Subway Access and School Participation: Evidence from Brazil*

Edilberto Almeida† Lucas Emanuel‡ Rafael H. M. Pereira § Giuseppe Trevisan¶

Abstract

Despite educational policies targeted to increase school participation, transportation costs can still pose a significant barrier to student retention. Using a subway expansion policy in a Brazilian city with pressing school participation issues, we evaluate how a significant expansion in public transit infrastructure can affect dropout rates in schools that are more exposed to this intervention. To establish causality, we leverage the exogenous variation in the staggered opening of subway stations and the proximity of schools to these new facilities using a difference-in-differences research design. We find robust evidence that schools located near subway stations retain significantly more students at both elementary and high school levels. This effect is entirely driven by public schools, which experience an average reduction in dropout rates by 24%. While the increase in retention rates does not affect student performance or relate to school and teacher quality, we find that students' background play a crucial role in moderating the policy's impact. Our findings underscore that transportation policies can have significant spillover effects on school participation, contributing to better educational outcomes, even within sprawling urban environments.

Keywords: Subway network, dropout rate, difference-in-differences.
JEL codes: I25, O18, R41.

Resumo

Embora políticas educacionais específicas tenham sido implementadas para diminuir a evasão escolar, os custos de transporte ainda podem representar uma barreira significativa para a retenção de alunos. Utilizando uma política de expansão da rede de metrô em uma cidade brasileira que apresenta problemas críticos de retenção escolar, este artigo avalia como uma expansão substantiva na infraestrutura de transporte público pode afetar as taxas de evasão em escolas mais expostas a essa intervenção. Para estabelecer uma relação de causalidade, utilizamos a variação exógena proveniente da abertura progressiva de estações de metrô e a proximidade das escolas a essas novas instalações, utilizando um desenho de diferença-em-diferenças. Nossos resultados indicam que as escolas localizadas próximas às estações de metrô retêm significativamente mais alunos, tanto no ensino fundamental quanto no ensino médio. Esse impacto é inteiramente impulsionado pelas escolas públicas, que apresentam uma redução nas taxas de evasão na ordem de 24%. Embora o aumento nas taxas de retenção não afete o desempenho dos alunos nem esteja relacionado com a qualidade das escolas e dos professores, o background dos alunos desempenha um papel crucial na moderação do impacto da política de transporte. Nossos achados destacam que políticas de transporte podem ter efeitos de transbordamento na retenção escolar, contribuindo para a melhoria de indicadores educacionais mesmo em regiões mais urbanizadas.

Palavras-chave: Rede de metrô, evasão escolar, diferença-em-diferenças.

Área temática: Área 1 - Teoria, métodos e modelos de economia regional

---

*We benefited from helpful comments by participants at several seminars. We are solely responsible for the paper's contents. Lucas Emanuel acknowledges financial support from Edital JOVEMPESQ/PRPPG/UFBA.

†Federal University of Pernambuco. E-mail: edilberto.almeida@ufpe.br.
‡Federal University of Bahia. E-mail: lucasemanuel@ufba.br.
§Institute for Applied Economic Research, Brasília, Brazil. E-mail: rafael.pereira@ipea.gov.br.
¶Federal University of Pernambuco. E-mail: giuseppe.trevisan@ufpe.br.

1 Introduction

Transportation costs can be a significant barrier for school participation (Duflo, 2001; Muralidharan and Prakash, 2017). Previous studies have found that reducing transportation costs, for example through road construction or opening of new schools nearby students, can substantially contribute to reduce school dropout rates (Aaronson and Mazumder, 2011; Burde and Linden, 2013; Adukia et al., 2020). Most of these studies, with few exceptions (Asahi and Pinto, 2022), focus their analyses on rural contexts with low population and school densities. Moreover, there is scant evidence on the extent to which improving access to schools in urban areas through investments in mass transportation can impact school participation and achievement (Trajkovski et al., 2021).

In this paper, we use a quasi-experimental approach to examine how the expansion of a subway system can impact school participation in a large urban area. Using a difference-in-differences design, we leverage the staggered opening of new subway stations to examine the policy's causal impact on the dropout rates of public and private schools at both elementary and high school levels. We exploit the subway expansion policy in Salvador, the fifth-largest city in Brazil with 2.4 million people, where school performance remains a longstanding issue and the city spearheads nationwide concerns among poor school participation rates. School dropout rates in Brazil were more than twice the average of OECD countries in 2011, ranking behind only Mexico and Turkey (OECD, 2014). Approximately 30% of Brazilian students drop out before completing elementary school, and 25% of high school students drop out before graduation (Bruns et al., 2011).

Along with the staggered opening of the subway stations, we explore the proximity of schools to the new subway lines to define the treatment and control groups, as schools located near the stations are arguably more comparable. To address the challenges inherent in difference-in-differences designs with heterogeneity in the timing of events and in treatment effects, we employ the estimator developed by Callaway and Sant'Anna (2021). Our results indicate that schools more directly exposed to the subway network significantly improve student retention at both the elementary and high school levels. Dropout rates respond in the year immediately following the stations' inauguration, and the effects persist for at least six years. When analyzing the impact across different school types, we find that the average effect is entirely driven by public schools. On average, these schools experience a reduction in dropout rates of approximately 20% to 24% at both educational levels. While not directly comparable, our estimates indicate that the impact of expanding the subway network on reducing school dropout rates is similar in magnitude to the effects observed from affirmative action policies for public school students (Mello, 2023). Despite the increased retention rates, the transportation policy does not impact student performance measures in these schools. We show that our estimates are robust to various econometric specifications and that our results are not subject to violations of the parallel trends assumption, strengthening the causal interpretation of our findings.

The richness of our data allows us to explore several potential drivers of school participation. Specifically, improving access to schools can impact educational outcomes through the following channels (Asahi and Pinto, 2022). First, there may be indirect effects on the supply and quality of education services. Schools that are more accessible to a larger population could potentially attract better teachers if commuting costs are lower. However, our findings indicate that schools do not respond by altering the quality of their teachers or by reducing the student-to-teacher ratio. The second channel involves changes in the equilibrium conditions of supply and demand in the school market. Increased accessibility can affect supply and demand dynamics, which in turn might influence academic achievement through factors such as class sizes, competition in application processes, classmate turnover, and peer composition (Krueger, 1999; Krueger and Whitmore, 2001; Lavy et al., 2012; Burke and Sass, 2013). Our results show no increase in demand for schooling, nor do schools alter the number of classes or the student-to-class ratio.

The third channel is that the reduction in home-to-school distances increases the number of accessible schools, giving families and students more options to choose schools according to their preferences (Dustan and Ngo, 2018; Herskovic, 2020). This is closely linked to the literature on school choice (Dustan and Ngo, 2018; Herskovic, 2020; Burgess et al., 2015; Ruijs and Oosterbeek, 2019). Although we cannot directly test this connection due to a lack of data on students' addresses, our findings indicate that the subway network does not bridge the gap for students from other cities to access urban public schools. Although we cannot test for it, we acknowledge that other factors, such as reductions in nearby crime and improvements in traffic conditions, could play a role in school participation. Nevertheless, our results suggest that public transportation infrastructure may enhance school participation through channels that particularly benefit public school students, who are likely to incur higher transportation costs compared to those attending private schools.

Finally, we evaluate potential moderators in school participation. Specifically, we focus on the observed treatment effects among public schools to explore how school quality and students' background interacts with better accessibility to schools. While our findings suggest that education quality does not play a role in student retention in these schools, we find evidence that students in early elementary grades from more advantaged family backgrounds are better positioned to capitalize on the benefits of improved transportation accessibility. This result is in line with Dustan and Ngo (2018), who found that students from higher socioeconomic status (SES) backgrounds are more likely to select elite schools following improvements to the suburban train network.

