote the fraction of $U = 1$ by treatment and outcomes as: $p_{ij} = Pr(U = 1|C = i, y = j)$, with $i, j = 0, 1$. On the basis of these parameters, a value of U is imputed to each individual and the ATT is estimated by nearest neighbor propensity score matching with U in the set of matching variables. The process is repeated 500 times. Γ is the average estimated odds ratio of U in the logit model of $Pr(y = 1|C = 0, U, X)$; Λ is the average estimated odds ratio of U in the logit model of $Pr(C = 1|U, X)$; "ATT" is the average of the simulated ATTs; "SE" is the standard error calculated as shown in Ichino et al., 2008, with proofs presented in Ichino et al., 2006. The first two rows show the ATT estimate with no confounding factor or with a confounder whose outcome and selection effects are insignificant, respectively. In the "confounder-like" rows, U has been calibrated to match the distribution of the corresponding covariate.

4.2 Additional Remarks on Robustness

Despite all the robustness exercises we have presented so far, one still may question why we did not consider controls for the mode of travel used by each student. As a matter of fact, we do have access to this information, as displayed in Table 7, as well as if the student is a newcomer. Intuitively, to control for those covariates could even give us more support to claim a causal connotation. Notwithstanding, an attentive reader should remember that more control is not always better in an attempt to draw causal conclusions, even when their inclusion might be expected to change the regression coefficients, since the inclusion of those controls could come at the expense of introducing some degree of selection bias into the model ("bad control"). In other words, bad control is a variable that could be itself the outcome variable in the analysis at hand (ANGRIST; PISCHKE, p. 64-68, 2009).

Table 7: Additional controls

	Long commuters		Short commuters	
	mean	stand. dev.	mean	stand. dev.
Additional student/household characteristics				
Student is a newcomer	0.721	0.449	0.714	0.452
1st test score	-0.098	0.874	0.037	1.009
Travel mode				
Responsible's vehicle	0.044	0.207	0.036	0.186
Carpool	0.000	0.000	0.009	0.097
Regular bus	0.264	0.442	0.105	0.307
School bus	0.005	0.070	0.007	0.085
Bicycle	0.020	0.140	0.0333	0.179
Walking	0.667	0.472	0.809	0.393
Observations	201		2,617	

Mode of transportation, for example, could be thought as an outcome variable, since the duration of commuting can somehow predict the mode of transportation used by a student. It is also the case of *1st test score*, since it may also be affected by the duration of commuting. In any case, as an additional exercise, we present the estimated ATTs including other relevant (though potentially troublesome) controls in our specification, in order to indicate the robustness of our results. From Table 8, one may notice that our previous results are non-sensitive to the additional information included in the model, as we expected, once even before matching both groups had statistically equal means for many of the variables considered here (see Table 7).

In this robustness exercise, we also included the variables $newcomer$, $1sttestscore$ and an interaction term of the form $newcomer \times 1st\ test\ score$ in the specification, since one may argue that parents of under-achieving students could enroll their children in schools with higher standards hoping the kid would benefit from better teachers or peer effect (HANUSHEK et al., 2003), for example. This would lead to an association between underachievement and long commuting (selection bias). Finally, we address the possibility of a strong unobserved heterogeneity related with neighborhood characteristics by including neighborhood fixed effects (FE). For instance, better neighborhoods might concentrate schools with better infrastructure (BAYER et al., 2007). The results for this exercise are presented in Table 8. Nevertheless, the statistical significance of the estimated ATTs persisted, suggesting the robustness of our results.

Table 8: Estimated ATT including additional controls

	ATT			
	(1)	(2)	(3)	(4)
OLS	-0.157**	-0.213***	-0.221***	-0.142**
	(0.062)	(0.067)	(0.072)	(0.063)
Nearest Neighbor	-0.227***	-0.237***	-0.196*	-0.190*
	(0.097)	(0.098)	(0.104)	(0.102)
4 nearest neighbors	-0.166**	-0.225***	-0.209***	-0.181**
	(0.078)	(0.079)	(0.079)	(0.081)
Kernel	-0.161***	-0.214***	-0.219***	-0.139**
	(0.067)	(0.070)	(0.072)	(0.075)
IPW	-0.171***	-0.223***	-0.209***	-0.159***
	(0.065)	(0.071)	(0.063)	(.057)
Additional student	YES	NO	NO	YES
Travel mode	NO	YES	NO	YES
Neighborhood FE	NO	NO	YES	YES
Observations	2816	2816	2666	*2666*

Note: Standard errors in parentheses. ***, ** and * represent $p<1\%$, $p<5\%$ and $p<10\%$ respectively.

