Commuting and School Absenteeism: Evidence from Brazil

Rubens Lopes, Raul Silveira-Neto
Federal University of Pernambuco (UFPE), Brazil

RESUMO

O absenteísmo escolar dificulta o desenvolvimento de habilidades e comportamentos não-cognitivos importantes para crianças em idade escolar. Assim, entender as causas desse problema é essencial para abordar a questão de forma eficaz. Poucos estudos exploraram o impacto do commuting no absenteísmo escolar, especialmente em países em desenvolvimento. Utilizando um rico banco de dados, este estudo emprega análise de inferência causal para explorar o impacto da distância de deslocamento no absenteísmo de estudantes em Recife, uma cidade densamente povoada e com desafios significativos de transporte, especialmente para os residentes de baixa renda em áreas periféricas. A estratégia de variável instrumental se baseia na política de matrícula do governo municipal, que desempenha um papel na influência da escolha da escola. Os resultados revelam uma associação significativa entre a distância do commuting e o absenteísmo, indicando que o dobro da distância leva a um aumento de 18,8% nas faltas. Esta pesquisa contribui para a literatura ao destacar o papel da frequência escolar na formação de resultados futuros.
Palavras-chave: Absenteísmo escolar, Deslocamento urbano, Resultados escolares
Área de submissão: Economia Regional e Urbana
Códigos JEL: I20, R41

ABSTRACT

School absenteeism hinders the development of important non-cognitive skills and behaviors in students. Thus, understanding the causes of this problem is essential for effectively addressing the issue. Few studies have explored the impact of commuting on absenteeism in school, particularly for developing countries. Using a rich survey dataset, this study employs causal inference analysis to explore the impact of travel distance on student absenteeism on Recife, a densely populated city with significant transportation challenges particularly for low-income residents in peripheral areas. The instrumental variable strategy leverages the city government's enrollment policy, which plays a role in influencing school choice. The findings reveal a significant association between commute distance and absenteeism, indicating that a doubling of the distance leads to an 18.8% increase in absences. This research contributes to the literature by highlighting the role of school attendance in shaping future outcomes.
Keywords: School absenteeism, Commuting, School outcomes
JEL codes: I20, R41

1 Introduction

The literature on education highlights the detrimental effects of school absenteeism on students' future outcomes (Ansari and Pianta, 2019; Gottfried, 2014). While academic performance is often the primary focus of research, grades alone do not fully reflect the importance of the educational experience. In addition to improving academic achievement, school is where students develop many positive non-cognitive skills and behaviors (Cornelissen and Dustmann, 2019). School absenteeism has been linked, for instance, to decreased educational engagement, social engagement, behavioral issues like alienation (Finn, 1989; Johnson, 2005; Newmann, 1981), and health-risk behaviors, including drug use (Hallfors et al., 2002).

If school attendance influences both cognitive and non-cognitive skills development, the determinants of school absence is a relevant research topic in itself. Previous research has found several explanations for school absenteeism: school characteristics (Krueger, 2003), teacher's absence (Banerjee and Duflo, 2007), parental education (Pontili and Kassouf, 2007), student's personality (Lounsbury et al., 2004), and weather (Goodman, 2014).

Nevertheless, there is a crucial aspect related to absenteeism that has not been given much attention: commuting. Although the labor literature has evidence that long commuting distance to work induces absenteeism (van Ommeren and Gutiérrez-i Puigarnau, 2011), and the education literature has found a link between travel time to school and academic performance (Kobus et al., 2015; Tigre et al., 2017), few previous research explored the direct relationship between commuting and absenteeism in school, especially for developing countries. Besides, as discussed on the literature review section, the available studies rarely use methods to infer causal relationships.

If absenteeism is causally linked to time spent traveling to school, it is possible to propose public policies that mitigate the adverse outcomes suffered by children who are frequently absent from school due to long commute times (Asahi, 2014). Moreover, this research also contributes to the ongoing discussion surrounding the optimal location of schools (Araya et al., 2012; De Armas et al., 2022; Pizzolato and Silva, 1997), as school agglomeration policies might negatively affect school accessibility, especially for low-income students (Moreno-Monroy et al., 2018).

The aim of this study is to analyze the impact of travel distance to school on absenteeism by using causal inference analysis in a rich survey dataset containing information on students, their families, and their schools.

Our data comprises schools in Recife, an important Brazilian city. Brazil is a developing country known for its urban areas with a high density of residents and remarkably high commuting costs (Fernández-Maldonado et al., 2014; Ingram and Carroll, 1981; Silveira Neto et al., 2015). Recife is the fourth most densely populated city in Brazil (IBGE, 2010), and is the capital of the state of Pernambuco, located in the Northeast region of the country. Founded in 1537, the city is the oldest state capital in the country, featuring a well-preserved central area. The subject of commuting is specially important for Recife, as between 1992 and 2009, the city saw a gradual increase of over 5 percentage points in the proportion of home-to-work trips that take more than one hour. This increase was concerning because the city has an outdated mass transportation system (Pereira and Schwanen, 2013). The city has a bus-based public transportation system, supported by an old suburban metro system inaugurated in the 1980s. Its roadway infrastructure is also poorly suited to accommodate the recent expansion of individual transportation, which may contribute to the cost of daily commuting (Lopes De Souza et al., 2023). The city of Recife was ranked as the most traffic-congested city in Brazil and the 24th in the world, according to the 2021 Traffic Index conducted by TomTom.[1] The transportation challenges of Recife seem to be worse for those living in the city's peripheral areas, where most of the low-income population resides (Costa et al., 2021).