The remainder of the paper is organized as follows. Section 2 provides the institutional background on the expansion of the subway network. Section 3 discusses the data, describes the variables, and outlines our definitions of the treatment and control groups. Section 4 details the identification strategy employed to estimate the impacts of the subway network on school participation. Section 5 presents the main results, and Section 6 concludes.

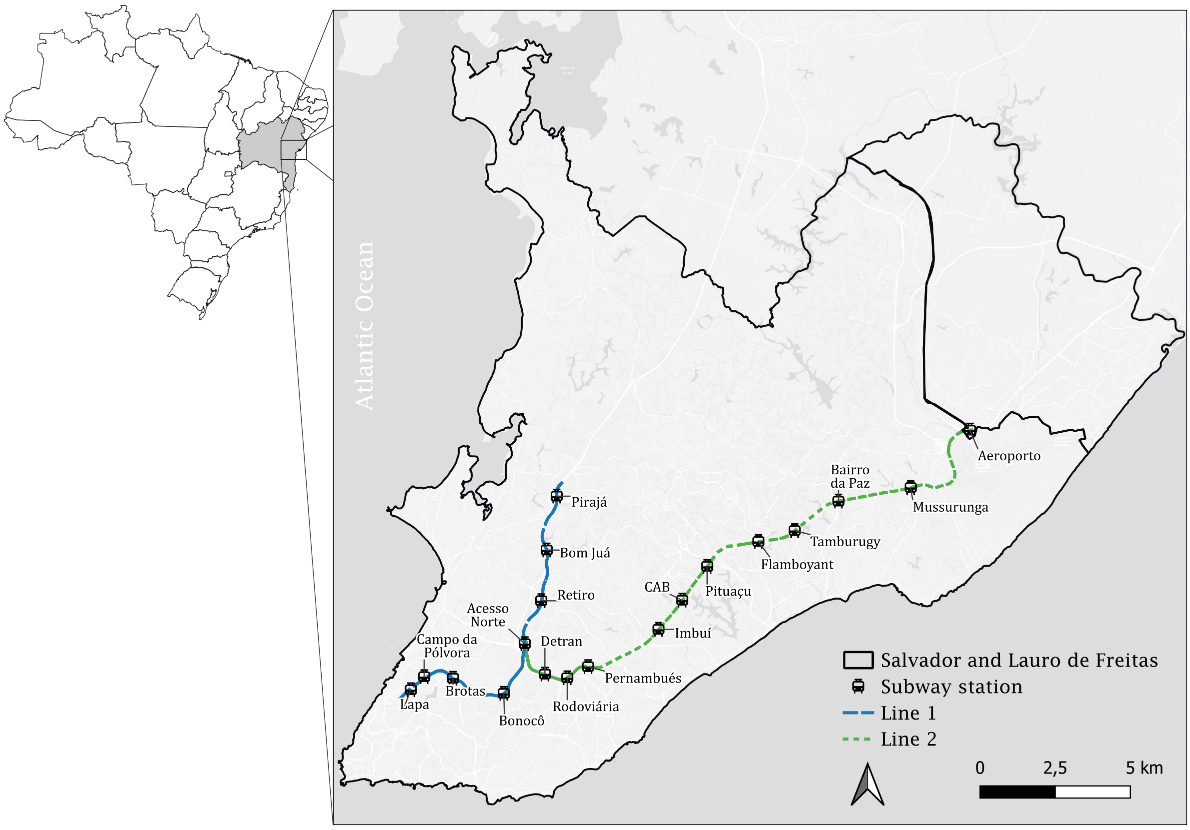
2 Institutional Background

2.1 The Subway Expansion Policy

According to data from the 2022 Census provided by the Brazilian Institute of Geography and Statistics (IBGE), Salvador, the capital of the state of Bahia, is the fifth most populous city in Brazil, with approximately 2.4 million inhabitants. Unlike other large Brazilian cities like São Paulo or Rio de Janeiro, Salvador's public transportation system has primarily relied on road transport. Previous research has shown that in the early 2010s the metropolitan region of Salvador had the sixth highest average commuting time among the country's largest metropolitan regions (behind only São Paulo, Rio de Janeiro, Recife, Belo Horizonte and Distrito Federal) and a worse commute than other large metropolitan regions in the world, such as Paris, Madrid, Chicago, Los Angeles, Santiago and Barcelona (Pereira and Schwanen, 2013; Pereira et al., 2021).

To improve urban accessibility and reduce commuting costs, the subway system was implemented in Salvador in 2014. The city's subway network has undergone various stages of implementation, eventually expanding to include 19 stations along its 33 km length (see Figure 1). The first phase was implemented in 2014 with the fist phase of Line 1 which includes five stations (Lapa, Campo da Pólvora, Brotas, Acesso Norte and Retiro). In 2015, three new stations were built on Line 1 (Bonocô, Bom Juá and Pirajá). The first two stations on Line 2 were inaugurated in 2016 (Detran and Rodoviária), integrating the historic city center with the new central business district (CBD) (Rodrigues et al., 2019). Eight more stations were inaugurated on Line 2 in 2017 (Bairro da Paz, Tamburugy, Imbuí, Pernambués, Flamboyant, Pituaçu, Mussurunga and CAB) and one in 2018 (Aeroporto).

Figure 1: Subway Network



Notes: This figure presents the subway system of Salvador and Lauro de Freitas.

With an average of 8 million passengers per month in 2022, the subway system also has 8 stations integrated with the metropolitan bus system, including the airport and the main bus terminal, which also integrates with the interstate bus system (CCR Metrô Bahia, 2022). According to CCR's 2018 data, Lapa and Pirajá are the two busiest stations in the entire subway system, with approximately 55,000 and 35,000 daily passengers, respectively. Additionally, the peak times occur in the morning (until 8:00 AM) and in the afternoon (between 4:00 PM and 6:30 PM), during which Pirajá and Lapa stations experience the highest passenger flows, respectively. In general, these patterns reflect the daily commute between suburban areas and the city center, facilitating access to employment, education, leisure, and other services typically concentrated in the CBD.

By enhancing access to services and workplaces, the subway system can affect the performance of economic agents in different ways. For instance, it can improve worker and firm productivity (Chen and Wu, 2024; Baek and Park, 2022), increase connectivity among individuals to foster innovation (Koh et al., 2022), influence real estate values (Tyndall, 2021; Zhou et al., 2021), and impact educational outcomes (Asahi and Pinto, 2022). In this context, the gradual expansion of Salvador's subway in a relatively short period of time (19 stations built from 2014 to 2018) introduces a staggered treatment timing that presents a unique quasi-experimental opportunity. This expansion allows us to examine the potential causal impact of increased accessibility to schools near subway stations on school outcomes, for example, reducing dropout rates.

2.2 Educational System

In Brazil, although federal education law guarantees the right to access education close to home, the process of admission to the public education system varies by state, particularly in terms of school choice. In the state of Bahia, in particular, families are allowed to choose which public school to enroll their children regardless of school location, even though families are encouraged to choose schools close to their home. In the municipality of Salvador, there are no restrictions on school choice either.

The educational system in Salvador comprises public schools (at the federal, state, and municipal levels) as well as private ones. In 2021, public schools constituted 41% of the basic education system, which includes early childhood education, elementary school, and high school (INEP, 2021). This percentage includes municipal schools (28%), state schools (13%), and two federal schools.