Given the results presented so far, if we consider the least estimated ATT, -0.139 standard deviation, as a lower bound (in absolute terms) for the true causal effect, its magnitude still should not be negligible to policymakers. For instance, in absolute terms, it is comparable to the effect of a variation in teacher quality and equivalent to a class size reduction of ten or more students (RIVKIN et al., 2005).

5 Concluding remarks

Recognizing that the additional time spent traveling from home to school could be otherwise allocated on studying or devoted to physical activities, leisure and sleeping, activities positively related with academic achievement (AHRBERG et al., 2012; PERKINSON-GLOOR et al., 2013), we carried out a thorough investigation on whether commuting duration alone could significantly affects students' performance. We took advantage of an unique survey undertaken by The Joaquim Nabuco Institute for Social Research in the city of Recife, one of the major cities in Brazil. As it is not feasible to randomly assign people to commutes of varying duration, we balanced treated and control samples using different propensity score estimators in order to isolate the effect of commuting duration from observed confounding variables. Conditioning on characteristics of students, family, teacher and school, we found that children with travel time greater than one hour had their test score decreased approximately by -0.2 standard deviations.

Aware of the possibility of our results being driven by a strong endogenous factor, we applied the falsification test presented in Ichino et al. (2008), since this method is the only to assesses the robustness of point estimates without relying on any parametric model for the outcome equation, and performed a set of empirical exercises in order to check the strength of our results. We believe that the results we obtained indicate that duration of commuting has a significant causal negative effect on academic performance, even when considering the possibility of strong endogeneity. For instance If we consider the least estimated ATT, -0.139 standard deviation, as a lower bound (in absolute terms) for the true causal effect, its magnitude still should not be negligible to policymakers. For instance, in absolute terms, it is comparable to the effect of a variation in teacher quality and equivalent to a class size reduction of ten or more students (RIVKIN et al., 2005).

To the best our knowledge the present work is the only to analyze this effect on academic performance while encompassing the use of a large and informational rich sample, and to put under scrutiny the possibility of omitted variable be driving our results. As urban policy has been one of the most discussed social-science topic worldwide (KING; PENDLEBURY, 2013), and considering the nonexistence of a robust work on the specific question we study, specially involving a developing country, our findings could serve to guide policies on urban mobility in order to promote schooling.

Finally, it is important to remember that the analysis presented above was not carried out without any caveats, and should be seen as an initial approach intended to shed light upon something that appears to be a relevant obstacle to a more efficient promotion of education. Firstly, despite our effort to isolate the effect of commuting from other unobserved confounders, commuting time assignment possibly has a considerably share of endogeneity (specially in cities unlikely Recife, where neighborhood mobility is not concern for the selection into treatment perspective). At last, we have no evidence to discuss through which mechanism a longer commuting duration may affect academic performance. Since depending on the mechanism different policies should be implemented, we encourage further studies on this topic. At last, even though we truly believe that the present work brings a genuine contribution to both education and urban studies literature, as strong evidence is provided to support a causal interpretation, much effort must still be devoted to this point in order to validate and generalize analogous results.

References

[1] ABADIE, A.; IMBENS, G.W. On the Failure of the Bootstrap for Matching Estimators. **Econometrica**, v. 76, n. 6, p. 1537-1557, 2008.

[2] AHRBERG, K.; DRESLER, M.; NIEDERMAIER, S.; STEIGER, A.; GENZEL, L. The interaction between sleep quality and academic performance. **Journal of psychiatric research**, v. 46, n. 12, p. 1618-1622, 2012.

[3] ALTONJI, J.; ELDER, T.; TABER, C. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. **Journal of Political Economy**, v. 113, n. 1, p. 151-184, 2005.

[4] ANGRIST, J.; LAVY, V. Using Maimondides' rule to estimate the effect of class size on scholastic achievement. **Quarterly Journal of Economics**, v. 114, n. 2, p. 533-575, 1999.