We used an instrumental variable approach similar to the one used by Kobus et al. (2015) and Tigre et al. (2017). Our instrument is the average distance to the two schools closest to the student's residence. As the city government of Recife induces parents to enroll their children in the closest or second closest school from home[2], the instrument should be a strong predictor of commute time and, at the same time, should not affect absenteeism except for its effect on commuting. To ensure the validity of the instrument, we employ additional control variables and perform verification tests that are discussed in Section 4.

Our results show that commute time has a significant impact on absenteeism. We found that, if commute distance doubles, absences increases by 18.8%. We ran several robustness tests and the results remained significant with little variance. This finding is important because it suggests that public policies that reduce commute time can have a positive impact on school attendance.

This paper is organized as follows: Section 2 presents the literature review. Section 3 presents a theoretical framework. Section 4 describes the data and the empirical strategy. Section 5 presents the results, and section 6 concludes.

2 Daily Commuting and School Attendance

2.1 Absenteeism and its consequences

The negative relationship between student absenteeism and academic performance has been established in the literature for many years. Studies conducted by Summers and Wolfe (1977) in Philadelphia and Monk and Ibrahim (1984) in upstate New York found that students who had a higher number of unexcused absences had a lower performance on tests. More recently, Gottfried (2009) studying data of Philapelphia students highlighted that, when absences are not excused, the negative impact on academic achievement is much greater compared to when absences are due to legitimate reasons such as health issues. The author used linear regression models with fixed-effects to analyse the data. The regression analysis studies conducted by Gottfried (2011) in Philapelphia, Gershenson et al. (2015) in North Carolina, and Morrissey et al. (2014) in Florida found similar results, showing that there is a significant negative relationship between absenteeism and test scores. On a comparable study, Gottfried and Kirksey (2017) analyzed elementary school data in California with time series regressions and reported that the most critical period is the 30-day window leading up to a test. Additionaly, Gottfried (2014) studied data for American kindergarten students using linear regression and found that chronic absenteeism also hurts reading outcomes. Ansari and Pianta (2019) revealed that absenteeism in the early years of school can have long-term negative effects on academic outcomes. They analysed a sample of American children who were followed from birth through high school using correlation analysis. Finally, Klein et al. (2022) found that all forms of absenteeism, regardless of the reason, are negatively associated with academic

[1]Source (In Portuguese). Accessed on June 14, 2023.

[2]Source (In Portuguese). Accessed on June 16, 2023.

achievement both during compulsory and post-compulsory schooling in Scotland. The authors used OLS linear regression with fixed-effects and first-difference models.

A number of studies have also explored the relationship between absenteeism and dropping out rates. Rumberger (1995) found that moderate to high levels of absenteeism are a significant predictor of dropout for middle school American students. Lehr et al. (2004) discovered that Minnesota students who are at risk of dropping out can be identified as early as third grade based on attendance patterns and other factors.

While academic performance is often considered the primary goal of school attendance, it is important to recognize that schools play a crucial role in children's overall development beyond academics (Gottfried and Ansari, 2021). The school environment itself is significant, as it is where children learn a range of positive non-cognitive skills and desired social behaviors. Ryan and Patrick (2001), for example, surveyed eighth-graders from two midwestern school districts regarding changes in motivation and engagement when they moved from seventh to eighth grade. They found that when students perceived their teacher as supportive, they exhibited more self-regulated learning and less disruptive behavior. Also, students' perception of being encouraged to interact with others in the classroom and to share their ideas was correlated with all indicators of motivation and engagement. West et al. (2016) found a positive correlation between school attendance and the development of qualities such as conscientiousness, self-control, grit, and growth mindset in students from fourth to eighth grade in Boston city. Cornelissen and Dustmann (2019) similarly discovered a positive effect of school attendance on student-teacher relationships, academic interest, and reduced disruptive behavior for 11-year-old students in the United Kingdom.

Inversely, several studies have shown a correlation between school absenteeism and decreased educational engagement, reduced social engagement, and behavioral problems such as alienation (Finn, 1989; Johnson, 2005; Newmann, 1981). Hallfors et al. (2002) conducted a meta-analysis with American school data. They investigated the predictors of drug use among students and found that the odds-ratio for absenteeism among seventh and eighth graders was twice as large as GPA score in predicting drug use. Eaton et al. (2008) used logistic regression analysis on data from 8 American states and found that students who were absent without permission were twice as likely to engage in health-risk behaviors, such as tobacco use and drug use. Gottfried (2014) conducted a large-scale regression analysis on a sample of 10,740 kindergarten students in the US and found that chronic absenteeism was associated with a decline in educational engagement (as measured by approaches to learning and eagerness to learn) and a decline in social engagement (as measured by an increase in internalizing behaviors). Fuhs et al. (2018) studied the relationship between school attendance, readiness skills, and executive function (i.e., working memory, inhibitory control, and the ability to focus attention) for preschool children in an American Midwestern city using correlation analysis. They found that children who were chronic absentees made significantly fewer gains in their executive function skills from fall to spring of their preschool year. Ansari and Pianta (2019) investigated the lasting effects of absenteeism during the earliest stages of a child's education. Through correlation analysis on a American national dataset, they discovered that absenteeism in the first ten years of a child's education was associated with lower academic performance and less favorable social-behavioral outcomes at the age of 15. These problematic outcomes include sexual practices and risky behaviors (e.g., drinking alcohol, smoking, getting into fights). Additionally, they were more likely to continue missing school later in their educational journey. Gottfried and Ansari (2021) conducted regression analysis on a large dataset of 14,370 children in the US and found that chronically absent children demonstrated roughly 1–1.5 fewer months of gains in working memory between kindergarten and third grade compared to their peers. They also reported greater school-related stress, higher levels of victimization and social anxiety, lower levels of motivation and grit, and lower levels of school belonging.

