The impact of congestion pricing policies on primary school students' performance: individual-level evidence from Milan's Area C

Léonard Moulin, Valeria Maria Urbano et Lorenzo Maraviglia

DOCUMENTS DE TRAVAIL 308

The impact of congestion pricing policies on primary school students' performance: individual-level evidence from Milan's Area C

Léonard Moulin, Valeria Maria Urbano, Lorenzo Maraviglia

May 2026

ined INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES https://doi.org/10.48756/ined-dt-308.0526

The impact of congestion pricing policies on primary school students' performance: individual-level evidence from Milan's Area C

Léonard Moulin*, Valeria Maria Urbano†, Lorenzo Maraviglia‡

Abstract

Congestion pricing policies have been implemented in several cities to reduce traffic congestion and mitigate environmental impacts in urban areas. However, the externalities of such policies may extend beyond traffic reduction, potentially generating indirect effects on health and, consequently, on children's educational outcomes. This study examines the impact of the congestion policy Area C introduced in Milan in 2012 on the academic performance of primary school pupils. Using a difference-in-differences design and individual-level data from academic years 2009/2010 to 2018/2019, we analyze students' outcomes across grades and subjects based on standardized tests from INVALSI. Our findings show statistically significant positive effects for second-grade students in both Mathematics and Italian, while no significant effects emerge for fifth-grade students. Moreover, the effects are heterogeneous across parental occupational backgrounds, with the largest gains observed among children from lower occupational backgrounds. Our results show that environmental regulation can generate meaningful equity-enhancing effects, narrowing early academic inequalities that mirror the socioeconomic structure of the city.

JEL codes: I24; Q53; Q58; R41.

Keywords: Low-emission zones; congestion policy; air pollution; student achievement; educational inequality; difference-in-differences.

*Institut national d'études démographiques (INED), F-93300 Aubervilliers, France. Université Paris-Est Créteil (Erudite). lenoard.moulin@ined.fr. Corresponding author.

†Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Via Raffaele Lambruschini, 20156, Milan, Italy. valeriamaria.urbano@polimi.it.

‡INVALSI, Via Ippolito Nievo, 35, 00153 Roma RM, Italy.

1 Introduction

Air pollution represents one of the most relevant environmental challenges in urban areas. A large body of evidence identifies vehicular traffic as one of the primary sources of air pollution in cities (Davis, 2008; Gallego et al., 2013), with detrimental effects on human health, cognitive functioning, and economic productivity (Chay and Greenstone, 2003; Zivin and Neidell, 2012; Chang et al., 2019). To address these challenges, several European cities have introduced low emission zones (LEZs). LEZs are designated urban areas in which access is restricted for vehicles that do not meet specific emission standards. Other restrictions implemented together with LEZs include road pricing systems, parking restrictions or dedicated lanes for low emission vehicles (Holman et al., 2015). A growing empirical literature shows that LEZs are effective in reducing traffic-related air pollution, generating significant environmental and health benefits (Pestel and Wozny, 2021; Margaryan, 2021; Lehe, 2019; Beria, 2016; Gibson and Carnovale, 2015). However, recent evidence on life satisfaction outcomes shows that LEZs led to a significant decline in self-reported well-being among individuals residing within the restricted areas, suggesting that the overall welfare effects of such policies may extend beyond improvements in environmental quality (Sarmiento et al., 2023). Given the wide spectrum of potential benefits and costs, it is crucial to explore the broader externalities associated with LEZs.

To this end, an emerging body of research has focused on the effects of these policies on children, highlighting their implications for health and the indirect consequences for academic achievement. Empirical studies demonstrated that air pollution has detrimental effects on children's health and development (Stafford, 2015; Klauber et al., 2024; Park et al., 2002; Mölter et al., 2014; Khreis et al., 2017; Orellano et al., 2017; Anenberg et al., 2022). Compared to adults, children are particularly susceptible to environmental hazards, as their respiratory and immune systems are still developing (Etzel, 2020). At the same time, ongoing urbanization has increased the number of children exposed to traffic-related pollution, as a growing share of young cohorts live and attend school in densely populated urban centers (Bishop and Corkery, 2017) where vehicular traffic is a major source of air pollution (Raysoni et al., 2011; Bennett et al., 2019; Kumar et al., 2017). The exposure of children to air pollutants may have effect on working memory and processing speed (Alvarez-Pedrerol et al., 2017), as well as on attention (Sunyer et al., 2017) and general cognitive development (Milojevic et al., 2021). Recent research has, therefore, begun to explore not only the direct health consequences of LEZs for children health (Simeonova et al., 2021), but also their broader effects on educational achievement.

Conte Keivabu and Rüttenauer (2022) explored the impact of stricter traffic regulations introduced in Central London in late 2015 and documented a substantial decline in pollution levels. While no overall improvement in attendance was detected across all schools, the analysis revealed significant reductions in absences among schools with a high share of students with low socio-economic status. Avila-Uribe et al. (2024) analyzed the introduction of the Ultra Low Emission Zone (ULEZ) in London in 2019 showing that improved air quality around schools led to an average increase of 0.09 standard deviations in standardized test scores, with larger gains in initially low-performing schools. Exploiting the staggered introduction of LEZs in North Rhine-Westphalia since 2008, Brehm et al. (2025) found that LEZ implementation increased transition rates to the academic track in secondary education by about one percentage point (2.4 percent). Similarly, Valdés et al. (2025) documented that the LEZ introduced in Madrid at the end of 2018 significantly improved air quality and generated a 0.20 standard deviation increase in university entrance exam performance among high schools located within the restricted area. Despite providing consistent preliminary evidence across different European contexts on the role of LEZs in shaping children's academic achievement, these studies rely on aggregated school-level outcomes, assessing how improvements in air quality translate into changes in average school performance. This limits the precision of estimates, prevents direct control for compositional changes in school enrollment over time, and does not allow for the analysis of how effects are distributed across students with different characteristics.

This paper examines the educational effects of Milan's Area C congestion policy using individual-level student data spanning academic years 2009/10 to 2018/19. Milan provides a particularly relevant setting for studying the broader externalities associated with the introduction of the LEZ. The city is Italy's financial and economic capital and one of the most polluted urban areas in Western Europe, partly because of its location in the Po Valley, where geographical and meteorological conditions favor the accumulation of pollutants. Concentrations of particulate matter and nitrogen dioxide have historically and repeatedly exceeded European regulatory thresholds with Milan ranking 5th among 1,000 European cities in terms of NO_2-attributable mortality and an annual average NO_2 concentration of 38 $\mu g/m^3$ (Nieuwenhuijsen et al., 2024), making air quality a persistent policy concern.

Introduced on January 16, 2012, Area C is a congestion policy implemented in Milan's city center. It combines access restrictions for the most polluting vehicles with a flat daily charge for vehicles allowed to enter, with the explicit objective of reducing both traffic congestion and atmospheric pollution (Beria, 2016). While the literature has generally treated LEZs and area-based congestion pricing as distinct policy instruments, the substantial overlap in the behavioural responses targeted by both policies suggests that they can be meaningfully analysed together (Axsen and Wolinetz, 2021). Given the hybrid design and dual objectives of Area C, and the significant convergence between these two policy approaches, in this paper we refer to Area C both as a congestion pricing scheme and as a LEZ.

The policy generated a sharp and well-documented reduction in traffic and air pollution (Lehe, 2019; Beria, 2016; Gibson and Carnovale, 2015; Percoco, 2013, 2014), exposing schools located within the charging zone to an abrupt improvement in their local environment. Crucially, the timing of the reform coincides with the availability of individual-level INVALSI data, allowing us to compare the evolution of test scores in treated and control schools before and after the policy. Although the pre-treatment window is limited to two academic years, the seven post-treatment years allow us to trace the dynamic evolution of effects and to assess whether improvements in environmental conditions translate into persistent changes in academic performance.

The contribution of this study to the literature on the educational achievement of LEZs is threefold. First, by examining the case of Milan's Area C, it extends the existing evidence to a new institutional and geographical context. Previous studies have focused on London, Madrid, and Germany. Providing evidence from Milan helps assess the validity of prior findings and strengthens the robustness of the emerging empirical literature across distinct policy designs and contexts. Despite previous studies have cautioned against drawing direct empirical comparisons between LEZ systems due to institutional and contextual differences (Veitch and Rhodes, 2024; Rotaris et al., 2010), documenting consistent effects across diverse settings is essential to reinforce the evidence on the broader societal benefits of LEZs. Existing studies on the effects of LEZs on educational achievement have focused on London, Madrid, and Germany. The LEZ systems in Madrid and Germany are primarily based on vehicle bans applied in city centers, targeting high-emission vehicles in Germany and transit vehicles (i.e. non resident vehicles whose origin and destination was not in the restricted area, with exemptions for electric and hybrid vehicles) in Madrid. In contrast, both Milan and London rely on policies that target vehicle emissions and traffic congestion, although they do so through different institutional designs. Milan's Area C differs from the London scheme in that it operates as a cordon system, requiring a toll upon entering the restricted area, whereas London adopts a zonal system, where charges also apply to vehicles circulating within the designated boundaries (De Palma and Lindsey, 2011; Veitch and Rhodes, 2024). These differences across LEZ systems make the analysis of a distinct policy context, such as Milan's Area C, valuable, as it helps strengthen the robustness of the emerging empirical literature across different policy designs and facilitates comparisons across systems.

Second, this study provides the first individual level analysis of the impact of a LEZ on academic achievement at the primary school level. Moving from aggregated school level outcomes to individual level data enables a direct assessment of how congestion reduction policies affect the distribution of learning outcomes across students, free from the compositional confounders that can bias school level estimates (Conte Keivabu and Rüttenauer, 2022; Valdés et al., 2025). Moreover, granular data allow to characterize treatment effect heterogeneity along socioeconomic, demographic, and educational dimensions, revealing which children benefit most from improvements in urban air quality.

Third, the paper provides a systematic comparison of effects across two grades within primary school (Grade 2 and Grade 5) and across two subjects (Mathematics and Italian). This design enables an examination of whether the benefits associated with improved air quality differ by developmental stage and by the cognitive demands of different academic disciplines. To the best of our knowledge, this is the first study to provide such a structured within-primary-school comparison, offering a more nuanced understanding of how LEZs may differentially influence academic outcomes across developmental stages and learning domains.

The remainder of the paper is organized as follows. Section 2 provides the institutional background, describing the main features of the Area C policy and its effects on air pollution. Section 3 introduces the INVALSI dataset, detailing the data description, sample selection, and descriptive statistics. Section 4 outlines the empirical strategy and methods. Section 5 presents the main results, including heterogeneity analysis, and discusses them in light of the existing literature. Section 6 reports a series of robustness checks. Section 7 explores potential mechanisms underlying the effects of LEZ implementation on students' academic performance. Section 8 concludes the paper.