[5] ANGRIST, J.; PISCHKE, J. **Mostly Harmless Econometrics: An Empiricist's Companion**. Princeton University Press, Princeton, 2009.

[6] ARPINO, B.; AASSVE, A. Estimating the causal effect of fertility on economic wellbeing: data requirements, identifying assumptions and estimation methods. **Empirical Economics**, v. 44, n. 1, p. 355-385, 2013.

[7] BANERJEE, A.; DUFLO, E. Addressing Absence. **Journal of Economic Perspective**, v. 20, n. 1, p. 117-132, 2006.

[8] BAYER, P.; FERREIRA, F.; MCMILLAN, R. A Unified Framework for Measuring Preferences for Schools and Neighborhoods, **Journal of Political Economy**, v. 115, n. 4, p. 588-638, 2007.

[9] BECKER, G. A Theory of the Allocation of Time. **The economic journal**, v. 75, n. 299, p. 493-517, 1965.

[10] BISSON, N.; TOBIN, S.; GRONDIN, S. Prospective and Retrospective Time Estimates of Children: A Comparison Based on Ecological Tasks. **PLoS ONE**, v. 7, n. 3, e33049, 2012.

[11] BLACK, D.; SMITH, J. How robust is the evidence on the effects of college quality? Evidence from matching. **Journal of Econometrics**, v. 121, n. 1-2, p. 99-124, 2004.

[12] BLUNDELL, R.; DEARDEN, L.; SIANESI, B. Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS. **Journal of the Royal Statistical Society, Series A**, v. 168, n. 3, p. 473-512, 2005.

[13] BRICKA, S. G.; SEN, S.; PALETI, R.; BHAT, C. R. An analysis of the factors influencing differences in survey-reported and GPS-recorded trips. **Transportation research part C: emerging technologies**, v. 21, n. 1, p. 67-88, 2012.

[14] BUSSO, M.; DINARDO, J.; MCCRARY, J. New evidence on the finite sample properties of propensity score reweighting and matching estimators. **Discussion Paper**, 2009. (Institute for the Study of Labor, Working Paper No. 3998)

[15] CALIENDO, M.; KOPEINIG, S. Some Practical Guidance for the Implementation of Propensity Score Matching. **Journal of Economic Surveys**, v. 22, n. 1, p. 31-72, 2008.

[16] CASASANTO, D.; BORODITSKY, L. Time in the mind: Using space to think about time, **Cognition**, v. 106, n. 2, p. 579-593, 2008.

[17] CURRIE, J.; MORETTI, E. Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings. **Quarterly Journal of Economics**, v. 118, n. 4, p. 1495-1532, 2003.

[18] DEE, T.S. Are there civic returns to education? **Journal of Public Economics**, v. 88, n. 9-10, p. 1697-1720, 2004.

[19] DEHEJIA, R.; WAHBA, S. Propensity score-matching methods for nonexperimental causal studies. **The Review of Economics and Statistics**, v. 84, n. 1, p. 151-156, 2002.

[20] DROIT-VOLET, S. Time perception in children: A neurodevelopmental approach. **Neuropsychologia**, v. 51, n. 2, p. 220-234, 2013.

[21] EDWARDS, F. Early to rise? The effect of daily start times on academic performance. **Economics of Education Review**, v. 31, n. 6, p. 970-983, 2012.

[22] GUTIÉRREZ-I-PUIGARNAU, E.; VAN OMMEREN, J. Labour supply and commuting. **Journal of Urban Economics**, v. 68, n. 1, p. 82-89, 2010.

[23] HANUSHEK, E.; KAIN, J.; MARKMAN, J.; RIVKIN, S. Does peer ability affect student achievement? **Journal of Applied Econometrics**, v. 18, n. 5, p. 527-544, 2003.

[24] HANUSHEK, E.; RAYMOND, M.E. Does school accountability lead to improved student performance? **Journal of policy analysis and management**, v. 24, n. 2, p. 297-327, 2005.

[25] HANUSHEK, E.; WÖSSMANN, L. **The Role of Education Quality for Economic Growth**. The World Bank, Washington DC, 2007.

[26] HASTINGS, J.; KANE, T.; STAIGER, D. Parental preferences and school competition: Evidence from a public school choice program. **Working Paper**, 2005. (National Bureau of Economic Research, Working Paper No 11805).