The reviewed literature suggests that school attendance is a significant factor for both cognitive and non-cognitive development in students. However, most studied apply naive methods that are not able to identify causal relationships. Thus, it is worth highlighting two studies that have attempted to address this issue. In 2015, Kobus et al. conducted a study on the impact of commute time on university presence and academic achievement using data from students of VU University in Amsterdam, the Netherlands. They used an instrumental variable approach, while controling for a series of socio-economic characteristics. Their instrument is the mean public transport travel time from the municipality of origin to the closest two Dutch university cities. They argue that this variable predicts the commuting time to the VU University, while having no impact on the potential students' decision to select the VU University – conditional on control variables. The municipality of origin should not directly affect university presence or academic achievement, as most parents have chosen a residence municipality ignoring the effect that this may have on their children's travel time to the university. However, the distance from the municipality of origin does influence commute times, particularly for students living with their parents or those who prefer to live in municipalities with social ties. They found a negative relationship between commute time and grades. Tigre et al. (2017), following a similar strategy, conducted a study using Brazilian data that investigated the causal impact of commuting duration on student performance. Their instrument is the average distance of current residence the the closes two schools. They argue that, as the city government asks parents to enroll their children in the closest school, and as they control for school and parenting quality, the effect of the distance from home to school should only affect achievement through commuting costs. They found that longer commutes had a negative impact on academic achievement. Neither studies, however, examine the connection between commuting and absenteeism, leaving a gap in the understanding of whether the link between commuting and grades is driven by absences' impact on academic performance or the influence of commuting on productivity due to fatigue and time loss.

2.2 Commuting and Absenteeism

If school attendance is important for students development and achievement, it is crucial to better comprehend the issue of absenteeism. One of the major barriers to school attendance is limited accessibility, with several studies confirming that long commute times, long distances, and inadequate transportation are significant predictors of absences.

Developed countries are not exempt from acessibility problems affecting school attendance. In 1982, Dexter published a study that surveyed 155 high school students of the city of Portland regarding absenteeism. The selected sample of students included only freshmen and sophomores considered to have excesive absenteeism (more than 10 absences). The author found that 34.3% of those who used buses to commute to school reported "Transportation" as a reason for their absences. More recently, Gottfried (2017) conducted a study in the US using linear regression on a national large-scale dataset and found that kindergarten students who took a school bus were less likely to be absent than those who used other modes of transportation. Similarly, García and Weiss (2018) used regression analysis and observed that a lack of adequate transportation was associated with a higher likelihood of absenteeism, especially chronic absenteeism, among 8th graders in the US. Stein and Grigg (2019) analyzed the attendance records of high school students in Baltimore and Maryland who commute to school. They used first-difference estimation method to address the issue of endogeneity and discovered that students who have to travel longer or face more complex commutes miss more days of school.

This accessibility challenge appears to be even more pronounced in developing countries. In a study conducted in Mozambique, Handa (2002) used probit regression to determine that the distance to a primary school significantly affects children's enrollment. The study found that reducing the travel time to the nearest school in 30 minutes could increase enrollment rates by 20 and 17 percentage points for both boys and girls, respectively. Similarly, Kazeem et al. (2010) used logistic regression on Nigerian data and found that living 20 or more minutes from the nearest school reduces the odds of attendance by 27 percent for primary schools and 52 percent for secondary schools. Duze (2011) used descriptive statistics to analyze the impact of the distance travelled by students on school attendance in Nigeria's states of Anambra, Enugu, and Ebonyi. The author found that 65.23% of primary school students and 76.09% of secondary school students travelled more than one kilometer to school, which negatively impacted their attendance. Thembo (2011), in Western Uganda, observed a negative correlation between the distance travelled by students from their homes to school and the daily school attendance.

Most studies in this literature do not estimate causality, but two studies that use causal analysis are worth highlighting. Vuri (2010), using multivariate probit with instrumental variable, found that an additional 10 minutes of travel time to school reduced the probability of attendance by 0.2 percentage points in rural Ghana. Burde and Linden (2013) conducted a randomized controlled trial in rural northwestern Afghanistan to assess the impact of village-based schools on primary school-age children. They randomly assigned village-based schools to 13 villages among a sample of 31 villages and found that reducing the distance to school increased enrollment rates by 35% for boys and 52% for girls, virtually eliminating the gender disparity in the treatment villages. This result is particularly noteworthy as cultural norms in Afghanistan do not permit girls to travel long distances alone.

Finally, there's a notable study conducted in Brazil by Moreno-Monroy et al. (2018). In 2016, the government of the state of São Paulo intended to close selected secondary schools as part of a plan to reduce the budgetary deficit. This policy was expected to impact over 300,000 students, with many being relocated to schools far from their homes. After public opposition, the decision was not implemented. Moreno-Monroy et al. (2018) conducted a study in response to this event, using simulation to estimate the effect of centralizing public secondary schools in the Metropolitan Region of São Paulo. The authors argued that given the significant spatial disparities in public transport and schooling provision, it is crucial to consider whether providing public transport subsidies to students can offset the limited availability of public schools in certain areas. The estimations indicated that the implementation of such a policy would result in longer commuting times for students in areas with low accessibility, leading to adverse consequences in terms of attendance rates, increased likelihood of dropout, and academic performance. As young individuals residing in areas with limited accessibility would be the most affected, this policy would exacerbate issues related to inadequate local schooling, restricted public transport access, and intense competition for enrollment in high-quality schools.