2 Area C and the impact on air pollution

The Area C charging zone was introduced in Milan on January 16, 2012, following a public consultation in 2011 that showed strong support for the scheme, with 79% of voters endorsing the congestion pricing program (Percoco, 2013). The policy combines access restriction for the most polluting vehicles with a flat daily charge for those allowed to enter, differentiated by category (general, resident, or commercial) (Hensher and Li, 2013; Lehe, 2019). Exemptions for certain vehicle types, such as motorcycles, along with restrictions on others, helped improve perceived fairness and contributed to public acceptance (Veitch and Rhodes, 2024).

Area C charging zone consists of the *Cerchia dei Bastioni* area, covering an 8 km 2 area in central Milan inhabited by more than 99,000 residents as of the 12/31/2019. The cordon includes 43 access points equipped with cameras to capture vehicle number plates upon entry, with access limitations and charges requiring daily payments. Charging hours are from 7:30 AM to 7:30 PM on weekdays.

Area C has been shown to reduce both traffic and air pollution. Building on a technical report published by the Municipality of Milan (Bedogni, 2013), Beria (2016) documented a sharp and persistent decline in emissions in 2012. To attribute these reductions to Area C, the author compared pollutant levels across the full day and during the charging hours, as well as across different pollutants (PM_{10}, NH_3, NO_x, NO_2, NMVOC, and CO_2). He found larger reductions during charging periods and for pollutants primarily associated with vehicle types most affected by the policy. Exploiting a temporary suspension of the policy as a natural experiment, Gibson and Carnovale (2015) applied causal inference methods to traffic volumes and pollution levels (CO_2, PM_{10}, and $PM_{2.5}$), finding significant increases in CO and PM_{10} in the range of 6–17%.

We present descriptive evidence through a longitudinal analysis of pollutant concentrations in the city center to document these effects. Data on air pollution level across years is provided as open data by the air pollution monitoring system of ARPA Lombardia (Regional Environmental Protection Agency of Lombardy), and sourced from a station located within Area C. The analysis covers the period 2009–2019 in order to match the availability of INVALSI data, which span academic years 2009/2010 to 2018/2019. This time window ensures that all calendar years included in the corresponding academic years are fully covered. The analysis focuses on particulate matter (PM_{10}) and nitrogen dioxide (NO_2), two pollutants commonly associated with vehicular traffic (Health Effects Institute, 2010) and widely used in previous studies on Area C (Beria, 2016; Gibson and Carnovale, 2015; Percoco, 2013) as well as in evaluations of congestion policies in other cities (Salas et al., 2021; Lebrusán and Toutouh, 2021; Galdon-Sanchez et al., 2023). NO_2 concentrations are observed at the hourly level, while PM_10 concentrations are available on a daily basis.

Tables 1 and 2 report differences in the annual mean concentrations of the two pollutants compared to the pre-treatment period (2009–2011), when the congestion policy had not yet been implemented. For the pollutant NO_2, the sample is restricted to weekdays between 7:00 and 20:00 in order to reflect the operating hours of Area C (Monday–Friday, 7:30–19:30). While for pollutant PM_10, as ARPA provides PM_{10} measurements at the daily level, we restrict the sample to weekdays in order to align the analysis with the operating days of Area C (Monday–Friday). The estimated differences on both pollutants are negative and statistically significant for all years considered, confirming previous empirical findings that air pollutant levels declined over time following the policy implementations (Beria, 2016; Gibson and Carnovale, 2015; Percoco, 2013).

3 Data and descriptive statistics

3.1 Data

INVALSI data are provided by the National Institute for the Evaluation of the Educational System, a government agency under the Ministry of Education responsible for quality assurance in primary and secondary education. Starting from the 2009/2010 school year, second- and fifth-grade students — the two primary education cohorts covered by INVALSI — have taken standardized national tests on an annual basis. Since the assessments are administered between May and June, throughout the paper we refer to each school year using the second calendar year (e.g., 2009/2010 is referred to as 2010). Participation of schools in these national tests is compulsory by law, which ensures full participation and limits selection bias in test-taking. The tests cover both Mathematics and Italian language and consist of a combination of closed questions (with a choice of predefined answers) and open questions. Test results distinguish between outcomes that refer to Italian language, including reading comprehension and grammatical knowledge, and Mathematics, including mathematical knowledge and mathematical reasoning. Scores used in the analysis consist of the percentage of correct answers and are standardized to have zero mean and unit variance within each school year.

INVALSI data also provide individual-level information on children's and parental characteristics including gender, origin, employment status of the parents, regularity and attendance to pre-primary schools (i.e. nursery and kindergarten). This additional information is collected through a family questionnaire sent to each family, a student questionnaire compiled by students and administrative records on students' characteristics provided by school staff.

The INVALSI dataset initially provided information on 495 primary schools located in the Milan Metropolitan Area. The geographic coordinates (latitude and longitude) of the school attended at the time of the assessment were assigned to each student.

The sample selection process consisted of two main steps. In the first step, we excluded schools with inaccurate or missing georeferencing information. Without precise geolocation data, it is not possible to assign schools to the treatment group (Area C) or the control group (Metropolitan Area), and therefore these observations cannot be used in the empirical analysis. This restriction reduces the sample to 395 schools, with 18 schools located in Area C and 377 schools located in the Milan Metropolitan Area. Information on school geolocation was not available in the original INVALSI database and was specifically constructed for the purposes of this study. In the Italian education system, schools are uniquely identified by administrative codes. However, these codes are not stable over time and may change following administrative modifications such as school closures, mergers, or divisions. This instability required the georeferencing procedure to be repeated at different points in time in order to minimize the number of schools lacking geographic information in specific years. Due to confidentiality and disclosure protection policies, detailed geolocation information is not publicly released in the INVALSI dataset, as it could allow the identification of individual schools. Instead, schools are classified according to broad territorial categories (e.g., "Area C", "Area B", and others). As a result, the analysis does not exploit exact spatial distances from Area C's area, but rather relies on this categorical classification.

In the second step, we excluded primary schools with no pre-treatment observations. Since our identification strategy relies on a difference-in-differences design, the availability of pre-policy data is essential to establish baseline outcomes and compare trends between treated and control schools prior to the intervention. This leads to a final analytical sample of 329 schools, including 17 in Area C and 312 in the Milan Metropolitan Area.

The spatial distribution of all primary schools initially identified in the Milan Metropolitan Area is displayed in Figure 1. Schools located within the Area C perimeter are represented by dark red dots, while schools in the wider Milan Metropolitan Area are shown in grey. The map also displays the boundaries of Area B, the broader LEZ that encompasses Area C, introduced in February 2019, in the final months of our study period.[1] Municipal administrative boundaries are shown in thin lines to provide geographic context.

3.2 Descriptive statistics

A comprehensive description of the covariates included in the dataset is provided in Table 3, which reports their distribution before and after the implementation of Area C for the treatment group (Area C) and the control group (Metropolitan Area). Overall, the table shows that student composition remains broadly stable over time in both areas, with some moderate changes along specific dimensions, particularly within Area C schools. Gender composition remains broadly balanced in both groups. In Area C, the share of students recorded as female decreases slightly from 51.0 percent before the policy to 48.8 percent after the policy, while it remains almost unchanged in the metropolitan area, from 48.7 to 48.9 percent. This pattern mainly reflects a small increase in missing information rather than a meaningful change in the gender composition of students.

Some differences are also observed along socio-economic and migration-background dimensions, although the overall patterns remain relatively stable. In Area C schools, the share of students with fathers employed as executives, senior officials, university teachers, or officers decreases from 21.0 to 19.4 percent, while the corresponding share in the metropolitan area changes from 5.8 to 5.1 percent. The share of fathers in professional occupations also declines slightly in Area C, from 28.6 to 26.0 percent, whereas it increases from 11.6 to 12.7 percent in the metropolitan area. For mothers, the share in executive or senior occupations remains broadly stable in Area C, moving from 8.0 to 8.6 percent, and remains much lower in the metropolitan area, increasing from 1.6 to 1.8 percent. These descriptive differences indicate that schools located inside Area C serve, on average, a more socioeconomically advantaged population than schools in the wider metropolitan area. Some differences in the direction of compositional changes are also observed across the two groups over time. These patterns motivate the inclusion of student-level controls in the empirical specifications, which account for observed changes in student composition over time

The table also reports the distribution of students by migration background. In Area C schools, the share of students recorded as being of Italian origin decreases from 92.7 to 88.1 percent after the introduction of the policy, while the shares of second-

[1] Given the overlap between the introduction of Area B (February 2019) and the final year of our study period, schools in the control group may also have been affected. As discussed in Section 6, this does not affect our results.

generation foreign-origin students change only marginally, from 5.4 to 5.9 percent. This pattern likely reflects, at least in part, an increase in missing information on migration background rather than a substantive change in the ethnic composition of Area C schools, as the share of students with missing origin data rises from 0.4 to 4.5 percent over the same period. A comparable pattern is observed in the metropolitan area, where the share of students recorded as being of Italian origin decreases from 85.2 to 80.5 percent, with missing data on migration background also increasing over the period. In contrast to Area C, however, this decline is accompanied by a rise in the share of second-generation foreign-origin students, from 8.9 to 12.3 percent, likely reflecting broader demographic trends in the wider urban area. Overall, these patterns suggest relatively limited changes in migration-background composition within Area C schools, while changes are somewhat more visible in the wider metropolitan area.

A comparable pattern is observed in the metropolitan area, where the share of students recorded as being of Italian origin decreases from 85.2 to 80.5 percent, with missing data on migration background also increasing over the period. In contrast to Area C, however, this decline is accompanied by a rise in the share of second-generation foreign-origin students, from 8.9 to 12.3 percent, likely reflecting broader demographic trends in the wider urban area.

Early childhood education variables display somewhat larger differences over time, although these patterns should be interpreted with caution given changes in missing information and reporting. In Area C schools, the share of students reported as having attended nursery increases from 32.3 to 43.0 percent, while it rises from 24.6 to 30.5 percent in the metropolitan area. These increases are at least partly mechanical, as the share of students with missing nursery information drops substantially over the period, from 37.6 to 15.2 percent in Area C and from 36.3 to 33.0 percent in the metropolitan area, suggesting that improved reporting partly drives the observed rise. By contrast, reported kindergarten attendance decreases in both areas, from 75.6 to 67.9 percent in Area C and from 87.3 to 72.9 percent in the metropolitan area. This decline is likely to reflect, at least in part, changes in missing information or reporting patterns rather than a substantive decrease in kindergarten attendance, as evidenced by the share of students with missing kindergarten data moving in opposite directions across the two groups, from 23.5 to 13.1 percent in Area C and from 10.7 to 21.0 percent in the metropolitan area. Finally, school progression remains highly regular in both groups: the share of students classified as regular or early entrants is 98.7 percent before and 96.8 percent after the policy in Area C, and remains stable at 96.9 percent in the metropolitan area. The slight decline in Area C is largely accounted for by an increase in missing information on track regularity, from 0.0 to 2.2 percent over the period.