[27] HASTINGS, J.; KANE, T.; STAIGER, D. Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery. **Working Paper**, 2006. (National Bureau of Economic Research, Working Paper No. 12145).

[28] HECKMAN, J.; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. **Review of Economic Studies**, v. 64, n. 4, p. 605-654, 1997.

[29] HECKMAN, J.; ROBB JR, R. Alternative methods for evaluating the impact of interventions: An overview. **Journal of Econometrics**, v. 30, n. 1-2, p. 239-267, 1985.

[30] HIRANO, K.; IMBENS, G. W. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. **Health Services and Outcomes research methodology**, v. 2, n. 3-4, 259-278, 2001.

[31] ICHINO, A.; MEALLI, F.; NANNICINI, T. (). From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity? **Discussion Paper**, 2006. (Institute for the Study of Labor, Working Paper No. 2149)

[32] ICHINO, A.; MEALLI, F.; NANNICINI, T. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? **Journal of Applied Econometrics**, v. 23, n. 3, p. 305-327, 2008.

[33] IMBENS, G. Sensitivity to exogeneity assumptions in program evaluation. **American Economic Review Papers and Proceedings**, v. 93, n. 2, p. 126-132, 2003.

[34] IMBENS, G.; WOOLDRIDGE, J. Recent Developments in the Econometrics of Program Evaluation. **Journal of Economic Literature**, v. 47, n. 1, p. 5-86, 2009.

[35] KING, C.; PENDLEBURY, D.A. **Research Fronts 2013: 100 Top-Ranked Specialties in the Sciences and Social Sciences**. Thomson Reuters, 2013.

[36] KRUEGER, A.B. Economic Considerations and Class Size, **The Economic Journal**, v. 113, n. 485, pages F34-F63, 2003.

[37] LANDEK, M.M. **An Examination of Commuter and Residential Student Time Allocation and Relationship to Student Retention**. Ph.D thesis, University of Illinois at Chicago, 2013. Available at: <https://indigo.uic.edu/handle/10027/9991>. [July 2, 2014].

[38] LOCHNER, L.; MORETTI, E. The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. **American Economic Review**, v. 94, n. 1, p. 155-189, 2004.

[39] MARTÍNEZ-GÓMEZ, D. et al. Active commuting to school and cognitive performance in adolescents: The avena study, **Archives of Pediatrics & Adolescent Medicine**, v. 165, n. 4, p. 300-305, 2011.

[40] MCLENNAN, P.; BENNETTS, M. The journey to work: a descriptive UK case study. **Facilities**, v. 21, p. 180-187, 2003.

[41] MICHALOPOULOS, C.; BLOOM, H.; HILL, C. Can Propensity-Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs? **Review of Economics and Statistics**, v. 86, n. 1, p. 156-179, 2004.

[42] MILLIGAN, K.; MORETTI, E.; OREOPOULOS, P. Does education improve citizenship? Evidence from the United States and the United Kingdom. **Journal of Public Economics**, v. 88, n. 9-10, p. 1667-1695, 2004.

[43] NOVACO, R.W.; GONZALEZ, O.I. Commuting and well-being. In: Amichai-Hamburger, Y. ed. **Technology and Psychological Well-being**. Cambridge University Press, New York, p. 174-205, 2011.

[44] PERKINSON-GLOOR, N.; LEMOLA, S.; GROB, A. Sleep duration, positive attitude toward life, and academic achievement: The role of daytime tiredness, behavioral persistence, and school start times, **Journal of Adolescence**, v. 36, n. 2, p. 311-318, 2013.

[45] RANDS, A. **Desigualdades regionais no Brasil**. Elsevier Brasil, 2011.

[46] RAUDENBUSH, S.W.; BHUMIRAT, C. The distribution of resources for primary education and its consequences for educational achievement in Thailand. **International Journal of Educational Research**, v. 17, n. 2, p.143-164, 1992.

[47] RIVKIN, S.; HANUSHEK, E.; KAIN, J. Teachers, schools, and academic achievement. **Econometrica**, v. 73, n. 2, p. 417-458, 2005.

[48] ROSENBAUM, P.; RUBIN, D. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. **Journal of the Royal Statistical Society Series B**, v. 45, n. 2, p. 212-218, 1983a.