Unfortunately, many studies either do not examine the direct relationship between commute distance and school attendance, or rely on simple correlation analysis without establishing causality. There is a notable scarcity of studies employing causal analysis in the context of developing countries, and to the best of our knowledge, none have been conducted for urbanized areas or specifically for Brazil. The purpose of our study is to bridge this gap by addressing these research deficiencies.

3 Theoretical Framework

In this section, we present a simple theoretical model that describes how absences can be related to commute time. We follow the modelling presented by Kobus et al. (2015) to a certain extent.

Suppose that a student maximize utility from school gains, P, and leisure time at home, H_l. School gains is positively related to time spent at school, which is measure by days present at school, D_u, and daily number of hours present, H_u. School gains is also positively related to hours studying at home, given by H_h. The student spends time commuting to school, given by t; and time is the only cost associated with traveling to school. Therefore, the student maximizes the function $U(P, H_l) = U(D_u, H_h, H_l)$, and his utility is subject to a time constraint $D_u(H_u + t) + H_h + H_l = M$, where M denotes the student's time budget.

We assume a Cobb-Douglas utility function given by $P = D_u^{\alpha} H_u^{\beta} H_h^{\gamma} \rightarrow PH_l^{\delta} = D_u^{\alpha} H_u^{\beta} H_h^{\gamma} H_l^{\delta}$, in which $\alpha, \beta, \gamma, \delta > 0$. It is easy to obtain that $\alpha > \beta$ is a necessary condition for the maximization problem. Therefore, the number of days elasticity of school gains is larger than the number of hours elasticity of school gains. This implies that hours of study face stronger diminishing returns than days going to school, which makes sense as students should get tired after several hours of study.

It can be shown that:

$D_u^* = \frac{M(\alpha-\beta)}{t(\alpha+\gamma+\delta)}$.

Recalling that $\alpha > \beta$, it follows that $\frac{\sigma D_u^*}{\sigma t} < 0$. This implies that, for any student, longer commute times is associated with fewer days present at school.

This simple framework illustrated our main hypothesis. We expect to find a positive relationship between commute times and school absenteeism.

4 Empirical Strategy

4.1 Specification

In this section, we propose an empirical model to measure the impact of accessibility on school absenteeism. We consider the following model:

$$Absences = f(Commuting + A + T + E) \tag{1}$$

In which the dependent variable, *Absences*, is the total of absences of student i, who is part of the j class of the k school. The main regressors is the commute distance of the student i to the school k, given by *Commuting*. This variable was assessed using the Google API, which provides the shortest route by foot[3] taken by the student when traveling to school.

[3]77% of students in the dataset commute to school on foot.

A is a vector of twenty control variables related to the students and the student's family. In T, we included three variables specifically related to the student's class. E is a vector of school-related regressors. All variables mentioned are listed in Table 1.

In this study, we employ a Poisson model estimation for our main regression analysis. The choice of the Poisson model is motivated by the nature of the outcome variable, which represents count data in the form of the number of absences. Poisson regression is specifically designed to handle count data and is well-suited for situations where the outcome variable exhibits a skewed distribution with non-negative integer values. By using the Poisson model, we can account for the inherent characteristics of count data and appropriately model the relationship between commuting and school absences.

We will also estimate this model using OLS. For this case, we will use the following specification:

$$Absences_{ijk} = \alpha + \beta\, Commuting_i + \Gamma A_i + \Psi T_j + \Lambda E_k + \varepsilon_{ijk} \tag{2}$$

In which i is the index of students, j the index of schools, and k the index of classes.

4.2 Endogeneity

The standard Poisson and OLS estimations for the model above may suffer from problems related to endogeneity, as omitted variable bias. To ensure we are capturing a causal effect, we employ a similar approach to the one used by Kobus et al. (2015) and Tigre et al. (2017). We utilize the average distance between students and the two nearest schools as an instrumental variable for the commuting distance.

The city government of Recife instructs parents to enroll their children in the school closest to their home. As a result, the average distance between the two closest schools to the residence should be a good predictor of the actual commuting distance of the student. At the same time, this variable should not affect the number of absences of students throughout the year, except for its effect on commuting. If these assumptions hold, we have a valid instrument.

However, two objections can be raised against this instrument. First, parents who are more concerned about their children's education may choose to live closer to better schools and also make an effort to ensure that their children do not miss classes. In this case, parental concern might be related both to the instrument and the endogenous variable, rendering the instrument invalid. Second, there may be heterogeneities regarding the family's socioeconomic status that affect both the locational choice of housing and school absenteeism. An example is the Bolsa Família, a federal government benefit that provides income transfers to low-income families, conditioned on the children regularly attending school. In this case, the family's socioeconomic status can determine the proximity to public schools while also influencing school absenteeism, precluding our causal identification.

To address the first objection, we include several control variables that control for parents' concern:

1. Parents supervise homework?
2. How was the child's school chosen?
3. What was the strategy adopted to get a quality school place?