Salvador is among the worst Brazilian capitals in the Basic Education Development Index (IDEB) and Human Capital Index (HCI). In 2019, Salvador exhibited one of the lowest HCI among Brazilian capitals, significantly below the national average of 0.601 percent (World Bank, 2022). Specifically, while capitals like Vitória, Curitiba, and Campo Grande presented high-performing HCI comparable to those of Chile (0.652 percent), China (0.653 percent), and Turkey (0.649 percent), Salvador recorded a much lower index of around 0.56 percent (World Bank, 2020). This positioned Salvador as the third lowest-ranking capital in terms of HCI, ranking higher only than Belém and Macapá, two capitals located in the northern region of Brazil. Additionally and particularly pertinent to our analysis, data from the 2019 School Census reveals that Salvador holds the fifth highest dropout rate among Brazilian capitals (of 27 including the Federal District) in public elementary school and it ranks eleventh in dropout rates for public high school.

3 Data

3.1 Data Sources and Variables

For the main analysis of the effects of subway network expansion on schools, we constructed an eleven-year panel (2009–2019) of aggregated school data using the Brazilian School Census, which is conducted annually by the National Institute for Educational Studies and Research Anísio Teixeira (INEP), a research agency linked to the Ministry of Education (MEC). This school-level data set covers all schools in Brazil, from preschools to high schools in a given year, whether they are municipally, state, federally funded, or private schools. The dataset includes information about students, class composition, and other school characteristics.

Our main educational outcome is the school dropout rate, segmented by stage of education. The dropout rate indicates the percentage of students who stopped attending school after the reference date of the School Census. For primary education, these rates are calculated separately for the early grades (1st to 5th grade) and upper grades (6th to 9th grade). Information on dropout rates are essential for monitoring and verifying education performance across the country, and they are provided annually by INEP through the Basic Education School Census. The calculation of dropout rates, as well as approval and failure rates, is based on student performance and movement information collected in the second stage of the School Census, called "Student Situation." At this stage, respondents provide performance information ("approved" or "failed"), movement ("deceased," "dropped out," or "transferred"), ongoing courses, and non-movement of students declared in the first stage of the School Census (Initial Enrollment). The enrollments considered for the calculation of these rates apply to primary education, regular secondary education, and integrated technical courses (integrated secondary education), for which the statuses "approved," "failed," or "dropped out" are reported.[1]

Our analysis focuses on the cities of Salvador and Lauro de Freitas — the two municipalities in the metropolitan region of Salvador directly affected by the subway system. To our school-level panel covering all schools in these two municipalities from 2009 to 2019, we add two additional pieces of data: the geolocation of schools, also collected from INEP, and data on the location and year of inauguration of each subway station, from CCR Metrô Bahia. In order to capture both the socioeconomic profile of students and the average quality of schools in terms of academic performance, we used INEP's data on School Socioeconomic Level Indicators and the Basic Education Development Index.

[1] For more details, see https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/taxas_de_rendimento_escolar_2022.pdf

School socioeconomic level indicators (INSE) are calculated by INEP, and are intended to categorize students within a socioeconomic group. This classification is based on the possession of household goods, family income, the hiring of services by the students' families, and the educational level of their parents. The socioeconomic level measure used in this study was derived from data collected from Prova Brasil and Aneb in 2011 and 2013 — questionnaires filled out by students —, as well as ENEM data from 2011 and 2013, covering the period before the introduction of the subway stations (INEP, 2014). Basic Education Development Index (*"Índice de Desenvolvimento da Educação Básica"* - IDEB), which ranges from 0 to 10, is calculated every two years by INEP. It combines student results in Portuguese Language and Mathematics in the SAEB (Basic Education Assessment System - *"Sistema de Avaliação da Educação Básica"*) and approval rates. SAEB consists of a series of large-scale external assessments conducted periodically using standardized instruments (cognitive tests and questionnaires) targeting specific stages of basic education (INEP, 2019).

3.2 Defining the Sample, Treated and Control Groups, and Descriptive Statistics

For each school, we calculate the Euclidean distance to the closest subway station. To focus on schools located in areas closer to the stations — with greater potential similarity in terms of neighborhood characteristics, geography, and people flow —, we include only schools situated within 4 km of any point along a subway line. This restriction leaves us with a main analysis sample of 1,490 schools out of the total 1,608 geocoded schools located in the municipalities of Salvador and Lauro de Freitas.

We define the treatment group as those schools located within a 2 km radius around the nearest subway station, so that the control group is composed of schools located up to 4 km away from the subway points. These schools in our comparison group are arguably more comparable to those near the stations, as they are subject to similar geography and neighborhood economic activity and tend to be influenced by the same provision of other types of public transport and travel behavior.[2] The definition of treatment based on an arbitrary coverage radius is a commonly adopted approach in empirical studies evaluating the effects of events with uncertain reach (Monteiro and Rocha, 2017; Zhu et al., 2016). It is also usual in other studies of subway use (Keijer and Rietveld, 2000) and, more related to our paper, studies that attempted to assess the impact of subway access on distance traveled to school (Herskovic, 2020).[3]

Keeping the data from the control group more similar to the treated group offers advantages such as higher comparability and lower risk of differential trends driven by other factors. However, this definition of the control group can also be problematic since the control schools are in the same spatial area and might also be directly or indirectly affected by the higher accessibility exposure, potentially biasing our estimates. We address this issue in two different ways in our robustness analyses.

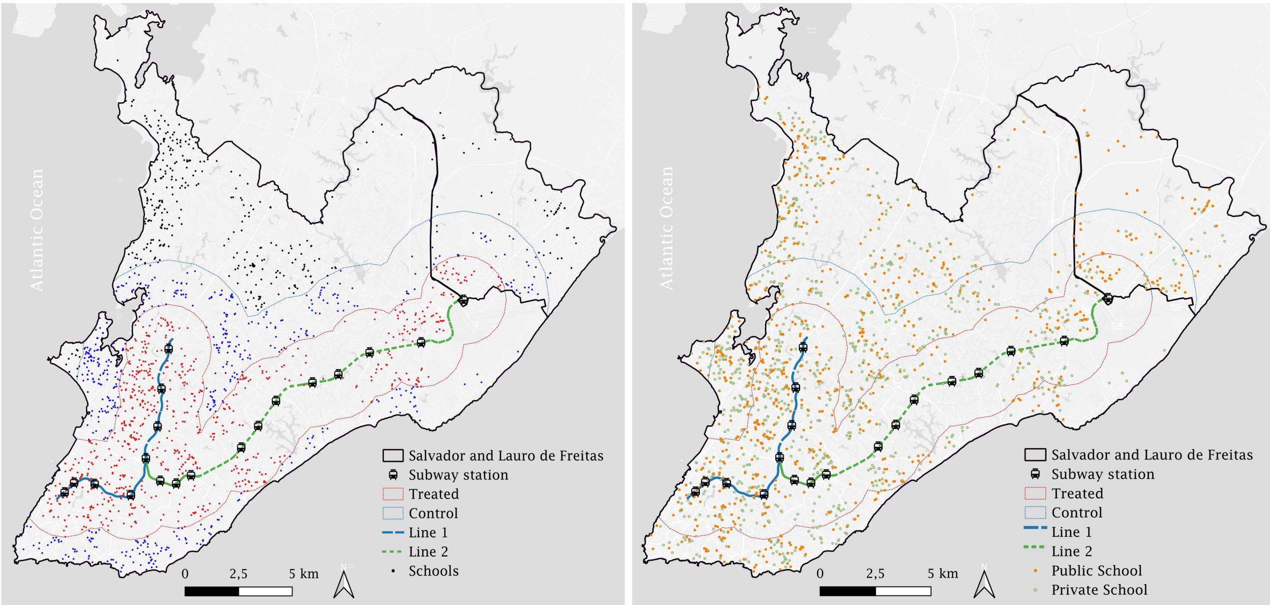
Panel A of Figure 2 shows the spatial distribution of all schools on a map of the Salvador and Lauro de Freitas municipalities. Treated schools (represented by red dots) are located within 2 km of each subway station, while control group schools (blue dots) are those located between 2 km and 4 km from a subway station. Panel B of Figure 2 shows that private schools (green dots) and public schools (orange dots) are well distributed throughout the city, with no particular concentration of either category in any specific area. Panel B, therefore, suggests that the potential benefits of constructing the subway lines may have impacted both groups of schools in terms of accessibility gains. We test the specific effect by type of school in this study.