[49] ROSENBAUM, P.; RUBIN, D. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41-55, 1983b.

[50] RUBIN, D. Estimating causal effects of treatments in randomized and nonrandomized studies. **Journal of Educational Psychology**, v. 66, n. 5, p. 688-701, 1974.

[51] RUBIN, D. **Multiple Imputation for Nonresponse in Surveys**. John Wiley & Sons, New York. ISBN 0-471-08705-X, 1987.

[52] SCHWARTZ, A.E.; ZABEL, J. The Good, the Bad, and the Ugly: Measuring School Efficiency Using School Production Functions. **In Measuring School Performance and Efficiency: Implications for Practice and Research**, ed. L. Stiefel, A. E. Schwartz, R. Rubenstein and J. Zabel. NY: Eye on Education, Inc. p. 37-66, 2005.

[53] SMITH, J.; TODD, P. Does matching overcome Lalonde's critique of nonexperimental estimators? **Journal of Econometrics**, v. 125, n. 1-2, p. 305-353, 2005.

[54] STOPHER, P.; FITZGERALD, C.; XU, M. Assessing the accuracy of the Sydney Household Travel Survey with GPS. **Transportation**, v. 34, n. 6, p. 723-741, 2007.

[55] TEMPLE, J. Growth effects of education and social capital in the OECD countries. **OECD Economic Studies**, v. 33, p. 57-101, 2001;

[56] WANG, L.; KRIEGER, A. Causal conclusions are most sensitive to unobserved binary covariates. **Statistics in Medicine**, v. 25, n. 13, p. 2257-2271, 2006.

[57] WENER, R.E.; EVANS, G.W. A Morning Stroll: Levels of Physical Activity in Car and Mass Transit Commuting. **Environment and Behavior**, v. 39, n.1, p. 62-74, 2007.

[58] WOOLDRIDGE, J.M. **Econometric analysis of cross section and panel data**. MIT press, Cambridge, 2010.

[59] ZHAO, Z. (2006). Matching Estimators and the Data from the National Supported Work Demonstration Again. **Discussion Paper**, 2006. (Institute for the Study of Labor, Working Paper No. 2375)

Appendices

Table 9: Description of the variables included in our estimations.

Variable name	Description
Student	
1st test score	Standardized test score obtained obtained by the student in the first math test applied by the Joaquim Nabuco Institute, during the second semester of 2013.
2nd test score	Standardized test score obtained by the student in the second math test applied by the Joaquim Nabuco Institute, during the second semester 2013.
Gender (male=1)	Indicates student gender.
Age	Age of the student in years.
Color (black=1)	Indicates whether the student declared her/himself as being black.
Student is a newcomer	Indicates whether the student is a newcomer in her/his currently school.
Has already failed	Indicates whether the student has ever failed a school year.
Works	Indicates whether the student is currently working
Family	
Mother is the responsible	Indicates whether the responsible for monitoring and assisting in school activities is the mother of the student.
School quality	Indicates whether the school choice was mainly based on information about school quality.
Responsible's education	Formal education of the responsible in years of study.
Per capita income (ln)	Logarithm of the household per capita income.
Bolsa Família	Indicates whether the household receives cash transfers from the federal program Bolsa Família.
Responsible works	Indicates whether the responsible works outside the household.
Teacher and School	
Temporary shutdown	Indicates whether the school suffered from temporary shutdown during 2013.
Teacher's experience	Indicates teacher's class-room experience in years.
Roll call	Indicates whether the teacher does roll call to control students' attendance.
Provides math book	Indicates whether the school provided math books to students free of charge*.
Mandatory tests	Indicates whether the teacher requires students to take mandatory math tests during the year as a form of assessment of academic performance.

Table 10: Results using other definitions for *commuting duration*.

	$\geq$ 1h	30 minutes	minutes
ATT_{OLS}	-0.212***	-0.083**	-.003***
	(0.066)	(0.035)	(0.001)
Observations	2816	2,816	2,816
R^2	0.074	0.073	0.074
F statistic	13.61***	13.38***	13.81***

Note: ***, ** and * represent p<1%, p<5% and p<10% respectively. Specifications included all variables shown in Table *Summary Statistics*.