Table 1: Control variables

Category	Variable
Students	Age
Students	Race
Students	Gender
Students	Has already failed
Students	Attends particular extra classes
Students	Household has a computer with internet access
Students	How often missed classes because were bullied
Students	Intends to go to college
Students	Is taken to school every day
Students	Math test score
Students	Means of transport used
Students	Parents supervise homework
Students	Portuguese test score
Students	Uses the internet for studying
Students	Works
Family	If considers neighborhood to be violent
Family	If parent works
Family	Mother is the parent
Family	parent's education
Family	Satisfied with the neighborhood
Class	Is there a high rate of absenteeism among teachers in the class evaluated by the Fundaj Survey?
Class	Teacher's experience
Class	Was this class interrupted this year?
School	If the school offers extra-class activities
School	If the school offers integral classes
School	If the school provides a math book
School	Is there a dropout program?
School	Is there a failure rate reduction program?
School	What the school does to prevent students from skipping classes

4. Would you like the child to study at a different public school?

5. Why do you live in this residence?

The first four variables can identify those parents most interested in the quality of the student's education. The last variable also fulfills this function, since one of the answer options is "Close to school". In addition, we perform three robustness tests in which we use specific subsamples that remove from the sample childrens' parents that showed concern about school quality and distance.

To address the second objection, we included several variables that control for socioeconomic status: if the parent works, parent's education, and if Household has a computer with internet access. We also show that our instrumental variable is balanced regarding to socioeconomic characteristics of the family. We first correlate each decile of the IV with the following variables: guardian's education, log(income), a binary indicator for Bolsa Família, and a binary indicator for formal employment. The correlations are shown in Table 2. We also run a regression of each of the variables mentioned above on the IV, and the results are shown in Table 3[4]. The results show that the IV is not correlated with the socioeconomic characteristics of the family.

Table 2: Correlation between IV and socioeconomic variables

	Parent's education	Log(Income)	Bolsa Família	Formal Employment
Percentile 10	0.0807	-0.0457	0.0588	0.1220
Percentile 20	0.0755	-0.0701	-0.0117	-0.0687
Percentile 30	-0.0388	-0.1137	0.0118	0.0196
Percentile 40	0.0033	-0.0429	-0.0558	0.0672
Percentile 50	-0.0554	0.0617	-0.0570	0.0977
Percentile 60	0.0239	-0.1224	0.0400	-0.0019
Percentile 70	-0.1275	0.0858	-0.0169	-0.1447
Percentile 80	-0.0438	0.0234	0.0285	-0.1794
Percentile 90	-0.1300	-0.0934	0.1174	-0.0635
Percentile 100	0.0456	-0.1193	0.0438	0.0669

[4] For this table specifically, we only considered significant estimates with a p-value<0.05.

Table 3: Regression of socioeconomic variables on IV

	(1) Log(IV)	(2) Log(IV)	(3) Log(IV)	(4) Log(IV)
Parent's education	0.00423			
	(0.00365)			
Log(Income)		-0.00576		
		(0.00380)		
Bolsa Família			-0.0248	
			(0.0270)	
Formal Employment				0.0596
				(0.0372)
N	2005	2008	2008	2008
r2	0.000681	0.000973	0.000421	0.00149

* p<0.05, ** p<0.01, *** p<0.001. Robust Standard errors in parentheses.

Once we control for parents' concern about the student education and socioeconomic conditions of the family, the remaning effect of the instrumental variable on school absences should be only through commuting.

4.3 Data

This study will use survey data provided by The Joaquim Nabuco Institute for Social Research (Fundaj), a local agency linked to the Brazilian Ministry of Education and Culture. The survey was conducted in 2017 and again in 2018 and it consisted of standardized Math and Portuguese tests, along with a series of socioeconomic questions, both objective and subjective. Interviewers collected a wide range of information on students and their school life through four questionnaires: one for students, one for the primary adult responsible for the child's academic life, one for the teachers and one for the school principal (Raposo et al., 2019). The questionnaire aimed at public school students ages 11 to 13. The survey was conducted in Recife, the wealthiest city in the North-Northeast region of Brazil, a relevant developing country.

In this study, we will focus our analysis on the survey data for the year 2018. The dataset is comprised of 2008 students and guardians, and 87 principals and schools.

5 Results

5.1 Main results

Table 4: Poisson regression analysis

	Poisson	IV Poisson GMM	IV Poisson cf
	(1)	(2)	(3)
	Absences	Absences	Absences
Absences			
Log(API distance)	0.0698*	0.188**	0.174**
	(0.0396)	(0.0943)	(0.0847)
Observations	1786	1786	1786
χ^2	35.51		

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors in parentheses.

In this section, we present the results and discussion of our analysis, focusing on the impact of travel time to school on absenteeism. We employed various regression models to explore this relationship, accounting for potential endogeneity concerns.

Estimation results are presented in Table 4. Column 1, which utilized a Poisson model, revealed a significant positive coefficient of 0.0698 for the logarithm of distance. This suggests that an increase in travel distance of 10% is associated with an approximately 7% more absences among students. However, given the possibility of endogeneity, we proceeded to estimate instrumental variable models. At this point we introduce our instrumental variable, the average distance to the two closest schools from the student's residence. In Column 2 we present our main specification and estimation model: an IV Poisson using the generalized method of moments (GMM). The GMM allows for a flexible functional form and does not require explicit specification of the control equation. It can also handle both linear and non-linear relationships between the endogenous variables and the instruments. The coefficient for Log(API distance) increased to 0.188 and remained statistically significant at the 5% level. This strengthened effect reinforces the substantial impact of longer commute times on absences. The result indicates that doubling the travel distance to school could lead to a 18.8% increase in absences.[5] To further validate our findings, we employed an IV Poisson model with a control function in Column 3. The results were consistent with the model in Column 2, showing a coefficient of 0.174, which was also statistically significant at the 5% level. This result provides additional support for the robustness of our findings, as the relationship between travel distance and absences holds even when using a different estimation technique.