Overall, while some compositional changes are observed over time, they appear largely driven by shifts in reporting patterns rather than changes in student characteristics, and are accounted for by the inclusion of individual-level controls in all empirical specifications.

The descriptive statistics show that schools located inside Area C differ from those in the wider metropolitan area in their initial student composition, particularly along socio-economic dimensions. At the same time, student characteristics remain broadly stable over time, with only moderate changes along specific dimensions and in both groups. These patterns are important for the empirical analysis for two reasons. First, they motivate the inclusion of student-level controls, including parental occupation, gender, migration background, early childhood education attendance, and grade regularity. Second, the availability of school identifiers allows us to include school fixed effects, so that identification relies on changes within schools over time rather than on permanent differences between schools. Year fixed effects further account for shocks common to all schools in a given year.

Table 4 complements the previous descriptive evidence by reporting average standardized test scores before and after the implementation of Area C, separately by grade and subject. In Area C schools, average scores increase after the introduction of the policy in all grade-subject combinations. The increases are particularly marked and statistically significant in Grade 2: Mathematics scores rise from 0.064 to 0.305 s.d. and Italian scores from 0.213 to 0.321 s.d., with p-values below 0.001 in both cases. In Grade 5, Mathematics scores also increase significantly, from 0.133 to 0.249 s.d. The increase in Grade 5 Italian is smaller, from 0.279 to 0.325 s.d., and only marginally significant.

By contrast, average test scores in the wider metropolitan area remain very close to zero before and after the policy. In Grade 2, Mathematics scores move from -0.002 to -0.012 s.d., with a p-value of 0.078, while Italian scores change from -0.007 to -0.012 s.d. and are not statistically different across periods. Similar patterns are observed in Grade 5, where Mathematics scores change from -0.004 to -0.010 s.d. and Italian scores from -0.009 to -0.013 s.d., with no statistically significant before-after differences. Except for Grade 2 Mathematics, where the before-after difference is marginally significant, these changes are not statistically significant.

These raw comparisons indicate larger before-after increases in Area C schools than in the wider metropolitan area, especially in Grade 2. However, they remain purely descriptive and should not be interpreted as causal evidence, since they do not account for permanent differences between schools, common year shocks, or changes in student composition.

4 Empirical framework

We estimate the impact of the introduction of Area C on student achievement using a two-way fixed effects (TWFE) difference-in-differences framework, exploiting variation across schools and over time.

$$Y_{ist} = \alpha + \beta\,(AreaC_s \times Post_t) + \mathbf{X}'_{ist}\gamma + \mu_s + \lambda_t + \varepsilon_{ist} \tag{1}$$

where Y_{ist} denotes the standardized INVALSI test score of student i enrolled in school s in year t. The treatment indicator $AreaC_s$ equals one for schools located within the Area C zone, while $Post_t$ is an indicator equal to one for years following the policy implementation in 2012. The coefficient of interest, β, captures the average effect of the policy on student performance. The vector $\mathbf{X}_{ist}$ includes a set of student-level covariates, such as parental occupation, gender, migration background, early childhood education attendance, and indicators for grade retention or irregular school trajectories. These controls account for compositional changes in the student population over time. We include school fixed effects μ_s and year fixed effects λ_t. School fixed effects control for time-invariant school characteristics, such as school size, location, and level of competition (Hanushek, 1986, 1998; Agasisti and Murtinu, 2012). Standard errors are clustered at the school level.

Identification relies on the assumption that, absent the introduction of Area C, treated and control schools would have followed parallel trends in test scores. Given the localized nature of the policy, this assumption requires that schools just inside and outside the Area C boundary evolve similarly in the absence of treatment. Figure 2 provides graphical evidence on this assumption by plotting the evolution of average standardized test scores for treated and control schools, pooling across grades and subjects. As shown in the left panel, pre-treatment trends appear broadly similar across groups, with no clear evidence of systematic divergence prior to the policy introduction. This pattern is consistent with the fitted linear trends displayed in the right panel, which indicate comparable pre-policy trajectories between treated and control schools. We complement this visual evidence with a formal test of parallel trends. A test for differential linear trends between treated and control schools in the pre-treatment period yields an F-statistic of 0.01 (p-value = 0.91), failing to reject the null hypothesis of parallel trends. These results provide support for the identifying assumptions underlying the TWFE specification, although they cannot fully rule out more subtle differences in underlying trends.

To further address potential concerns about differential pre-treatment dynamics, we complement the baseline DiD estimates with the Synthetic Difference-in-Differences (SDID) estimator proposed by Arkhangelsky et al. (2021). This approach combines features of difference-in-differences and synthetic control methods by constructing weights over control units and pre-treatment periods. Intuitively, the SDID procedure reweights control schools so that their weighted average better reproduces the pre-treatment trajectory of treated schools. In this setting, the SDID estimator provides a flexible way to construct a counterfactual path for treated schools that is more closely aligned with their observed pre-policy evolution. This is particularly relevant in contexts where simple DiD estimates may be sensitive to small differences in trends across groups.

We next examine the dynamic effects of the policy using an event-study specification of the DiD model:

$$Y_{ist} = \alpha + \sum_{\substack{k=-2 \\ k\neq -1}}^{7} \beta_k (AreaC_s \times D_t^k) + \mathbf{X}'_{ist}\gamma + \mu_s + \lambda_t + \varepsilon_{ist} \tag{2}$$

where $D_t^k = \mathbb{1}_{\{t-2012=k\}}$ indicates the number of years relative to the introduction of Area C. The year immediately preceding the policy ($k = -1$) is omitted and serves as the reference period. The coefficients β_k trace the evolution of treatment effects over time. Estimates for $k < 0$ provide a direct test of the parallel trends assumption, as they capture differences between treated and control schools prior to the policy. Estimates for $k \geq 0$ capture the evolution of treatment effects following the introduction of Area C.

Finally, we explore whether the impact of Area C varies across student characteristics by estimating the following interaction model:

$$Y_{ist} = \alpha + \beta DID_{st} + \delta(DID_{st} \times Z_{ist}) + \theta Z_{ist} + \mathbf{W}'_{ist}\gamma + \mu_s + \lambda_t + \varepsilon_{ist} \tag{3}$$

where $DID_{st} = AreaC_s \times Post_t$, Z_{ist} denotes a student-level characteristic of interest, and $\mathbf{W}_{ist}$ denotes the set of controls other than Z_{ist}. We consider heterogeneity along several dimensions, including gender, parental occupation, migration background, early childhood education, and indicators of school progression. The coefficient δ captures differential treatment effects across groups and can be interpreted as the difference in treatment effects between students with $Z_{ist} = 1$ and those with $Z_{ist} = 0$, the latter being the reference group.

5 Results

5.1 Baseline difference-in-differences estimates

Table 5 reports the baseline DID coefficients, providing the estimated impact of Area C on student achievement across grades and subjects. Across the full sample, the DID estimates reveal positive and statistically significant effects for second-grade pupils, while the coefficients for fifth-grade students are consistently close to zero and not statistically significant. For Grade 2, exposure to Area C improves both Mathematics and Italian performance with significant improvements in test performance: Mathematics scores increase by 0.256 standard deviations (s.d.) and Italian scores by 0.116 s.d. The higher effect on Mathematics is consistent with previous studies on air pollution effects on student academic performances (Bernardi and Keivabu, 2024; Amanzadeh et al., 2020; Duque and

Gilraine, 2022). The subject-specific differences can be explained by different cognitive tasks associated with different subjects (Smith and Stein, 1998; Tekkumru-Kisa et al., 2015). Analytical subjects, such as mathematics, typically require problem-solving and higher-order reasoning, whereas language-based subjects often rely more on memorization and lower cognitive effort. Since high-demand tasks require more complex cognitive processing, air pollution is theoretically expected to exert a stronger negative effect on performance in these domains. In terms of magnitude, the effect on Mathematics scores is approximately twice as large as that on Italian scores. This pattern is consistent with prior studies providing causal evidence on the impact of pollution on cognitive outcomes, which similarly found effects on English scores that are roughly half the size of those observed for Mathematics (Duque and Gilraine, 2022).

In contrast, the coefficients for Grade 5 in Mathematics and Italian are smaller in magnitude and not statistically significant. The pattern of sizeable effects in Grade 2 but no detectable impact in Grade 5 is consistent with the idea that air pollution reductions have greater impacts on younger children. Previous studies suggested that children are particularly vulnerable to air pollution exposure due to the greater malleability of cognitive development at younger ages (Persico and Venator, 2021). The lack of significant effects for Grade 5 students may indicate that, below a certain threshold of pollution reduction, improvements are insufficient to generate measurable academic gains among older students.

The magnitude of our estimates is within the range documented in the existing literature on environmental regulation and educational outcomes. For instance, Valdés et al. (2025), examining the introduction of a LEZ in Madrid and its impact on high-stakes university admission exams, found a significant increase of 0.20 s.d. in average test scores. Similarly, Avila-Uribe et al. (2024), in their study on the effects of London's LEZ on elementary school performance, reported a significant improvement of 0.09 s.d. in test scores. In comparison, the estimated effect is smaller in magnitude than the impact on INVALSI test scores documented in studies on early childcare interventions (Corazzini et al., 2021) and the negative effects of COVID-19 school closures on academic performance (Contini et al., 2025). However, it is comparable to findings from research examining the effects of extended instructional time on INVALSI test scores (Battistin and Meroni, 2016).

5.2 Synthetic difference-in-differences estimates

To strengthen the credibility of these results, we also estimate the effects using the SDID estimator. SDID complements the baseline DID framework by constructing unit and time weights that improve the alignment between treated schools and the control group in the pre-treatment period. By explicitly targeting balance before the policy, this approach provides a more credible counterfactual and allows us to assess whether the baseline results are robust to an alternative identification strategy. Figure 3 illustrates how the reweighting procedure affects the alignment of pre-treatment trajectories. The figure plots the evolution of standardized test scores for treated schools and for the reweighted control group. Across subjects and grades, the SDID weights substantially reduce pre-existing differences between Area C schools and those in the wider metropolitan area, yielding a closer match prior to the introduction of the policy. While some discrepancies remain, the improved pre-treatment fit suggests that the SDID estimator provides a more credible basis for constructing the counterfactual evolution of treated schools. Figure 4 provides additional insight into the construction of the SDID estimator by plotting, for each control school, the adjusted outcome difference relative to the treated group, with marker size reflecting the corresponding unit weight. The figure reveals substantial heterogeneity across control schools, with both positive and negative adjusted differences. The SDID estimate is given by the weighted average of these school-level differences, with the more heavily weighted units contributing more to the overall effect.

The resulting estimates, presented in Table 6, are broadly consistent with the baseline findings, although reduced in magnitude. For Grade 2 Mathematics, the SDID coefficient remains positive and statistically significant, with an estimated effect of 0.219 s.d. This estimate is slightly smaller than the baseline DID effect but remains sizeable relative to the literature. By contrast, the Grade 2 Italian estimate becomes small and not statistically significant, indicating weaker support for an effect in this subject once pre-treatment balance is enforced more tightly. For Grade 5, SDID consistently yields no detectable impact: point estimates are close to zero in both subjects. The absence of effects in Grade 5 therefore appears robust across identification strategies. Overall, the SDID results reinforce the conclusion that the introduction of Area C generated significant improvements in early academic performance, with the greater and most robust gains concentrated in Grade 2 Mathematics.