[2]We assess the robustness of our results by re-estimating the coefficients of interest with alternate definitions of the treatment perimeter. Including schools located further from subway stations in the treatment group tends to push the estimates toward zero. Results are available upon request.

[3]As a reference, in Salvador city, the average commuting time was 33.9 minutes (or 0.565 h) in 2010 (Pereira and Schwanen, 2013) and the "comfortable" walking speed is 3.5 km/h (Duim et al., 2017). Thus, using a simple calculation, the average commuting distance is $3.5 \times 0.565 = 2.04$km.

Figure 2: Treatment and Control Groups

a Treated and control **b** Public and Private Schools



Notes: Figure 2, panel A, shows the spatial distribution of schools on a map of the Salvador and Lauro de Freitas municipalities. According to panel B, there is no specific clustering of public or private schools around subway stations. The two subway lines that existed at the end of 2018 are displayed for reference.

Table 1 shows descriptive statistics for some selected characteristics in our sample, for all schools and separately for public and private schools. The table highlights that public and private schools exhibit very similar average distances to the nearest subway station, with public schools averaging 1.91 km and private schools averaging 1.87 km. However, significant differences are observed in dropout rates. The average dropout rate for public schools is consistently higher across all education stages. For instance, in the early years of elementary school, the dropout rate for public schools is 3.14% compared to 0.77% for private schools. Similarly, in the final years of elementary school, public schools have an average dropout rate of 5.28%, while private schools have an average of only 0.5%. In high school, the disparity is even more pronounced, with public schools experiencing a dropout rate of 13.91% compared to just 0.39% for private schools. These statistics underscore the disparities in educational outcomes between public and private schools.

Table 1: Descriptive statistics

	All schools		Public schools		Private schools	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
	(1)	(2)	(3)	(4)	(5)	(6)
School-subway distance (km)	1.891	0.891	1.908	0.905	1.871	0.873
Dropout rate						
Elementary (early grades)	1.951	4.108	3.139	4.209	0.768	3.638
Elementary (upper grades)	3.392	5.156	5.278	5.370	0.497	3.050
High school	8.615	11.78	13.91	12.47	0.391	1.247
Approval rate						
Elementary (early grades)	91.72	9.414	87.05	9.604	96.37	6.460
Elementary (upper grades)	78.43	16.65	68.62	13.17	93.47	7.896
High school	75.16	18.31	64.08	14.31	92.38	7.104
Repetition rate						
Elementary (early grades)	6.327	7.577	9.808	8.018	2.862	5.149
Elementary (upper grades)	18.18	14.15	26.10	11.79	6.028	7.097
High school	16.22	12.10	22.01	11.14	7.228	6.946
Class size						
Elementary (early grades)	19.86	7.378	24.11	4.770	15.99	7.203
Elementary (upper grades)	26.55	9.138	29.59	5.733	22.92	10.94
High school	32.34	8.499	32.54	6.187	32.07	11.01
Proportion of schools with:						
Filtered water	0.986	0.119	0.981	0.135	0.991	0.0950
Access to the sewage network	0.942	0.233	0.911	0.284	0.980	0.140
Number of schools in the sample:	1,490		656		834	

Notes: This table reports the mean and standard deviation (S.D.) for our main variables, for all schools and separately for public and private schools in our sample. Summary statistics are calculated at the school-year level from 2009 to 2019, for the schools in areas within a 4 km radius of the subway stations.

4 Empirical Strategy

This study aims to estimate the impact of the construction and expansion of a subway network on school dropout rates over time. Ideally, we would compare the educational outcomes of schools in affected areas with a counterfactual scenario in which the subway stations were not present. However, because this counterfactual cannot be directly observed, we employ a quasi-experimental approach to address this issue.

Using a difference-in-differences research design, we compare the changes in dropout rates between schools located within a 0–2 km radius of the subway stations (the treated group) against those within a 2–4 km radius (the control group) exploiting variation in the timing of subway stations installation. The underlying identifying assumption is, absent the subway stations installation, dropout rates would have evolved similarly in both groups over time. However, due to the heterogeneity in treatment timing — since subway stations were inaugurated in different years —, using a plain-vanilla two-way fixed effects (TWFE) estimator in a difference-in-differences design may lead to bias and may not accurately capture causal effects. As pointed out by Goodman-Bacon (2021), the TWFE estimator can generate problematic negative weights, leading to misleading comparisons between early-treated and late-treated units.

To address the issue of 'forbidden comparisons' in staggered DiD designs, we use the event-study estimator proposed by Callaway and Sant'Anna (2021) for DiD models with variation in treatment timing. This method allows us to estimate the impact of subway network expansion while accounting for heterogeneous treatment effects across cohorts and potential anticipation effects. We use schools near the subway stations in *Lauro de Freitas*, *Campinas*, and *Águas Claras* as the never treated group, as these stations were only inaugurated in 2023.[4] A key advantage of using Callaway and Sant'Anna (2021)'s estimator is its flexibility in incorporating covariates, thereby avoiding the need for restrictive assumptions such as homogeneous treatment effects across covariates or the absence of covariate-specific trends. The treatment effect on the treated (ATT) may vary between treatment groups or cohorts g and over time t and, thus, proposes the group-time ATT, denoted by $ATT(g,t)$. In this approach, each treated school i can be classified into a cohort $G_i = g \in \{2014, ..., 2018\}$ based on the first year of exposure to the subway station, denoted by g.

This empirical strategy enables us to more accurately capture the effects of subway network expansion on school dropout rates, providing robust estimates that reflect the true impact of increased accessibility on educational outcomes. We model the relationship of interest as in the following DiD-based regression:

$$Y_{it} = \left[\sum_{\tau=-6}^{-1} \beta_\tau I(t_t - t^* = \tau) + \sum_{\tau=1}^{5} \beta_\tau I(t_t - t^* = \tau) \right] + X_{it}'\Theta + \gamma_i + \theta_t + \epsilon_{it} \tag{1}$$

where Y_{it} is the outcome of interest (school dropout rate) for school i at year t; γ_i and θ_t are school and time fixed effects, respectively. The indicator $I(t_t - t^* = \tau)$ represents the time (in years) relative to the year of the subway stations installation near to school, while X_{it} is a vector containing school-level covariates with yearly variation, such as access to filtered water and sewage networks. ϵ_{it} is an idiosyncratic error term. By estimating $\beta_{\tau<0}$ and checking statistical significance, we can empirically test the plausibility of the identifying assumption. Thus, under conditional parallel trends assumption, the coefficients $\beta_{\tau\geq 0}$ represent the causal effects of the subway network on the outcome of interest. Standard errors are clustered at the school level.