Our results demonstrate a significant positive association between travel distance to school and absenteeism. Longer commute times are linked to a higher number of absences among students, indicating that reducing travel distance can potentially improve school attendance rates. These findings hold important implications for policymakers and educators, particularly in densely populated urban areas facing transportation challenges, similar to Recife.

[5]The average distance to school is 2,031.368 meters and the average number of absences is 4.049 days.

Effective public policies should prioritize initiatives aimed at reducing commute time. Improving transportation infrastructure, optimizing school locations, and reinforcing school choice mechanisms that consider proximity to students' residences are potential strategies to consider. By addressing the challenges associated with long travel times, policymakers can mitigate the adverse effects on students' attendance and educational outcomes.

5.2 Robustness tests

5.2.1 Alternative methods

Alternative methods can offer valuable insights and provide a comprehensive evaluation of the research findings. The main method employed in this study is IV Poisson, which it effective in addressing endogeneity issues and accommodating count data. However, to ensure the robustness of the results, it is important to consider alternative methods. Results are presented in Table 5.

Ordinary Least Squares is the default model for regression analysis, making it a suitable benchmark against which to compare the IV Poisson results (column 1). 2SLS (Two-Stage Least Squares) extends OLS by incorporating instrumental variables to address endogeneity (column 2). Tobit ML (Maximum Likelihood) is specifically designed for censored dependent variables, making it relevant when dealing with truncated data (column 3). Lastly, Tobit 2-step combines the strengths of Tobit models and instrumental variable techniques to account for potential endogeneity and censoring simultaneously (column 4). Notably, Tobit models assume a continuous dependent variable, which may not be suitable for count data, as is the case in our study. However, the obtained results are still useful for comparative analysis. Additionally, it is worth mentioning that a log transformation was applied to the explained variables in Table 5 to ensure comparability with the Poisson estimations.

The results show that the main findings (Table 4) are robust to alternative methods.

Table 5: Alternative regression methods

	OLS (1) Log(Absences)	2SLS (2) Log(Absences)	Tobit ML (3) Log(Absences)	Tobit 2-step (4) Log(Absences)
main				
Log(API distance)	0.0624***	0.125**	0.225***	0.225***
	(0.0215)	(0.0520)	(0.0773)	(0.0756)
Observations	1786	1786	1786	1786
R^2	0.0307	0.0262		
First stage F		0.0160		
Anderson-Rubin p-value		394.9		

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors in parentheses.

5.2.2 Alternative regressors

Table 6 displays the results of the robustness tests conducted using alternative regressors in the IV Poisson framework. The four columns represent different specifications of the IV Poisson model (1)-(4). Column 1 shows our prefered especification: commuting is measured as the walking distance from the student's home to school obtained via Google API using IV Poisson with control function (also shown in column 2 of Table 4). In a alternative specification, we used the number of walking minutes, also obtained via Google API (column 2). In Column 3, we used the number of minutes to commute from home to school as reported by the student's guardian. Finally, in Column 4, we used the Euclidean distance between the school and the student's home.

The estimations show that the main findings are robust to alternative specifications of the commuting variable, as shown by the similar magnitudes and statistical significance in columns 2 and 4. The specification that utilizes the guardian's self-reported commuting time (column 2) is the one that diverges the most, likely due to the subjective nature of self-reported data, which makes it more susceptible to measurement errors. Nonetheless, the result remains positive and significant.

Table 6: Alternative regressor variables

	IV Poisson GMM (1) Absences	IV Poisson GMM (2) Absences	IV Poisson GMM (3) Absences	IV Poisson GMM (4) Absences
Absences				
Log(API distance)	0.188**			
	(0.0943)			
Log(API time)		0.198**		
		(0.0997)		
Log(guardian time)			0.410*	
			(0.212)	
Log(euclidian distance)				0.167*
				(0.0894)
Observations	1786	1786	1756	1737

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors in parentheses.

5.2.3 Alternative samples

Even though we included several control variables and estimated the regression via two-staged least squared to obtain precise and causal estimations, it is also important to perform robustness tests to to assess the reliability of our findings. Thus, after the obtaining the main results, we repeated the regression analysis using specific subsamples (Table 7).

First, we will redo the analysis excluding those who answered that they chose the school based on quality (column 1). If the parent is more concerned about the child's education than the average parent, this preoccupation will affect both the travel time to school and absences. By removing those concerned parents and obtaining similar results, we can make a stronger case that we are capturing a causal link between commute time and school absenteeism.

Second, we also filter our sample to keep only those that affirmed that the school choice was either "referral from previous school" or "where there was vacancy" (column 2).

Third, we will remove those who affirmed that lived where they live for the reason of being near school. We strongly believe that this reason is confounding variables that may influence students' school attendance. By removing it, we can further confirm that our findings were not guided by other reasons other than commute time (column 3).

The results show that the main findings are robust to alternative samples. The coefficients remain positive and significant across all models.

Table 7: Alternative samples

	IV Poisson GMM	IV Poisson GMM	IV Poisson GMM
	(1)	(2)	(3)
	Absences	Absences	Absences
Absences			
Log(API distance)	0.226**	0.513***	0.182*
	(0.104)	(0.199)	(0.0945)
Observations	1550	664	1739

* p<0.10, ** p<0.05, *** p<0.01. Robust standard errors in parentheses.

6 Conclusions

Overall, the regression results provide evidence of a significant positive relationship between travel time to school and absenteeism. The coefficients consistently suggest that longer commute times are associated with a higher number of absences among students. These findings support the hypothesis that commute time plays a crucial role in students' attendance and highlight the importance of reducing travel time to promote better school attendance rates.