5.3 Dynamic effects: event-study estimates

To examine more finely the evolution of effects over time, we estimate an event-study model in which the impact of Area C is captured by a series of year-specific indicators relative to the year preceding the introduction of the congestion policy (2011). The coefficients plotted in Figure 5 represent differences in standardized test scores between schools located inside Area C and schools located in the metropolitan area, for each year relative to 2011, controlling for school and year fixed effects as well as the same set of individual covariates used in the baseline DID specifications. The pre-treatment coefficients do not reveal any clear differential trend prior to the introduction of the policy. For 2010, the estimate is close to zero across all four specifications and never statistically significant. This absence of a marked gap between treated and untreated schools in the only available pre-policy year is consistent with the parallel-trends assumption underlying our identification strategy, while also highlighting that the pre-treatment window is short.

Following the implementation of the policy, the dynamic patterns for Grade 2 broadly align with the average effects estimated using DID. In Mathematics, all post-treatment coefficients are positive from 2013 onward and range between roughly 0.10 and 0.37 s.d. Effects in 2014 and 2016 stand out as relatively large (about 0.28 and 0.37 s.d., respectively), becoming significant. In Italian, the dynamic profile is similar but somewhat smaller in magnitude: coefficients are positive throughout most of the post-treatment period and become statistically significant at the 10% level in the middle years (in 2014 and 2016), with effects reaching approximately 0.19–0.29 s.d. Overall, while the year-by-year trajectory remains irregular, several post-treatment effects in Grade 2 are statistically significant. This reinforces the interpretation of a positive and sustained impact of Area C on early primary outcomes. For Grade 5, the post-treatment coefficients are also mostly positive but remain smaller in magnitude and less precisely estimated. In both Mathematics and Italian, effects generally lie between 0.10 and 0.20 s.d., with no clear or persistent pattern. The event-study does not provide strong additional evidence of an effect of Area C on upper-primary outcomes, in line with the DID and SDID estimates.

These patterns are consistent with the evolution of pollution levels following the introduction of Area C. As shown in Tables 1 and 2, the largest and most statistically significant reductions in both NO_2 and PM_{10} concentrations occur in the middle of the post-treatment period, particularly in 2014 and 2016. Consistently, these are precisely the years in which we also observe the strongest and most robust effects on student achievement in Grade 2. Later years such as 2018 and 2019, where continued declines in pollution is observed,are more distant from the policy implementation and do not coincide with similarly pronounced educational effects. This temporal alignment suggests that the gains in student performance emerge most clearly when reductions in air pollution are both substantial and statistically significant relative to the pre-treatment period, providing further evidence of a link between improved environmental conditions and early academic outcomes.

5.4 Heterogeneity

Figure 6 examines whether the impact of Area C varies across student characteristics. A clear pattern of heterogeneous effects emerges.

The most pronounced differences arise along socio-economic dimensions. In particular, students whose parents hold executive occupations exhibit negative estimated effects. This contrasts with the positive average treatment effects documented in the baseline results, suggesting relatively more favorable impacts for students from less advantaged backgrounds. These findings point to a clear socio-economic gradient in treatment effects. Heterogeneity also emerges along dimensions related to early educational attendance. Students who attended kindergarten tend to display negative or smaller estimated effects, whereas those without kindergarten attendance exhibit more favorable outcomes. These patterns suggest that the policy may partially offset differences in early educational environments, by generating relatively stronger gains among students with less favorable initial conditions. Differences by gender are negligible, and heterogeneity along migration background is less systematic across subjects, with some estimates being imprecise. Overall, these results indicate that the distribution of gains is primarily structured along socio-economic background and early educational investments, rather than along broader demographic characteristics.

The heterogeneity results suggest that the effects of Area C are unevenly distributed across students, but tend to reduce inequalities in academic outcomes. While some groups, such as students from more advantaged backgrounds, exhibit negative estimated effects, the overall positive average impact implies relatively larger gains among students with less favorable initial conditions. This pattern is consistent with a reduction in early achievement gaps.

The results of the heterogeneity analysis are consistent with prior studies documenting differences across students from diverse socio-economic backgrounds, often using school-level data and alternative proxies for socio-economic status. Brehm et al. (2025), examining the impact of LEZs on children's educational outcomes using school-level data from one German state, found that estimated effects are slightly stronger in socio-economically disadvantaged areas. In their heterogeneity analysis, socio-economic characteristics are measured at the neighborhood level, showing that areas with below-median purchasing power and above-median shares of unemployed and foreign-born residents experience significantly larger effects. Avila-Uribe et al. (2024), in their study on the effects of London's LEZ on elementary school performance, use eligibility for free school meals as a proxy for household income. Their heterogeneity analysis revealed a more pronounced increase in test scores among schools with higher eligibility rates, underscoring the potential of such policies to reduce educational disparities.

6 Robustness

Tables 7 to 10 assess whether the main estimates are sensitive to a broad range of alternative specifications, sample definitions, and inference methods, such as the use of class rather than school fixed effects, balanced-panel estimates, multiple imputation for missing covariates, exclusion of schools in the outbound zone, alternative standardization of the outcome, school-specific linear trends, matched DID estimation, and wild-cluster-bootstrap inference. Overall, the robustness analysis reinforces the central conclusion of the paper: the positive effects of Area C are concentrated in Grade 2, especially in

Mathematics, whereas estimated effects in Grade 5 are smaller and stastically weaker.

The baseline specification, reported in column (1), includes school and year fixed effects and no additional controls. The estimated effect for Grade 2 Mathematics is large and significant, at 0.253 s.d., while the corresponding estimate for Grade 2 Italian is smaller (0.112 s.d.) but remains positive and statistically significant. These estimates are very close to those obtained in the preferred specification reported in the main analysis (column (2)), which additionally includes student-level covariates. This similarity indicates that the results are not driven by observable differences in student composition. By contrast, the Grade 5 estimates are positive but not significant, consistent with the absence of effects in subsequent grades.

When replacing school fixed effects with class fixed effects (column (3)), the estimates for Grade 2 remain positive and statistically significant in both subjects, with magnitudes very similar to previous specifications (0.249 s.d. in Mathematics and 0.114 s.d. in Italian). This suggests that the results are robust to a more granular control for unobserved heterogeneity at the class level. In contrast, the Grade 5 coefficients remain positive but not statistically significant.

Inference based on wild-cluster bootstrap with 1,000 replications (column (4)) yields results that are consistent with those obtained under conventional clustering. Such bootstrap procedures are commonly used to assess the robustness of inference with clustered standard errors, especially when the number of clusters may raise concerns (Cameron et al., 2008).The estimated effects for Grade 2 remain statistically significant in both Mathematics (0.256 s.d.) and Italian (0.116 s.d.), alleviating concerns about the reliability of conventional clustered standard errors in settings with a limited number of clusters. For Grade 5, estimates remain not statistically significant.

Restricting the sample to a balanced panel of schools observed in all years (column (5)) leaves the Mathematics results largely unaffected, with a sizeable and statistically significant effect in Grade 2 (0.252). The Italian estimate, however, becomes smaller and is no longer statistically significant, suggesting some sensitivity to sample composition. For Grade 5, estimates remain small and not statistically significant.

Addressing missing data through multiple imputation with 20 imputations (column (6)) produces estimates that are very similar to those obtained in earlier specifications. Multiple imputation by chained equations is a standard approach for handling missing covariate information in applied settings (White et al., 2011). The Grade 2 effects remain positive and statistically significant in both Mathematics (0.265 s.d.) and Italian (0.132 s.d.), indicating that missing covariates are unlikely to drive the results. As before, Grade 5 estimates remain imprecise and not statistically significant.

Excluding schools located in the outbound zone (column (7)) restricts the comparison to schools within Area C and the rest of the city of Milan, thereby focusing on a more comparable set of treated and control units. This leads to a modest attenuation of the

Grade 2 estimates: the effect in Mathematics remains positive and statistically significant (0.215 s.d.), whereas the Italian estimate becomes smaller and is no longer statistically significant. This pattern suggests that the estimates on Italian test results are statistically weaker and more sensitive to sample composition, while the effect on Mathematics remains robust. For Grade 5, estimates continue to be not statistically significant.

Excluding schools located in the area immediately surrounding Area C provides an additional robustness check aimed at limiting potential spillover contamination among nearby control schools (column (8)). The results remain fully consistent with the main findings. For Grade 2, the estimated effects are slightly larger than in the baseline specification and remain statistically significant in both subjects, reaching 0.282 s.d. in Mathematics and 0.135 s.d. in Italian. This increase is consistent with the possibility that including nearby schools in the control group attenuates the estimated effect. By contrast, for Grade 5, the coefficients remain positive but not statistically significant in both Mathematics and Italian.

Using a de-standardized outcome measure—constructed by dropping the year dimension from the grade–subject (column (9)) does not affect the results. The estimated effects for Grade 2 remain positive and statistically significant in both subjects, with magnitudes comparable to the baseline (0.276 s.d. in Mathematics and 0.122 s.d. in Italian), indicating that the findings are not driven by the normalization procedure. Consistent with previous columns, Grade 5 estimates remain imprecise and not statistically significant.

Introducing school-specific linear time trends (column (10)) provides a further robustness check for differential pre-existing trends. While the estimated effects for Grade 2 decline in magnitude, the Mathematics coefficient remains statistically significant (0.148 s.d.), whereas the Italian estimate declines and is no longer statistically significant. This attenuation is consistent with the demanding nature of this specification and suggests that part of the effect may reflect differential pre-existing trends, although the core result in Mathematics remains robust. As in all previous specifications, Grade 5 estimates remain not statistically significant.

Finally, the propensity-score-matched DID specification (column (11)) corroborates the main findings. The estimated effect in Grade 2 Mathematics remains positive, sizeable, and statistically significant (0.232 s.d.), and the Italian estimate is also positive and statistically significant (0.115 s.d.). This indicates that differences in observable characteristics between treated and control schools are unlikely to drive the results. In line with the rest of the analysis, Grade 5 estimates remain small and not statistically significant.

In addition, Appendix Figure A.1 provides further support using randomization inference, a permutation-based approach that assesses whether the observed estimate is extreme relative to the distribution obtained under alternative assignments of treatment (Heß, 2017). The placebo distributions are centered around zero for all subject–grade combinations. For Grade 2, the estimated effects lie in the upper part of the placebo distribution, particularly in Mathematics, suggesting that effects of this magnitudes are unlikely to arise under random reassignment of treatment. By contrast, for Grade 5 outcomes, the estimated effects fall closer to the bulk of the placebo distribution, reinforcing the conclusion that they are not statistically significant.