5 Results

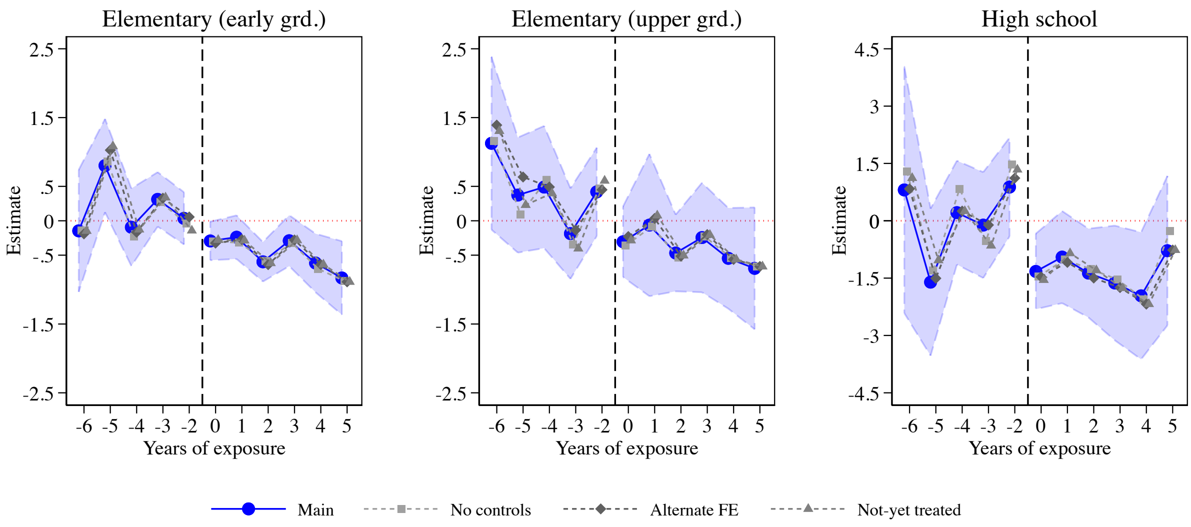
We begin this section by presenting our main results, specifically that dropout rates decline in schools located near the subway stations. We show the impacts are robust and vary across public and private schools. Furthermore, we exclude several potential mechanisms that could explain the impacts of subway accessibility and evaluate possible moderators influencing the increase in school retention rates.

5.1 Impacts on Dropout Rates

A necessary condition for establishing a causal relationship between subway access and dropout rates is that pre-treatment trends in potential outcomes of treated and control schools are parallel. Although the underlying identification assumption is not directly testable, we can assess its plausibility. We leverage the extended period before and after the installation of subway stations to check for anticipation effects and examine dynamic treatment effects. Figure 3 reports the event study results for each educational category, i.e. elementary (early and upper grades) and high-school levels. The blue dots represent pre- and post-treatment effects of subway access derived from our main specification (Equation 1), along with 90% confidence intervals. Treatment effects are measured in terms of relative differences from the baseline outcome of treated schools in the year just before installation. As observed in all graphs, the wide confidence intervals for estimated coefficients in the pre-treatment period formally discard the presence of anticipatory effects, thus strengthening the reliability of our empirical strategy. The absence of pre-trends in dropout rates suggests that transit spillovers did not affect schools during the construction of subway stations.

[4] We also test the robustness of our estimates by using only not-yet-treated units. Results are available upon request.

Figure 3: Dynamic Effects of Subway Access on Dropout Rates



Notes: This figure presents presents event-study estimated coefficients. Shaded areas represent 90% confidence interval.

In the year following the opening of subway stations, treated schools experience a significant decrease in dropout rates among students in the early grades of elementary education. The effect appears to be progressive in further years, suggesting the improvement in accessibility to schools through the new subway network helps to retain these students. Although the magnitudes of effects in post-treatment years are similar between elementary grade groups, we cannot affirm that students in treated schools from the upper elementary grades benefit from the policy. Conversely, high-school dropout rates sharply decline in schools located near subway stations, a trend that persists for several years. In the first treatment year, dropout rate significantly falls by 1.5 percentage point. For each grade group, we show that point estimates remain remarkably robust when excluding covariates, including municipality linear trends, and adding not-yet treated observations to the control group.

The descriptive statistics show the distribution of public and private schools varies across different grade groups. Consequently, our main treatment effect estimates may mask some nuances related to school type, as students in public and private schools, e.g., may have distinct preferences for using public transportation. As shown in the right panel of Figure 2, the spatial distribution of these schools around the subway stations appears quasi-random. This lack of school-type clustering thus endorses the plausibility of our heterogeneity analysis. In Table 2, we average treatment effects across all treatment cohorts and periods using a single-coefficient DD approach based on Equation 1, showing results for our main sample and separately for public and private schools. The results exhibited in panels B and C reveal a contrasting pattern of subway accessibility effects across school types.

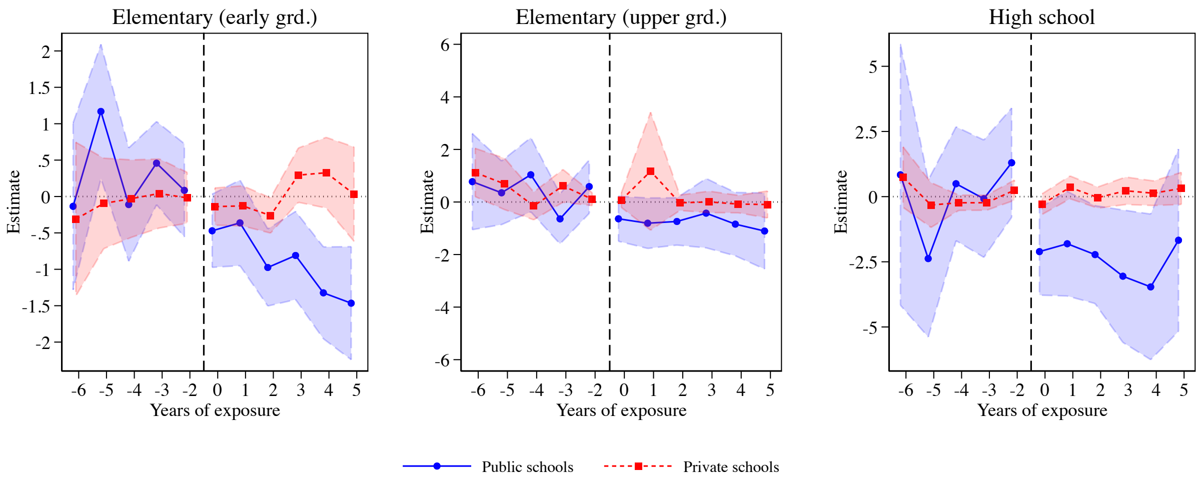
Table 2: Effects of Subway Access on Dropout Rates

	Elementary (early grd.)				Elementary (upper grd.)				High school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. All schools												
Estimate	-0.43**	-0.46***	-0.46**	-0.43***	-0.41	-0.43	-0.40	-0.32	-1.33**	-1.33*	-1.47**	-1.19*
	(0.17)	(0.17)	(0.19)	(0.16)	(0.33)	(0.33)	(0.36)	(0.32)	(0.65)	(0.69)	(0.71)	(0.65)
Baseline rate (%)	1.95				3.42				6.93			
Observations	6,421	6,421	6,421	6,439	4,070	4,070	4,070	4,076	2,322	2,322	2,322	2,323
Panel B. Public schools												
Estimate	-0.83***	-0.86***	-0.69*	-0.82***	-0.83*	-0.83*	-0.63	-0.71	-2.33**	-2.32**	-1.88	-2.24**
	(0.29)	(0.28)	(0.36)	(0.29)	(0.50)	(0.48)	(0.57)	(0.48)	(1.08)	(1.05)	(1.33)	(1.06)
Baseline rate (%)	3.42				5.57				11.60			
Observations	3,481	3,481	3,481	3,481	2,625	2,625	2,625	2,625	1,457	1,457	1,457	1,457
Panel C. Private schools												
Estimate	-0.01	-0.01	-0.01	-0.03	0.24	0.23	0.26	0.27	0.08	0.05	0.15	0.29
	(0.16)	(0.15)	(0.17)	(0.14)	(0.33)	(0.32)	(0.36)	(0.32)	(0.20)	(0.20)	(0.23)	(0.27)
Baseline rate (%)	0.57				0.37				0.47			
Observations	2,940	2,940	2,940	2,958	1,445	1,445	1,445	1,451	865	865	865	866
School FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✓		✓	✓	✓		✓	✓	✓		✓	✓
Municipality linear trend			✓				✓				✓	
Not-yet treated obs.				✓				✓				✓