It is worth noting that the results are based on a dataset comprising schools in Recife, a densely populated Brazilian city facing transportation challenges. Therefore, the findings may have particular relevance for urban areas with similar characteristics. The robustness of the results is further supported by the statistical significance and consistency across different models.

The regression analysis provides important insights for policymakers and educators. The results suggest that public policies aimed at reducing commute time to school may have a positive impact on addressing absenteeism issues. By improving transportation infrastructure, optimizing school locations, and reinforcing school choice mechanisms that consider proximity to students' residences, it may be possible to mitigate the adverse effects of long travel times on students' attendance.

However, it is important to acknowledge the limitations of the study. The analysis is based on observational data, and although instrumental variable techniques help address endogeneity concerns, there may still be unobserved factors that influence both travel time and absences.

In conclusion, the results indicate a significant relationship between travel time to school and absenteeism in Recife, Brazil. The findings underscore the need for targeted interventions and policies to reduce commute time and improve school attendance. Further research exploring similar relationships in different settings would contribute to a more comprehensive understanding of the impact of commute time on students' educational experiences.

References

Ansari, A. and Pianta, R. C. (2019). School absenteeism in the first decade of education and outcomes in adolescence. *Journal of School Psychology*, 76:48–61.

Araya, F., Dell, R., Donoso, P., Marianov, V., Martínez, F., and Weintraub, A. (2012). Optimizing location and size of rural schools in Chile: F. Araya et al. *International Transactions in Operational Research*, 19(5):695–710.

Asahi, K. (2014). The Impact of Better School Accessibility on Student Outcomes. *SERC DISCUSSION PAPER*.

Banerjee, A. V. and Duflo, E. (2007). The Economic Lives of the Poor. *Journal of Economic Perspectives*, 21(1):141–167.

Burde, D. and Linden, L. L. (2013). Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools. *American Economic Journal: Applied Economics*, 5(3):27–40.

Cornelissen, T. and Dustmann, C. (2019). Early School Exposure, Test Scores, and Noncognitive Outcomes. *American Economic Journal: Economic Policy*, 11(2):35–63.

Costa, C., Ha, J., and Lee, S. (2021). Spatial disparity of income-weighted accessibility in Brazilian Cities: Application of a Google Maps API. *Journal of Transport Geography*, 90:102905.

De Armas, J., Ramalhinho, H., and Reynal-Querol, M. (2022). Improving the accessibility to public schools in urban areas of developing countries through a location model and an analytical framework. *PLOS ONE*, 17(1):e0262520.

Dexter, R. L. (1982). An investigation of the causes of absenteeism in an inner-city school.

Duze, C. (2011). Average Distance Travelled To School by Primary and Secondary School Students in Nigeria and Its Effect on Attendance. *African Research Review*, 4(4).

Eaton, D. K., Brener, N., and Kann, L. K. (2008). Associations of Health Risk Behaviors With School Absenteeism. Does Having Permission for the Absence Make a Difference? *Journal of School Health*, 78(4):223–229.

Fernández-Maldonado, A. M., Romein, A., Verkoren, O., and Parente Paula Pessoa, R. (2014). Polycentric Structures in Latin American Metropolitan Areas: Identifying Employment Sub-centres. *Regional Studies*, 48(12):1954–1971.

Finn, J. D. (1989). Withdrawing from School. *Review of Educational Research*, page 26.

Fuhs, M. W., Nesbitt, K. T., and Jackson, H. (2018). Chronic Absenteeism and Preschool Children's Executive Functioning Skills Development. *Journal of Education for Students Placed at Risk (JESPAR)*, 23(1-2):39–52.

García, E. and Weiss, E. (2018). Student Absenteeism: Who Misses School and How Missing School Matters for Performance. *Economic Policy Institute*.

Gershenson, S., Jacknowitz, A., and Brannegan, A. (2015). Are Student Absences Worth the Worry in U.S. Primary Schools? *Education Finance and Policy*, 12(2):137–165.

Goodman, J. (2014). Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time. Technical Report w20221, National Bureau of Economic Research, Cambridge, MA.

Gottfried, M. A. (2009). Excused Versus Unexcused: How Student Absences in Elementary School Affect Academic Achievement. *Educational Evaluation and Policy Analysis*, 31(4):392–415.

Gottfried, M. A. (2011). The Detrimental Effects of Missing School: Evidence from Urban Siblings. *American Journal of Education*, 117(2):147–182.

Gottfried, M. A. (2014). Chronic Absenteeism and Its Effects on Students' Academic and Socioemotional Outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, 19(2):53–75.

Gottfried, M. A. (2017). Linking Getting to School With Going to School. *Educational Evaluation and Policy Analysis*, 39(4):571–592.

Gottfried, M. A. and Ansari, A. (2021). Detailing New Dangers: Linking Kindergarten Chronic Absenteeism to Long-Term Declines in Executive Functioning. *The Elementary School Journal*, 121(3):484–503.

Gottfried, M. A. and Kirksey, J. J. (2017). "When" Students Miss School: The Role of Timing of Absenteeism on Students' Test Performance. *Educational Researcher*, 46(3):119–130.

Hallfors, D., Vevea, J. L., Iritani, B., Cho, H., Khatapoush, S., and Saxe, L. (2002). Truancy, Grade Point Average, and Sexual Activity: A Meta-Analysis of Risk Indicators for Youth Substance Use. *Journal of School Health*, 72(5):205–211.