Appendix Table A.1 reports two additional placebo exercises. Panel A assigns treatment status to untreated schools, keeping the number of treated units equal to that in the original Area C sample, while Panel B randomizes the timing of the intervention within the pre-treatment period. In both cases, the estimated placebo coefficients are small and not statistically significant across all subjects and grades. This absence of spurious effects provides further reassurance that the main results are unlikely to be driven by pre-existing differences between treated and control schools or by coincidental timing.

We also assess the robustness of our results to the introduction of Area B — the broader LEZ encompassing Area C — in February 2019, which may partially affect the control group in the last year of our study period. However, given that INVALSI tests are administered in spring 2019, the effective exposure window amounts to only a few weeks. Table A.2 in the Appendix confirms that excluding 2019 does not affect the robustness of the estimates.

Finally, a further concern relates to the limited number of treated schools (N=17), which raises the question of whether the SDID estimates are driven by a small number of influential units. To address this, we implement a jackknife procedure that sequentially excludes each treated school and re-estimates the model, providing a direct assessment of whether any single unit drives the results (Arkhangelsky et al., 2021). Table A.3 in the Appendix reports the results. Estimates remain stable across all leave-one-out samples and are consistent with the baseline SDID results reported in Table 6, confirming that the findings are not driven by any single treated school.

The robustness checks consistently support the central conclusion of the paper. The positive effects of Area C are concentrated in early grades, with strong and robust gains in Grade 2 Mathematics and more moderate and less precisely estimated effects in Italian. In contrast, there is no evidence of meaningful effects in Grade 5. Overall, the results appear robust across a wide range of specifications, samples, and inference methods, as well as across placebo tests and randomization inference exercises, and are unlikely to be driven by spurious correlations or specific modeling choices.

7 Discussion of potential mechanisms

This section discusses the potential mechanisms through which Area C may have affected students' academic performance. Our estimates capture the overall effect of Area C on student performance, but they do not allow us to isolate the specific mechanisms through which this effect operates. We therefore discuss plausible mechanisms rather than provide direct mediation tests. The most direct channel operates through improvements in i) environmental conditions around schools, including reductions in air pollution, ii) indirect channels through adults and the school environment, iii) traffic-related noise. The factors may affect children's health, cognitive development, attention, and ultimately academic performance. The remainder of this section is organized around the three main mechanisms described above, with an additional subsection dedicated to other potential channels through which these effects may arise.

7.1 Children's health and cognitive development

Previous studies have documented the detrimental effect of air pollution on children's health and development, highlighting their higher vulnerability compared to adults. Exposure to traffic-related pollutants has been associated with adverse health outcomes in children, including impaired respiratory function, asthma, bronchitis, and other respiratory illnesses (Barone-Adesi et al., 2015; Alotaibi et al., 2019; Coneus and Spiess, 2012; Currie et al., 2009). More recent evidence showed that moderate improvements in air quality following the introduction of LEZs in Germany were associated with reductions in prescriptions for respiratory diseases among young children (Klauber et al., 2024). Using brain imaging data, Cserbik et al. (2020) showed that annual exposure to fine particulate matter is correlated with significant differences in brain morphology, providing neurobiological evidence on the impact of pollution on children's brain development.

Evidence from school and cognitive-based exposure measures points in the same direction. Alvarez-Pedrerol et al. (2017); Milojevic et al. (2021) demonstrated that chronic exposure to air pollutants, including particulate matter and nitrogen dioxide, is linked with decreased cognitive development, particularly in terms of working memory and processing speed. Sunyer et al. (2015) showed that traffic-related air pollution measured directly in school courtyards and classrooms negatively affects children's working memory and inattention. Conducting a study on primary school children in Barcelona, Sunyer et al. (2017) demonstrated that daily ambient levels in terms of nitrogen dioxide were negatively associated with all attention processes. A review of empirical studies on air pollution around schools similarly concludes that fine particulate matter and nitrogen dioxide are associated with deficits in working memory and school achievement (Gartland et al., 2022).

Although our data do not allow us to directly test health-related mechanisms, the larger impact on Mathematics, together with the presence of effects in Grade 2 but not in Grade 5, provides indirect support for a cognitive channel linking air pollution to academic performance.

First, the larger effect for Mathematics may reflect the distinct cognitive demands of these disciplines with mathematics requiring problem-solving and higher-order reasoning

(Smith and Stein, 1998; Tekkumru-Kisa et al., 2015). This pattern is consistent with previous evidence showing stronger impacts of air pollution on performance in analytical subjects (Bernardi and Keivabu, 2024; Amanzadeh et al., 2020; Duque and Gilraine, 2022). Since more cognitively demanding tasks are likely to be more sensitive to environmental stressors, air pollution may have stronger effects in these domains. Second, the presence of significant effects in Grade 2 but not in Grade 5 suggests that reductions in air pollution matter more for younger children, in line with evidence that cognitive development is more malleable and, therefore, more vulnerable to environmental exposure at younger ages (Persico and Venator, 2021).

7.2 Indirect channels through adults and the school environment

A second mechanism that may explain the impact of Area C on educational achievement operates through teachers and parents. Improved air quality may enhance their cognitive functioning, reduce absenteeism, and increase productivity, consistent with evidence on the effects of air pollution on workers (Archsmith et al., 2018; Lu, 2020). In addition, better health conditions and lower stress levels among parents may enable them to provide greater support to their children. As a result, children's overall learning environment may improve, leading to indirect positive effects on educational outcomes through this channel.

Air pollution may also affect learning conditions more directly through school attendance and classroom disruption. Several studies document that pollution exposure increases school absenteeism (Currie et al., 2009; Marcon et al., 2014). Recent evidence also suggested that short-term exposure to $PM_{2.5}$ can increase student absences, disciplinary incidents, and teacher sick leave, thereby disrupting learning conditions within schools (Chung et al., 2025). In the context of traffic regulation, Conte Keivabu and Rüttenauer (2022) showed that stricter traffic restrictions in Central London reduced pollution and were associated with lower absenteeism in schools serving more disadvantaged students. While we cannot observe absences, disciplinary incidents, or teacher health in our data, this literature suggests that improvements in air quality may affect student achievement not only through children's own health and cognition, but also through the functioning of the school environment.

7.3 Noise exposure

A third channel through which Area C may have affected children's educational achievement relates to the acoustic environment. Because Area C reduced traffic in Milan's city center, it may also have reduced traffic-related noise around treated schools. Evidence from school-based studies supports this channel. Chronic aircraft noise exposure has been associated with lower reading comprehension and recognition memory among children (Stansfeld et al., 2005; Clark et al., 2013), while higher outdoor noise exposure at school has been linked to lower standardized scores in language and mathematics among primary-school pupils (Pujol et al., 2014). Studies focusing on road traffic noise found that even small increases in noise levels around schools are associated with reduced working memory and slower development of attentional functions (Foraster et al., 2022). More recent evidence highlighted the joint effects of air pollution and noise on cognitive outcomes (Thompson et al., 2024).

Although our analysis does not directly test this mechanism, similar arguments to those discussed for health and cognitive channels may apply. In particular, the larger impact on Mathematics, together with the presence of effects in Grade 2 but not in Grade 5, provides indirect support for a cognitive channel through which reductions in noise exposure may contribute to improved academic performance. However, because Area C affected traffic, air pollution, and potentially noise at the same time, our estimates cannot separately identify the contribution of each environmental channel.

7.4 Other potential mechanisms

Other potential mechanisms may have contributed to the observed effects, although the available evidence suggests that they are unlikely to be the primary drivers of our results. The implementation of Area C in Milan's city center may have affected neighborhood amenities, liveliness and perceptions, potentially influencing school composition through residential and school choices. However, such compositional responses are more likely to materialize gradually, and the descriptive statistics reported in Table 3 indicate that student characteristics remain broadly stable over the period, with only moderate changes along specific dimensions. Our empirical specifications also include student-level covariates, including parental occupation, gender, migration background, early childhood education attendance, and indicators of grade retention or irregular school trajectories, which helps account for observed changes in student composition over time.

A similar mechanism may operate through teachers, as improved neighborhood conditions could affect teachers' location or school choices, potentially altering the composition of the teaching workforce across schools. However, the lack of data on teacher characteristics prevents us from directly testing or controlling for this channel. While these mechanisms may complement the primary effects of improved air quality on children's health and development, we expect their magnitude to be secondary. Future research could further investigate these channels by jointly analyzing residential and labor mobility alongside environmental improvements.

8 Conclusion

This paper examines the impact of Milan's Area C congestion policy on the academic performance of primary school students, using individual-level data from the INVALSI national standardized assessment system spanning academic years 2009/10 to 2018/19. By exploiting the sharp introduction of the congestion policy on 2012, we estimate the causal effect of improved urban air quality on children's educational outcomes through a two-way fixed effects difference-in-differences design, complemented by synthetic difference-in-differences estimates and event-study specifications.

Our central finding is that the introduction of Area C generated significant improvements in early academic achievement, with effects concentrated among second-grade students. Exposure to the congestion policy leads to a statistically significant increase of 0.256 s.d. in Mathematics scores and 0.116 s.d. in Italian scores for Grade 2 students. These results are robust across a wide range of alternative specifications, sample definitions, and inference methods, including class fixed effects, balanced panel estimates, multiple imputation, matched difference-in-differences, and wild cluster bootstrap inference. By contrast, no statistically significant effects emerge for fifth-grade students in either subject, a pattern consistent with the view that younger children's cognitive development is more malleable and hence more sensitive to reductions in environmental pollution.

Moreover, the heterogeneity analysis reveals that the gains from Area C are not uniform across students. Estimated improvements are larger among children whose parents hold lower-status occupations, suggesting that treatment effects vary by socioeconomic background. These findings point to an equity-enhancing dimension of congestion pricing that has received limited attention in the existing literature: beyond reducing traffic and improving air quality, such policies may generate significant redistributive effects on children's human capital formation, narrowing early inequalities that closely mirror the socioeconomic structure of the city.

These results carry important implications for the design and evaluation of urban environmental policies. First, they suggest that the benefits of congestion policies extend beyond the direct reduction of traffic and pollution emphasized in prior work on Milan (Boniardi et al., 2025), pointing to broader societal effects. When assessing the welfare effects of such policies, standard cost-benefit analysis that focus exclusively on congestion relief, fuel savings, and health outcomes are likely to substantially underestimate their total social value. The human capital gains documented here,particularly among disadvantaged children, represent an additional and previously overlooked source of social return that should be incorporated into policy evaluation frameworks. Second, the concentration of effects among younger children implies that the timing of environmental interventions matters for their educational impact. Policies that improve air quality during the early years of primary school, when cognitive development is most malleable, are likely to yield larger and more persistent gains than interventions targeting older students. This has direct implications for the sequencing and prioritization of urban air quality policies, suggesting that early childhood and early primary school environments deserve particular attention in the design of urban mobility strategies. Third, the equity-enhancing nature of the effects documented here provides an additional argument for congestion policies in cities characterized by pronounced socioeconomic segregation. In contexts where disadvantaged families are disproportionately concentrated in high-pollution urban areas with limited access to compensating resources, environmental regulation can serve as a *de facto* equalizing force, reducing early achievement gaps without requiring direct redistribution through the educational system. This complementarity between environmental and social policy deserves greater recognition in congestion policy evaluation discussions.