Notes. Each cell reports the coefficient of a separate regression. Robust standard errors clustered at the school level are in parentheses. Significance codes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The positive effects of subway access on student retention rates shown in our main results (panel A) are solely attributed to treated public schools. The baseline estimate for early elementary grades (column 1) shows that dropout rate falls by .83 percentage points in public schools, representing a 24% decrease relative to the baseline rate. Public schools also experience a statistically and economically significant reduction in dropout rates among students attending high school grades. Our preferred specification shows a 2.33 percentage points (20%) decline in dropout rates. The coefficients are robustly estimated across different specifications. The treatment effects for upper elementary grades are less precisely estimated but similar in magnitude to those for early grades, as illustrated in columns (5)–(8). Conversely, the provision of the subway network does not affect student retention rates in private schools in any grades. The event study results presented in Figure 4 validate the DD identifying assumption and confirm the flat trends in dropout rates among private schools, while showing significant post-treatment shifts in the outcome among public schools. Table 3 indicates that, despite the increased retention of students in school, performance outcomes such as approval and repetition rates and age-grade distortion remain unaffected.

Figure 4: Dynamic Effects of Subway Access on Dropout Rates by School Type



Notes: This figure presents presents event-study estimated coefficients. Shaded areas represent 90% confidence interval.

Table 3: Effects of Subway Access on Students' Performance

	Elementary (early grd))			Elementary (upper grd.)			High school		
	Approv. rate	Repet. rate	Age-grade distortion	Approv. rate	Repet. rate	Age-grade distortion	Approv. rate	Repet. rate	Age-grade distortion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. All schools									
Estimate	-0.05	0.47	0.25	0.01	0.40	0.22	0.59	0.74	-0.54
	(0.41)	(0.36)	(0.58)	(0.79)	(0.74)	(0.61)	(0.98)	(0.94)	(0.79)
Observations	6,421	6,421	6,529	4,070	4,070	3,917	2,322	2,322	2,390
Panel B. Public schools									
Estimate	0.14	0.69	0.32	-0.74	1.58	0.72	1.85	0.48	-0.28
	(0.70)	(0.60)	(0.86)	(1.20)	(1.12)	(0.84)	(1.49)	(1.46)	(1.13)
Observations	3,481	3,481	3,546	2,625	2,625	2,415	1,457	1,457	1,462
Panel C. Private schools									
Estimate	-0.29	0.29	0.17	0.92	-1.15	-0.45	-1.25	1.17	-0.87
	(0.39)	(0.37)	(0.73)	(0.82)	(0.75)	(0.86)	(1.05)	(0.99)	(1.05)
Observations	2,940	2,940	2,982	1,445	1,445	1,502	865	865	928
School FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Each cell reports the coefficient of a separate regression. Robust standard errors clustered at the school level are in parentheses. Significance codes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Drawing on evidence from studies exploring the effects of transportation infrastructure expansion on student outcomes (Dustan and Ngo, 2018; Herskovic, 2020), our findings highlight the significant potential of these policies to influence multiple facets of student attainment. In particular, schools that are more closely integrated with these infrastructures may retain more students. However, in our context, significant declines in dropout rates appear to be exclusively experienced by public schools as a result of subway network provision. Although we lack specific information on students' actual use of subway transport and cannot claim it as the main mechanism, we conduct several additional analyses to rule out other candidate drivers of these effects and to identify key moderators.

5.2 Additional Results

5.2.1 Potential Channels

We begin by exploring potential drivers of school achievement documented in the literature that might also explain the positive effects on school retention. For transparency, we present the results both in aggregate and separately for public and private schools.

The decrease in dropout rates among public schools could be attributed to changes in their infrastructure. For instance, improvements in commuting time and better school access facilitated by the subway network may have led to increased demand for education, prompting schools to expand the number of classes to better accommodate students. To assess these potential channels, Table 4 estimates the effects of subway accessibility on total enrollment and the number of classes, separately for elementary grades and high school. The results reported in columns (1)–(5) indicate that treated public schools do not significantly change these dimensions. In columns (6) and (7), we examine the impact of subway accessibility on a proxy for students' school choice. Since we lack information on students' addresses, we measure school choice as the number of students living in a city other than where the school is located. The coefficients are noisily estimated, strongly suggesting that schools are not attracting students from other cities. We acknowledge that this analysis is limited as we cannot observe whether subway stations have changed the school choice behavior of students *within* the city of Salvador.

Table 4: Effects of Subway Access on School Characteristics

	Share of students			Student-to-class		N. of classes		Live in other city		
	Elem. (early)	Elem. (upper)	High school	Elem. grades	High school	Elem. grades	High school	Elem. (early)	Elem. (upper)	High school
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. All schools										
Estimate	-0.00	0.01	-0.01	0.01	0.86	0.01	-0.02*	0.05	0.31	2.11
	(0.01)	(0.01)	(0.01)	(0.20)	(0.63)	(0.01)	(0.01)	(0.55)	(0.70)	(2.43)
Observations	6,386	4,052	2,321	9,014	2,433	9,547	2,625	5,816	3,695	2,149
Panel B. Public schools										
Estimate	-0.01	0.02	-0.02	-0.29	0.56	0.01	-0.02	0.52	0.03	1.17
	(0.01)	(0.02)	(0.02)	(0.23)	(0.68)	(0.01)	(0.02)	(0.95)	(0.68)	(1.94)
Observations	3,477	2,617	1,457	5,479	1,462	5,625	1,571	3,102	2,387	1,380
Panel C. Private schools										
Estimate	0.00	0.00	-0.01	0.49	1.26	0.01	-0.03	-0.37	0.95	4.18
	(0.01)	(0.02)	(0.02)	(0.36)	(1.17)	(0.01)	(0.02)	(0.49)	(1.38)	(5.33)
Observations	2,909	1,435	864	3,534	971	3,920	1,054	2,711	1,308	769
School FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Each cell reports the coefficient of a separate regression. Robust standard errors clustered at the school level are in parentheses. Significance codes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Next, we examine the effects of subway stations on school staffing characteristics. The rationale for this exercise is that more accessible schools could improve teacher-school matching by attracting and retaining better teachers due to reducing commuting costs.[5] Table 5 reports the treatment effects on teachers characteristics, such as student-to-teacher ratio, share of high-skilled teachers (i.e., with higher

[5] In Brazil, most public school teachers are hired through a public examination process, which involves several stages and is highly competitive. Since the position guarantees job stability, the turnover rate for these professionals is low compared to those in private schools.

education degree), and the share of teachers in elementary grades and high school level. Among public schools, estimated coefficients are statistically insignificant in all regressions. Our findings do not support the role of these mechanisms.