Handa, S. (2002). Raising primary school enrolment in developing countries The relative importance of supply and demand. *Journal of Development Economics*.

IBGE (2010). Indicadores sociais municipais: uma análise dos resultados do universo do censo demográfico 2010.

Ingram, G. K. and Carroll, A. (1981). The spatial structure of Latin American cities. *Journal of Urban Economics*, 9(2):257–273.

Johnson, G. M. (2005). Student Alienation, Academic Achievement, and WebCT Use. *Educational Technology & Society*, page 11.

Kazeem, A., Jensen, L., and Stokes, C. (2010). School Attendance in Nigeria: Understanding the Impact and Intersection of Gender, Urban-Rural Residence, and Socioeconomic Status. *Comparative Education Review*, 54(2):295–319.

Klein, M., Sosu, E. M., and Dare, S. (2022). School Absenteeism and Academic Achievement: Does the Reason for Absence Matter? *AERA Open*, 8:233285842110711.

Kobus, M. B., Van Ommeren, J. N., and Rietveld, P. (2015). Student commute time, university presence and academic achievement. *Regional Science and Urban Economics*, 52:129–140.

Krueger, A. B. (2003). Economic Considerations and Class Size. *The Economic Journal*, 113(485):F34–F63.

Lehr, C. A., Sinclair, M. F., and Christenson, S. L. (2004). Addressing Student Engagement and Truancy Prevention During the Elementary School Years: A Replication Study of the Check & Connect Model. *Journal of Education for Students Placed at Risk (JESPAR)*, 9(3):279–301.

Lopes De Souza, A., Batista Duarte, L., and Da Mota Silveira Neto, R. (2023). Acessibilidade espacial ao emprego como determinante da participação, empregabilidade e informalidade: Um estudo empírico para a Região Metropolitana do Recife. *Revista Brasileira de Estudos Regionais e Urbanos*, 16(2):184–213.

Lounsbury, J. W., Steel, R. P., Loveland, J. M., and Gibson, L. W. (2004). An Investigation of Personality Traits in Relation to Adolescent School Absenteeism. *Journal of Youth and Adolescence*, 33(5):457–466.

Monk, D. H. and Ibrahim, M. A. (1984). Patterns of Absence and Pupil Achievement. *American Educational Research Journal*, 21(2):295–310.

Moreno-Monroy, A. I., Lovelace, R., and Ramos, F. R. (2018). Public transport and school location impacts on educational inequalities: Insights from São Paulo. *Journal of Transport Geography*, 67:110–118.

Morrissey, T. W., Hutchison, L., and Winsler, A. (2014). Family income, school attendance, and academic achievement in elementary school. *Developmental Psychology*, 50(3):741–753.

Newmann, F. (1981). Reducing Student Alienation in High Schools: Implications of Theory. *Harvard Educational Review*, 51(4):546–564.

Pereira, R. H. M. and Schwanen, T. (2013). Tempo de deslocamento casa-trabalho no Brasil (1992-2009): Diferenças entre regiões metropolitanas, níveis de renda e sexo.

Pizzolato, N. and Silva, H. F. (1997). The Location of Public Schools: Evaluation of Practical Experiences. *International Transactions in Operational Research*, 4(1):13–22.

Pontili, R. M. and Kassouf, A. L. (2007). Fatores que afetam a freqüência e o atraso escolar, nos meios urbano e rural, de São Paulo e Pernambuco. *Revista de Economia e Sociologia Rural*, 45(1):27–47.

Raposo, I. P. d. A., Menezes, T. A. d., Lima, R. C. d. A., and da Nóbrega, R. Z. (2019). Peer effects and scholastic achievement: spatial models estimates using the student friendship network at the classroom level. *Economia Aplicada*, 23(1):5–24.

Rumberger, R. W. (1995). Dropping Out of Middle School: A Multilevel Analysis of Students and Schools. *American Educational Research Journal*, 32(3):583–625.

Ryan, A. M. and Patrick, H. (2001). The Classroom Social Environment and Changes in Adolescents' Motivation and Engagement During Middle School. *American Educational Research Journal*, 38(2):437–460.

Silveira Neto, R., Duarte, G., and Páez, A. (2015). Gender and commuting time in São Paulo Metropolitan Region. *Urban Studies*, 52(2):298–313.

Stein, M. L. and Grigg, J. A. (2019). Missing Bus, Missing School: Establishing the Relationship Between Public Transit Use and Student Absenteeism. *American Educational Research Journal*, 56(5):1834–1860.

Summers, A. A. and Wolfe, B. L. (1977). Do Schools Make a Difference? *The American Economic Review*, 67(4):639–652.

Thembo, K. N. (2011). School accessibility and students' attendance in secondary Schools in Kasese district, Uganda. *Kyambogo University*.

Tigre, R., Sampaio, B., and Menezes, T. (2017). THE IMPACT OF COMMUTING TIME ON YOUTH'S SCHOOL PERFORMANCE. *Journal of Regional Science*, 57(1):28–47.

van Ommeren, J. N. and Gutiérrez-i Puigarnau, E. (2011). Are workers with a long commute less productive? An empirical analysis of absenteeism. *Regional Science and Urban Economics*, 41(1):1–8.

Vuri, D. (2010). The Effect of Availability of School and Distance to School on Children's Time Allocation in Ghana: The Effect of Availability of School and Distance to School. *LABOUR*, 24:46–75.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F. O., and Gabrieli, J. D. E. (2016). Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling. *Educational Evaluation and Policy Analysis*, 38(1):148–170.