Several limitations of this study warrant acknowledgment. First, the pre-treatment window is limited to two academic years, which constrains the power of formal pre-trend tests and means that identification relies more heavily on the plausibility of the parallel trends assumption than would be ideal. While our event-study estimates do not reveal systematic pre-existing differences between treated and control schools, this evidence should be interpreted with appropriate caution. It is noteworthy the possibility that unobserved time-varying factors may affect trends in academic performance in treated and control schools. For instance, changes in competitive pressure over time may influence school quality (Agasisti and Murtinu, 2012), with potentially heterogeneous effects across suburban, peripheral, and central urban areas. Second, the number of treated schools is relatively small with seventeen primary schools located within the Area C perimeter. This may limit statistical precision and may affect the generalizability of findings within the area.

Notwithstanding these limitations, this study contributes to a growing body of evidence that environmental regulation generates meaningful externalities for children's development. By providing the first individual-level causal estimates of the educational effects of a congestion policy, and by documenting the equity-enhancing nature of these effects in a unexplored context, it strengthens the empirical findings on the broader role of air pollution on children's development. As European cities continue to expand and tighten LEZs, the evidence accumulated across diverse settings will be essential for informing both the design of these policies and the broader public debate about their social value.

References

Agasisti, T., Murtinu, S., 2012. 'perceived'competition and performance in italian secondary schools: new evidence from oecd–pisa 2006. British Educational Research Journal 38, 841–858.

Alotaibi, R., Bechle, M., Marshall, J.D., Ramani, T., Zietsman, J., Nieuwenhuijsen, M.J., Khreis, H., 2019. Traffic related air pollution and the burden of childhood asthma in the contiguous united states in 2000 and 2010. Environment international 127, 858–867.

Alvarez-Pedrerol, M., Rivas, I., López-Vicente, M., Suades-González, E., Donaire-Gonzalez, D., Cirach, M., de Castro, M., Esnaola, M., Basagaña, X., Dadvand, P., et al., 2017. Impact of commuting exposure to traffic-related air pollution on cognitive development in children walking to school. Environmental pollution 231, 837–844.

Amanzadeh, N., Vesal, M., Ardestani, S.F.F., 2020. The impact of short-term exposure to ambient air pollution on test scores in iran. Population and Environment 41, 253–285.

Anenberg, S.C., Mohegh, A., Goldberg, D.L., Kerr, G.H., Brauer, M., Burkart, K., Hystad, P., Larkin, A., Wozniak, S., Lamsal, L., 2022. Long-term trends in urban no2 concentrations and associated paediatric asthma incidence: estimates from global datasets. The Lancet Planetary Health 6, e49–e58.

Archsmith, J., Heyes, A., Saberian, S., 2018. Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation. Journal of the Association of Environmental and Resource Economists 5, 827–863.

Arkhangelsky, D., Athey, S., Hirshberg, D.A., Imbens, G.W., Wager, S., 2021. Synthetic difference-in-differences. American economic review 111, 4088–4118.

Avila-Uribe, A., Roth, S., Shields, B., 2024. Putting Low Emission Zone (LEZ) to the Test: The Effect of London's LEZ on Education. Technical Report. IZA Institute of Labor Economics.

Axsen, J., Wolinetz, M., 2021. Taxes, tolls and zev zones for climate: Synthesizing insights on effectiveness, efficiency, equity, acceptability and implementation. Energy Policy 156, 112457.

Barone-Adesi, F., Dent, J.E., Dajnak, D., Beevers, S., Anderson, H.R., Kelly, F.J., Cook, D.G., Whincup, P.H., 2015. Long-term exposure to primary traffic pollutants and lung function in children: cross-sectional study and meta-analysis. PloS one 10, e0142565.

Battistin, E., Meroni, E.C., 2016. Should we increase instruction time in low achieving schools? evidence from southern italy. Economics of Education Review 55, 39–56.

Bedogni, M., 2013. Emissioni atmosferiche da traffico stradale in Area C nel periodo gennaio–dicembre 2012. Technical Report. Agenzia Mobilità Ambiente e Territorio, Comune di Milano. Milan, Italy.

Bennett, J., Davy, P., Trompetter, B., Wang, Y., Pierse, N., Boulic, M., Phipps, R., Howden-Chapman, P., 2019. Sources of indoor air pollution at a new zealand urban primary school; a case study. Atmospheric Pollution Research 10, 435–444.

Beria, P., 2016. Effectiveness and monetary impact of milan's road charge, one year after implementation. International Journal of Sustainable Transportation 10, 657–669.

Bernardi, F., Keivabu, R.C., 2024. Poor air quality at school and educational inequality by family socioeconomic status in italy. Research in Social Stratification and Mobility 91, 100932.

Bishop, K., Corkery, L., 2017. Designing cities with children and young people: Beyond playgrounds and skate parks. Taylor & Francis.

Boniardi, L., Nobile, F., Stafoggia, M., Michelozzi, P., Ancona, C., 2025. Assessing the impact of traffic restriction interventions on school air quality: a citizen science-based modelling study. Environmental Research 277, 121562.

Brehm, J., Pestel, N., Schaffner, S., Schmitz, L., 2025. From low emission zone to academic track: Environmental policy effects on educational achievement in elementary school. Journal of Environmental Economics and Management 132, 103165.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. The review of economics and statistics 90, 414–427.

Chang, T.Y., Graff Zivin, J., Gross, T., Neidell, M., 2019. The effect of pollution on worker productivity: evidence from call center workers in china. American Economic Journal: Applied Economics 11, 151–172.

Chay, K.Y., Greenstone, M., 2003. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. The quarterly journal of economics 118, 1121–1167.

Chung, S., Persico, C., Liu, J., 2025. The Effects of Daily Air Pollution on Students and Teachers. Technical Report. National Bureau of Economic Research.

Clark, C., Head, J., Stansfeld, S.A., 2013. Longitudinal effects of aircraft noise exposure on children's health and cognition: a six-year follow-up of the uk ranch cohort. Journal of Environmental Psychology 35, 1–9.

Coneus, K., Spiess, C.K., 2012. Pollution exposure and child health: Evidence for infants and toddlers in germany. Journal of Health Economics 31, 180–196.

Conte Keivabu, R., Rüttenauer, T., 2022. London congestion charge: the impact on air pollution and school attendance by socioeconomic status. Population and Environment 43, 576–596.

Contini, D., Di Tommaso, M.L., Muratori, C., Piazzalunga, D., Schiavon, L., 2025. A lost generation? the impact of covid-19 on high school students' achievements: D. contini et al. The Journal of Economic Inequality , 1–28.

Corazzini, L., Meschi, E., Pavese, C., 2021. Impact of early childcare on immigrant children's educational performance. Economics of Education Review 85, 102181.

Cserbik, D., Chen, J.C., McConnell, R., Berhane, K., Sowell, E.R., Schwartz, J., Hackman, D.A., Kan, E., Fan, C.C., Herting, M.M., 2020. Fine particulate matter exposure during childhood relates to hemispheric-specific differences in brain structure. Environment international 143, 105933.

Currie, J., Neidell, M., Schmieder, J.F., 2009. Air pollution and infant health: Lessons from new jersey. Journal of health economics 28, 688–703.

Davis, L.W., 2008. The effect of driving restrictions on air quality in mexico city. Journal of Political Economy 116, 38–81.

De Palma, A., Lindsey, R., 2011. Traffic congestion pricing methodologies and technologies. Transportation Research Part C: Emerging Technologies 19, 1377–1399.

Duque, V., Gilraine, M., 2022. Coal use, air pollution, and student performance. Journal of Public Economics 213, 104712.

Etzel, R.A., 2020. The special vulnerability of children. International journal of hygiene and environmental health 227, 113516.

Foraster, M., Esnaola, M., López-Vicente, M., Rivas, I., Álvarez-Pedrerol, M., Persavento, C., Sebastian-Galles, N., Pujol, J., Dadvand, P., Sunyer, J., 2022. Exposure to road traffic noise and cognitive development in schoolchildren in barcelona, spain: A population-based cohort study. PLoS medicine 19, e1004001.

Galdon-Sanchez, J.E., Gil, R., Holub, F., Uriz-Uharte, G., 2023. Social benefits and private costs of driving restriction policies: The impact of madrid central on congestion, pollution, and consumer spending. Journal of the European Economic Association 21, 1227–1267.

Gallego, F., Montero, J.P., Salas, C., 2013. The effect of transport policies on car use: Evidence from latin american cities. Journal of Public Economics 107, 47–62.

Gartland, N., Aljofi, H.E., Dienes, K., Munford, L.A., Theakston, A.L., Van Tongeren, M., 2022. The effects of traffic air pollution in and around schools on executive function and academic performance in children: a rapid review. International Journal of Environmental Research and Public Health 19, 749.

Gibson, M., Carnovale, M., 2015. The effects of road pricing on driver behavior and air pollution. Journal of Urban Economics 89, 62–73.

Hanushek, E., 1998. The evidence on class size. Earning and learning: How schools matter .

Hanushek, E.A., 1986. The economics of schooling: Production and efficiency in public schools. Journal of economic literature 24, 1141–1177.

Health Effects Institute, 2010. Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects. Panel on the Health Effects of Traffic-Related Air Pollution .

Hensher, D.A., Li, Z., 2013. Referendum voting in road pricing reform: A review of the evidence. Transport Policy 25, 186–197.

Heß, S., 2017. Randomization inference with stata: A guide and software. The Stata Journal 17, 630–651.

Holman, C., Harrison, R., Querol, X., 2015. Review of the efficacy of low emission zones to improve urban air quality in european cities. Atmospheric Environment 111, 161–169.

Khreis, H., Kelly, C., Tate, J., Parslow, R., Lucas, K., Nieuwenhuijsen, M., 2017. Exposure to traffic-related air pollution and risk of development of childhood asthma: a systematic review and meta-analysis. Environment international 100, 1–31.

Klauber, H., Holub, F., Koch, N., Pestel, N., Ritter, N., Rohlf, A., 2024. Killing prescriptions softly: Low emission zones and child health from birth to school. American Economic Journal: Economic Policy 16, 220–248.

Kumar, P., Rivas, I., Sachdeva, L., 2017. Exposure of in-pram babies to airborne particles during morning drop-in and afternoon pick-up of school children. Environmental pollution 224, 407–420.

Lebrusán, I., Toutouh, J., 2021. Car restriction policies for better urban health: a low emission zone in madrid, spain. Air Quality, Atmosphere & Health 14, 333–342.

Lehe, L., 2019. Downtown congestion pricing in practice. Transportation Research Part C Emerging Technologies 100, 200–223.