Table 5: Effects of Subway Access on Teacher Characteristics

	Student-to-teacher			Share of teachers	
	Elementary grades (1)	High school (2)	High-skilled teachers (3)	Elementary grades (4)	High school (5)
Panel A. All schools					
Estimate	0.66**	0.00	-0.00	0.00	-0.01
	(0.29)	(0.23)	(0.01)	(0.01)	(0.01)
Observations	8,997	2,426	10,423	8,995	2,426
Panel B. Public schools					
Estimate	0.60	-0.01	0.00	0.00	-0.01
	(0.42)	(0.41)	(0.01)	(0.01)	(0.02)
Observations	5,479	1,462	6,757	5,479	1,462
Panel C. Private schools					
Estimate	0.79**	-0.00	-0.02	0.01	-0.02
	(0.37)	(0.15)	(0.02)	(0.01)	(0.02)
Observations	3,517	964	3,665	3,515	964
School FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓

Notes. Each cell reports the coefficient of a separate regression. Robust standard errors clustered at the school level are in parentheses. Significance codes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

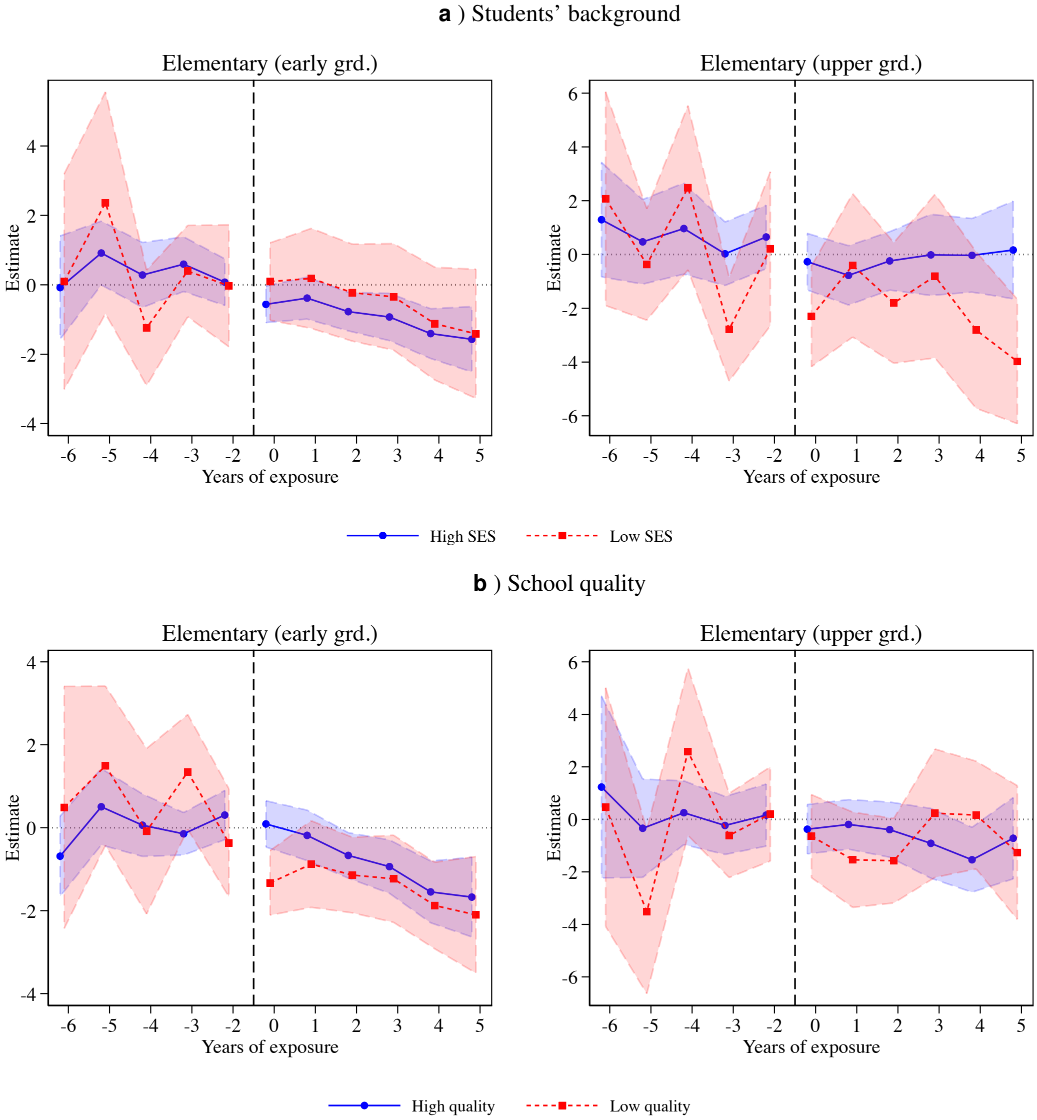
While we have ruled out several potential drivers and lack data to pinpoint the primary mechanism behind the decline in dropout rates, we acknowledge that other factors could still explain the observed patterns among public schools. For example, the expansion of subway network might influence crime rates around nearby schools by increasing pedestrian traffic. Given that higher homicide rates near schools are associated with increased student dropouts (Koppensteiner and Menezes, 2021), increased visibility could deter criminal activity. Additionally, the installation of the subway stations could improve traffic flow in the surrounding area, thereby reducing commuting times for car and bus users.

5.2.2 Effect Heterogeneity

In this section, we analyze the heterogeneous treatment effects of subway access on dropouts across school characteristics, leveraging the comprehensive data provided by the INEP source. The primary advantage of this dataset is that it allows us to measure pre-subway levels of school quality based on an objective index and assess students' background. However, the dataset is limited in that most information is available only for public schools at the elementary levels. To compare schools of different quality levels, we split the sample into high- and low-quality schools based on their pre-treatment IDEB scores. Similarly, using the INSE index described in Section 3, we divide schools at the median value. Specifically, we categorize schools based on whether they enroll predominantly high SES or low SES students.

Figure 5 presents the event-study results on dropout rates for the subsample of public schools. The top-left graph reveals that, although point estimates for high and low SES schools in early elementary grades appear similar in later years, significant effects on student retention rates are observed only in high SES schools. This finding complements Dustan and Ngo (2018)'s results, which suggest that students from high SES backgrounds are more likely to choose elite schools following improvements in the suburban train network. That is, students from more advantageous backgrounds in non-elite schools may also benefit from improved access to transportation infrastructure, given their likely stronger inclination toward pursuing higher educational levels than their less advantaged counterparts. For upper elementary grades, however, we cannot determine the moderating effect of students' backgrounds on dropout rates.

Figure 5: Heterogeneous Dynamic Effects of Subway Access on Dropout Rates Among Public Schools



Notes: This figure presents presents event-study estimated coefficients. Shaded areas represent 90% confidence interval.

The bottom graphs display the results separately for high-quality and low-quality public schools. While the dynamic treatment effects in upper elementary grades are flat and not statistically significant, a reduction in dropout rates is evident among early grade students in schools of both quality levels. While prior studies indicate that students tend to choose more accessible schools over lower-performing ones in response to transportation policies, our findings suggest that education quality does not play a role in student retention in public schools. This is particularly relevant for families who cannot afford expensive, higher-quality private schools and therefore place greater value on obtaining a school degree for their children. Altogether, it appears that family background plays a crucial role in leveraging the benefits of nearby schools influenced by transportation infrastructure.

6 Final Remarks

This study provides robust evidence on the impact of transportation infrastructure investments, specifically the construction and expansion of a subway network, on school participation in a large urban area. Utilizing geocoded school-level data and precise locations of new subway stations, we assess causality by estimating a staggered difference-in-differences regression. This approach effectively addresses challenges related to treatment effect heterogeneity and differential timing.

Our main findings show that schools located near the subway network significantly increase student retention. In particular, the positive impacts of subway expansion are most evident in public schools, where dropout rates fall considerably following the opening of subway stations. In contrast, private schools do not experience similar benefits. The transportation policy does not lead to changes in school infrastructure, staff quality, or the composition of students and classes. Additionally, the improvement in school participation does not translate into better student performance measures. On the other hand, our findings align with previous research showing that students from more advantaged family backgrounds are better positioned to capitalize on the benefits of improved transportation accessibility.