Lu, J.G., 2020. Air pollution: A systematic review of its psychological, economic, and social effects. Current opinion in psychology 32, 52–65.

Marcon, A., Pesce, G., Girardi, P., Marchetti, P., Blengio, G., de Zolt Sappadina, S., Falcone, S., Frapporti, G., Predicatori, F., De Marco, R., 2014. Association between pm10 concentrations and school absences in proximity of a cement plant in northern italy. International journal of hygiene and environmental health 217, 386–391.

Margaryan, S., 2021. Low emission zones and population health. Journal of health economics 76, 102402.

Milojevic, A., Dutey-Magni, P., Dearden, L., Wilkinson, P., 2021. Lifelong exposure to air pollution and cognitive development in young children: the uk millennium cohort study. Environmental Research Letters 16, 055023.

Mölter, A., Agius, R., De Vocht, F., Lindley, S., Gerrard, W., Custovic, A., Simpson, A., 2014. Effects of long-term exposure to pm10 and no2 on asthma and wheeze in a prospective birth cohort. J Epidemiol Community Health 68, 21–28.

Nieuwenhuijsen, M., De Nazelle, A., Pradas, M.C., Daher, C., Dzhambov, A.M., Echave, C., Gössling, S., Iungman, T., Khreis, H., Kirby, N., et al., 2024. The superblock model: A review of an innovative urban model for sustainability, liveability, health and well-being. Environmental research 251, 118550.

Orellano, P., Quaranta, N., Reynoso, J., Balbi, B., Vasquez, J., 2017. Effect of outdoor air pollution on asthma exacerbations in children and adults: systematic review and multilevel meta-analysis. PloS one 12, e0174050.

Park, H., Lee, B., Ha, E.H., Lee, J.T., Kim, H., Hong, Y.C., 2002. Association of air pollution with school absenteeism due to illness. Archives of pediatrics & adolescent medicine 156, 1235–1239.

Percoco, M., 2013. Is road pricing effective in abating pollution? evidence from milan. Transportation Research Part D: Transport and Environment 25, 112–118.

Percoco, M., 2014. The impact of road pricing on housing prices: Preliminary evidence from milan. Transportation Research Part A: Policy and Practice 67, 188–194.

Persico, C.L., Venator, J., 2021. The effects of local industrial pollution on students and schools. Journal of Human Resources 56, 406–445.

Pestel, N., Wozny, F., 2021. Health effects of low emission zones: evidence from german hospitals. Journal of Environmental Economics and Management 109, 102512.

Pujol, S., Levain, J.P., Houot, H., Petit, R., Berthillier, M., Defrance, J., Lardies, J., Masselot, C., Mauny, F., 2014. Association between ambient noise exposure and school performance of children living in an urban area: A cross-sectional population-based study. Journal of Urban Health 91, 256–271.

Raysoni, A.U., Sarnat, J.A., Sarnat, S.E., Garcia, J.H., Holguin, F., Luèvano, S.F., Li, W.W., 2011. Binational school-based monitoring of traffic-related air pollutants in el paso, texas (usa) and ciudad juárez, chihuahua (méxico). Environmental Pollution 159, 2476–2486.

Rotaris, L., Danielis, R., Marcucci, E., Massiani, J., 2010. The urban road pricing scheme to curb pollution in milan, italy: Description, impacts and preliminary cost–benefit analysis assessment. Transportation Research Part A: Policy and Practice 44, 359–375.

Salas, R., Perez-Villadoniga, M.J., Prieto-Rodriguez, J., Russo, A., 2021. Were traffic restrictions in madrid effective at reducing no2 levels? Transportation Research Part D: Transport and Environment 91, 102689.

Sarmiento, L., Wägner, N., Zaklan, A., 2023. The air quality and well-being effects of low emission zones. Journal of Public Economics 227, 105014.

Simeonova, E., Currie, J., Nilsson, P., Walker, R., 2021. Congestion pricing, air pollution, and children's health. Journal of Human resources 56, 971–996.

Smith, M.S., Stein, M.K., 1998. Reflections on practice: Selecting and creating mathematical tasks: From research to practice. Mathematics teaching in the middle school 3, 344–350.

Stafford, T.M., 2015. Indoor air quality and academic performance. Journal of Environmental Economics and Management 70, 34–50.

Stansfeld, S.A., Berglund, B., Clark, C., Lopez-Barrio, I., Fischer, P., Öhrström, E., Haines, M.M., Head, J., Hygge, S., Van Kamp, I., et al., 2005. Aircraft and road traffic noise and children's cognition and health: a cross-national study. The lancet 365, 1942–1949.

Sunyer, J., Esnaola, M., Alvarez-Pedrerol, M., Forns, J., Rivas, I., López-Vicente, M., Suades-González, E., Foraster, M., Garcia-Esteban, R., Basagaña, X., et al., 2015. Association between traffic-related air pollution in schools and cognitive development in primary school children: a prospective cohort study. PLoS medicine 12, e1001792.

Sunyer, J., Suades-González, E., García-Esteban, R., Rivas, I., Pujol, J., Alvarez-Pedrerol, M., Forns, J., Querol, X., Basagaña, X., 2017. Traffic-related air pollution and attention in primary school children: short-term association. Epidemiology 28, 181–189.

Tekkumru-Kisa, M., Stein, M.K., Schunn, C., 2015. A framework for analyzing cognitive demand and content-practices integration: Task analysis guide in science. Journal of Research in Science Teaching 52, 659–685.

Thompson, R., Stewart, G., Vu, T., Jephcote, C., Lim, S., Barratt, B., Smith, R.B., Karim, Y.B., Mussa, A., Mudway, I., et al., 2024. Air pollution, traffic noise, mental health, and cognitive development: A multi-exposure longitudinal study of london adolescents in the scamp cohort. Environment International 191, 108963.

Valdés, M.T., Espadafor, M.C., Conte Keivabu, R., 2025. Can a low emission zone improve academic performance? evidence from a natural experiment in the city of madrid. Population and Environment 47, 8.

Veitch, E., Rhodes, E., 2024. A cross-country comparative analysis of congestion pricing systems: Lessons for decarbonizing transportation. Case Studies on Transport Policy 15, 101128.

White, I.R., Royston, P., Wood, A.M., 2011. Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine 30, 377–399.

Zivin, J.G., Neidell, M., 2012. The impact of pollution on worker productivity. American Economic Review 102, 3652–3673.

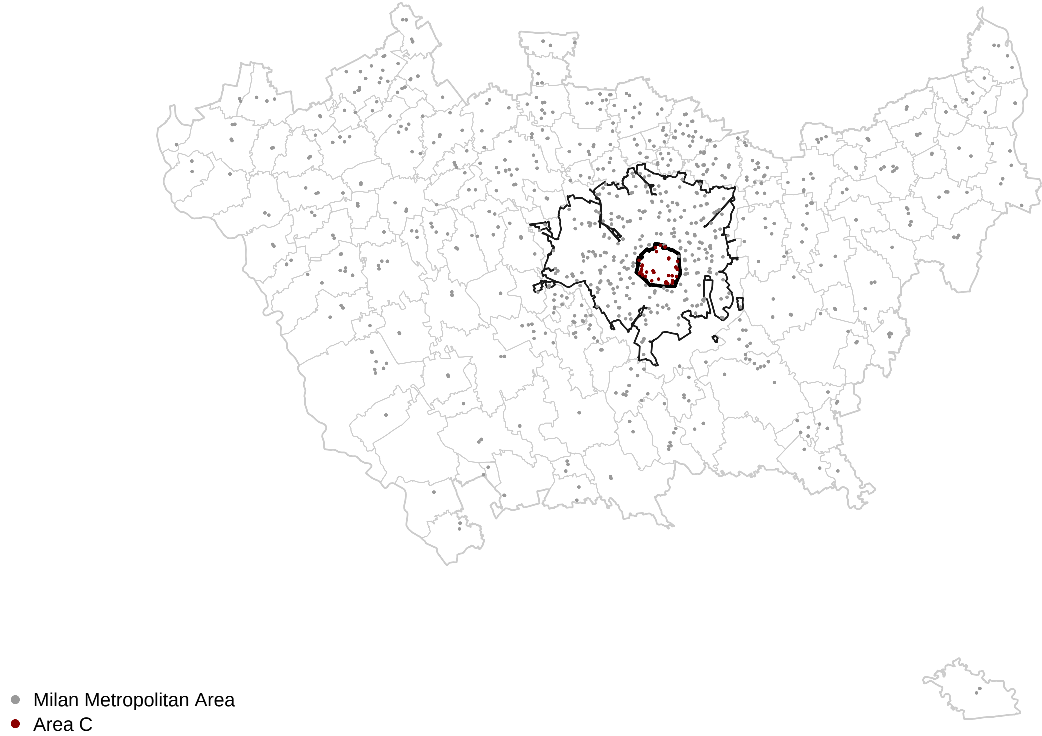

Figure 1: Map of primary schools in the Metropolitan City of Milan

Notes: Administrative boundaries of municipalities are shown in grey. The boundaries of Area B and Area C are highlighted in black. Primary schools are represented as dots: grey dots indicate schools located outside Area C, while dark red dots indicate schools within Area C. The map displays all primary schools initially identified in the Milan Metropolitan Area prior to sample selection. The number of dots may exceed the number of schools in the analytical sample, as a single school may operate across multiple buildings or locations.

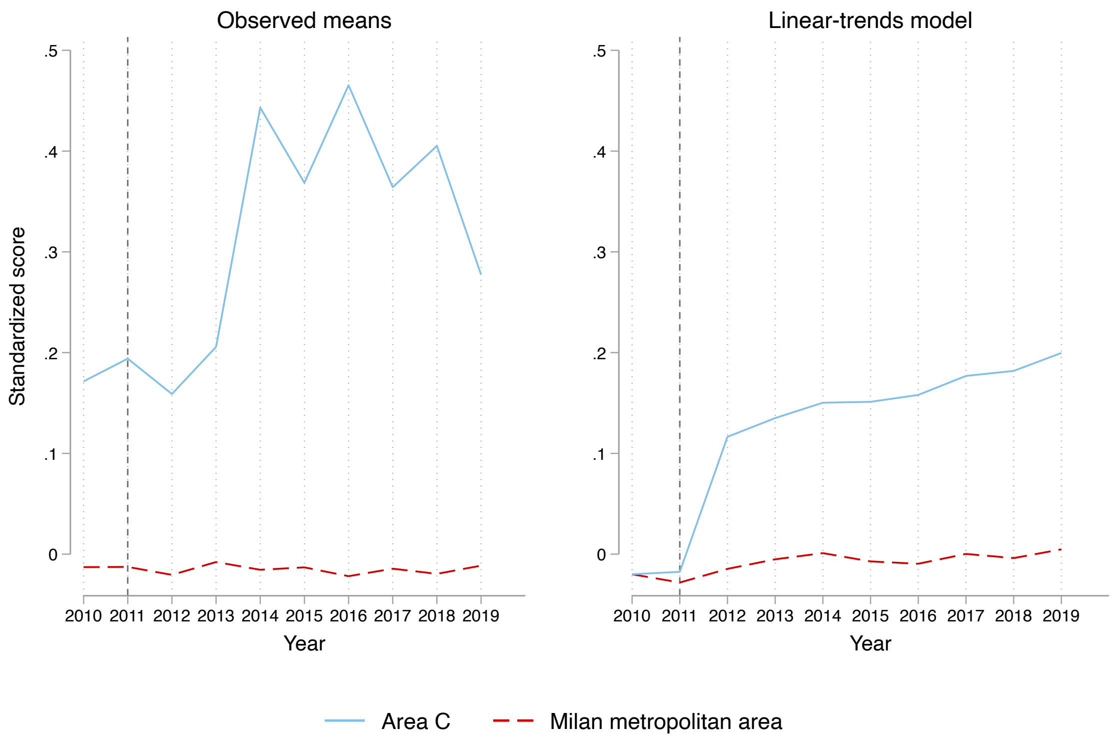

Figure 2: Evolution of standardized test scores in area C and the metropolitan area

Notes: The figure illustrates the evolution of average standardized test scores in schools located inside Area C and in those in the wider metropolitan area of Milan between 2010 and 2019. The left-hand figure reports the observed means for both groups. The right-hand figure shows predicted values from a linear-trend model estimated separately by exposure status. The year 2012 marks the introduction of Area C.