While our results add to the growing body of literature on the relationship between transportation infrastructure and educational outcomes (Herskovic, 2020; Trajkovski et al., 2021), it is important to recognize the study's limitations. A primary limitation is the inability to directly measure students' actual use of the subway system, which restricts our ability to precisely attribute the observed changes in dropout rates to enhanced transportation access. Additionally, our analysis is based on school-level data, which may mask individual-level variations and other contextual factors that influence school participation. Furthermore, the study is focused on a specific urban context, which may limit the generalizability of the findings to other cities with different transportation dynamics and educational systems. Overall, our findings suggest that public transportation infrastructure may induce school participation through channels that disproportionately benefit public school students. While we cannot unveil such mechanisms, our findings provide suggestive evidence that the subway network improves school participation for those who intrinsically face higher transportation costs.

Despite these limitations, our findings offer valuable insights into the potential of urban infrastructure development to influence social outcomes, particularly in education. The results highlight the significant role that transportation policies can play in reducing educational disparities by increasing access to schools. However, the observed differences in effects between public and private schools suggest that future policies should be tailored to address the specific needs of different educational sectors to fully realize the benefits of such investments. Continued research is required to explore the long-term impacts of transportation infrastructure on educational attainment and to better understand the underlying mechanisms.

References

Aaronson, Daniel, & Mazumder, Bhashkar (2011). "The impact of Rosenwald schools on black achievement," *Journal of Political Economy*, 119 (5): 821–888.

Adukia, Anjali, Asher, Sam, & Novosad, Paul (2020). "Educational investment responses to economic opportunity: evidence from Indian road construction," *American Economic Journal: Applied Economics*, 12 (1): 348–376.

Asahi, Kenzo, & Pinto, Ignacia (2022). "Transit, academic achievement and equalisation: evidence from a subway expansion," *Journal of Economic Geography*, 22 (5): 1045–1071.

Baek, Jisun, & Park, WooRam (2022). "The impact of improved passenger transport system on manufacturing plant productivity," *Regional Science and Urban Economics*, 96: p. 103805.

Bruns, Barbara, Evans, David, & Luque, Javier (2011). *Achieving world-class education in Brazil: The next agenda*, World Bank Publications.

Burde, Dana, & Linden, Leigh L (2013). "Bringing education to Afghan girls: A randomized controlled trial of village-based schools," *American Economic Journal: Applied Economics*, 5 (3): 27–40.

Burgess, Simon, Greaves, Ellen, Vignoles, Anna, & Wilson, Deborah (2015). "What parents want: School preferences and school choice," *The Economic Journal*, 125 (587): 1262–1289.

Burke, Mary A, & Sass, Tim R (2013). "Classroom peer effects and student achievement," *Journal of Labor Economics*, 31 (1): 51–82.

Callaway, Brantly, & Sant'Anna, Pedro HC (2021). "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225 (2): 200–230.

CCR Metrô Bahia (2022). "Metrô em numeros."

Chen, Guoliang, & Wu, Wuqing (2024). "The effect of subways on firm-level productivity," *Economics Letters*, 235: p. 111536.

Duflo, Esther (2001). "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," *American economic review*, 91 (4): 795–813.

Duim, Etienne, Lebrão, Maria Lucia, & Antunes, José Leopoldo Ferreira (2017). "Walking speed of older people and pedestrian crossing time," *Journal of Transport & Health*, 5: 70–76.

Dustan, Andrew, & Ngo, Diana KL (2018). "Commuting to educational opportunity? School choice effects of mass transit expansion in Mexico City," *Economics of Education Review*, 63: 116–133.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing," *Journal of econometrics*, 225 (2): 254–277.

Herskovic, Luis (2020). "The effect of subway access on school choice," *Economics of Education Review*, 78: p. 102021.

INEP (2014). "Nota técnica - Indicador de nível socioeconômico (inse) das escolas, 2011-2013,"Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

INEP (2019). "Nota técnica - Índice de Desenvolvimento da Educação Básica – Ideb,"Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

INEP (2021). "Sinopse Estatística da Educação Básica 2021 - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira."

Keijer, MJN, & Rietveld, Piet (2000). "How do people get to the railway station? The Dutch experience," *Transportation planning and technology*, 23 (3): 215–235.

Koh, Yumi, Li, Jing, & Xu, Jianhuan (2022). "Subway, collaborative matching, and innovation," *Review of Economics and Statistics* 1–45.

Koppensteiner, Martin Foureaux, & Menezes, Lívia (2021). "Violence and Human Capital Investments," *Journal of Labor Economics*, 39 (3): 787–823.

Krueger, Alan B (1999). "Experimental estimates of education production functions," *The Quarterly Journal of Economics*, 114 (2): 497–532.

Krueger, Alan B, & Whitmore, Diane M (2001). "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR," *The Economic Journal*, 111 (468): 1–28.

Lavy, Victor, Silva, Olmo, & Weinhardt, Felix (2012). "The good, the bad, and the average: Evidence on ability peer effects in schools," *Journal of Labor Economics*, 30 (2): 367–414.

Mello, Ursula (2023). "Affirmative action and the choice of schools," *Journal of Public Economics*, 219: p. 104824.

Monteiro, Joana, & Rocha, Rudi (2017). "Drug battles and school achievement: evidence from Rio de Janeiro's favelas," *Review of Economics and Statistics*, 99 (2): 213–228.

Muralidharan, Karthik, & Prakash, Nishith (2017). "Cycling to school: Increasing secondary school enrollment for girls in India," *American Economic Journal: Applied Economics*, 9 (3): 321–350.

OECD (2014). *Drop-out rates in Brazil and OECD countries, 2011*, OECD, https://doi.org/https://doi.org/10.1787/9789264208988-graph51-en.

Pereira, Rafael H. M., & Schwanen, Tim (2013). "Tempo de deslocamento casa-trabalho no Brasil (1992-2009): diferenças entre regiões metropolitanas, níveis de renda e sexo,"Technical report, Texto para discussão 1813, Instituto de Pesquisa Econômica Aplicada - IPEA.

Pereira, Rafael H. M., Warwar, Lucas, Parga, João, Bazzo, João, Braga, Carlos Kauê Vieira, Herszenhut, Daniel, & Saraiva, Marcus (2021). "Tendências e desigualdades da mobilidade urbana no Brasil I : o uso do transporte coletivo e individual," *Texto para Discussão IPEA*, 2673.

Rodrigues, Flavio, Silveira Neto, R, & Miranda, Filipe (2019). "Identification of Employment Subcenters in Brazilian Metropolitan Regions," preprint on webpage at https://www.anpec.org.br/encontro/2019/submissao/files_I/i10-5bbc4f56d19a715e4bcb780479f6a76a.pdf.

Ruijs, Nienke, & Oosterbeek, Hessel (2019). "School choice in Amsterdam: Which schools are chosen when school choice is free?" *Education Finance and Policy*, 14 (1): 1–30.

Trajkovski, Samantha, Zabel, Jeffrey, & Schwartz, Amy Ellen (2021). "Do school buses make school choice work?" *Regional Science and Urban Economics*, 86: p. 103607.

Tyndall, Justin (2021). "The local labour market effects of light rail transit," *Journal of Urban Economics*, 124: p. 103350.

World Bank (2020). *The Human Capital Index 2020 update: human capital in the time of COVID-19*, World Bank.

World Bank (2022). *Brazil Human Capital Review: Investing in People*, World Bank.

Zhou, Zhengyi, Chen, Hong, Han, Lu, & Zhang, Anming (2021). "The effect of a subway on house prices: Evidence from Shanghai," *Real Estate Economics*, 49 (S1): 199–234.

Zhu, Hongjia, Deng, Yongheng, Zhu, Rong, & He, Xiaobo (2016). "Fear of nuclear power? Evidence from Fukushima nuclear accident and land markets in China," *Regional Science and Urban Economics*, 60: 139–154.