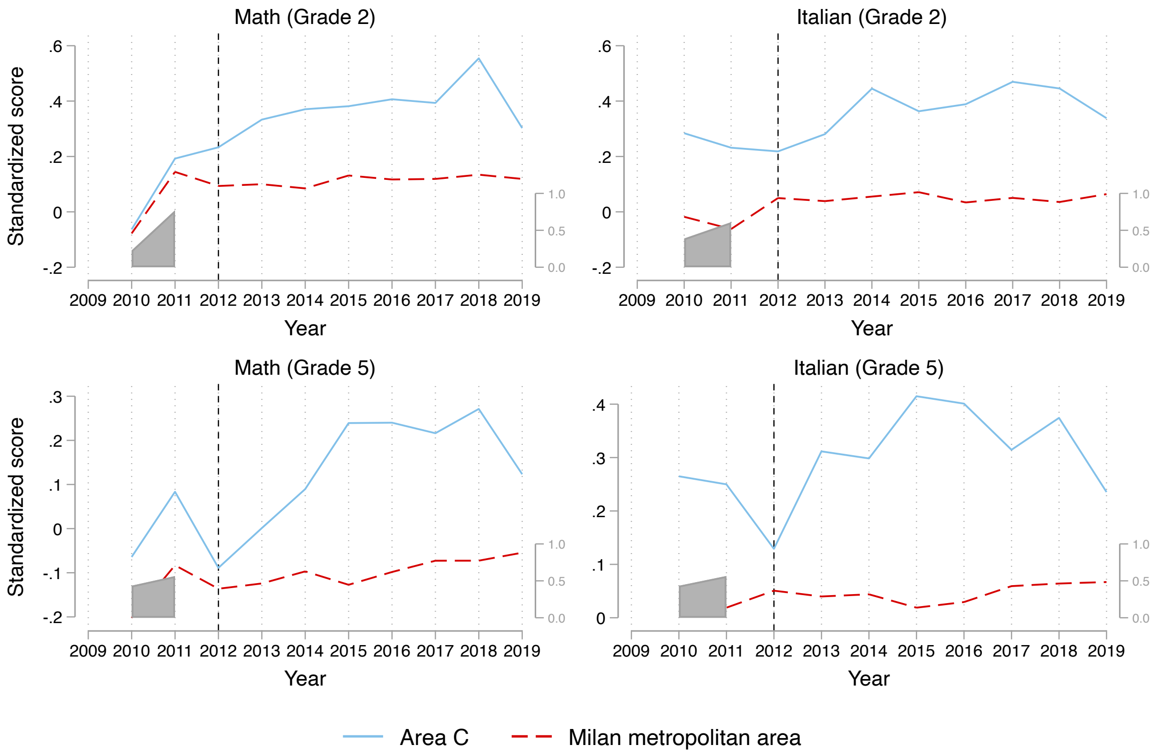

Figure 3: Synthetic difference-in-differences reweighted trends

Notes: The figure plots the evolution of standardized test scores for treated schools and for the SDID-reweighted control group. The reweighting procedure reduces pre-treatment level differences between Area C schools and the wider metropolitan area, yielding a closer match in the pre-intervention period. Post-treatment divergences therefore reflect dynamic SDID estimates of the policy's effect.

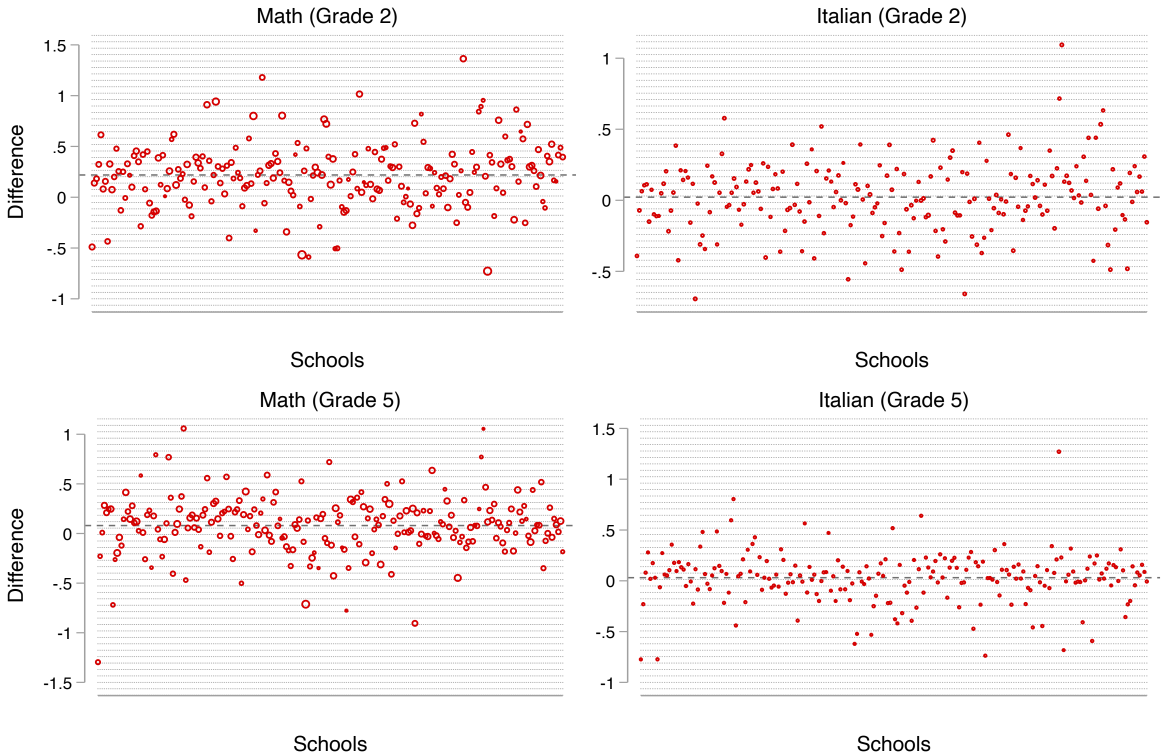

Figure 4: Synthetic difference-in-differences unit weights

Notes: Each point represents a single school and shows its post-treatment difference relative to the outcome predicted by the SDID synthetic control. The size of each point reflects the unit weight assigned by SDID when constructing the synthetic control. The vertical dispersion captures the variation in school-level post-treatment gaps, and the distribution of point sizes summarizes how SDID allocates weights across control schools.

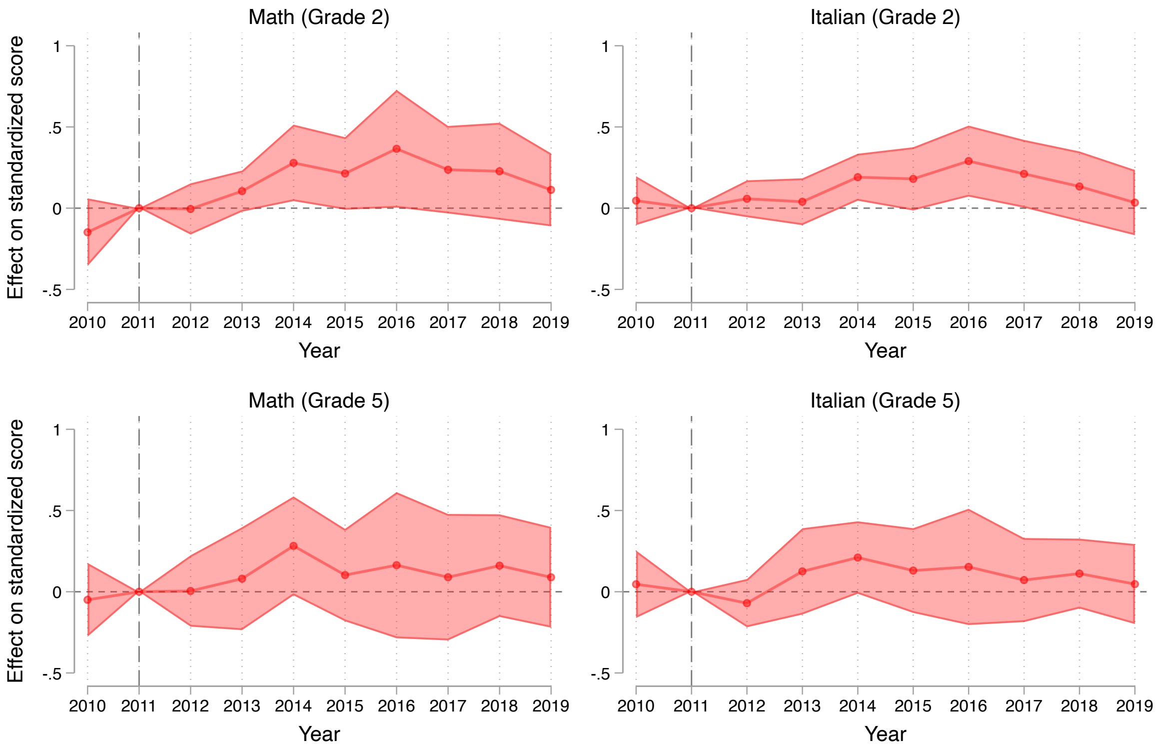

Figure 5: Event-study estimates of the effects of area C on student achievement

Notes: The figure reports event-study estimates of the impact of Area C on standardized test scores, expressed relative to the year before implementation. Estimates come from specifications including school and year fixed effects and the full set of individual controls. Shaded areas indicate 90% confidence intervals based on standard errors clustered at the school level.

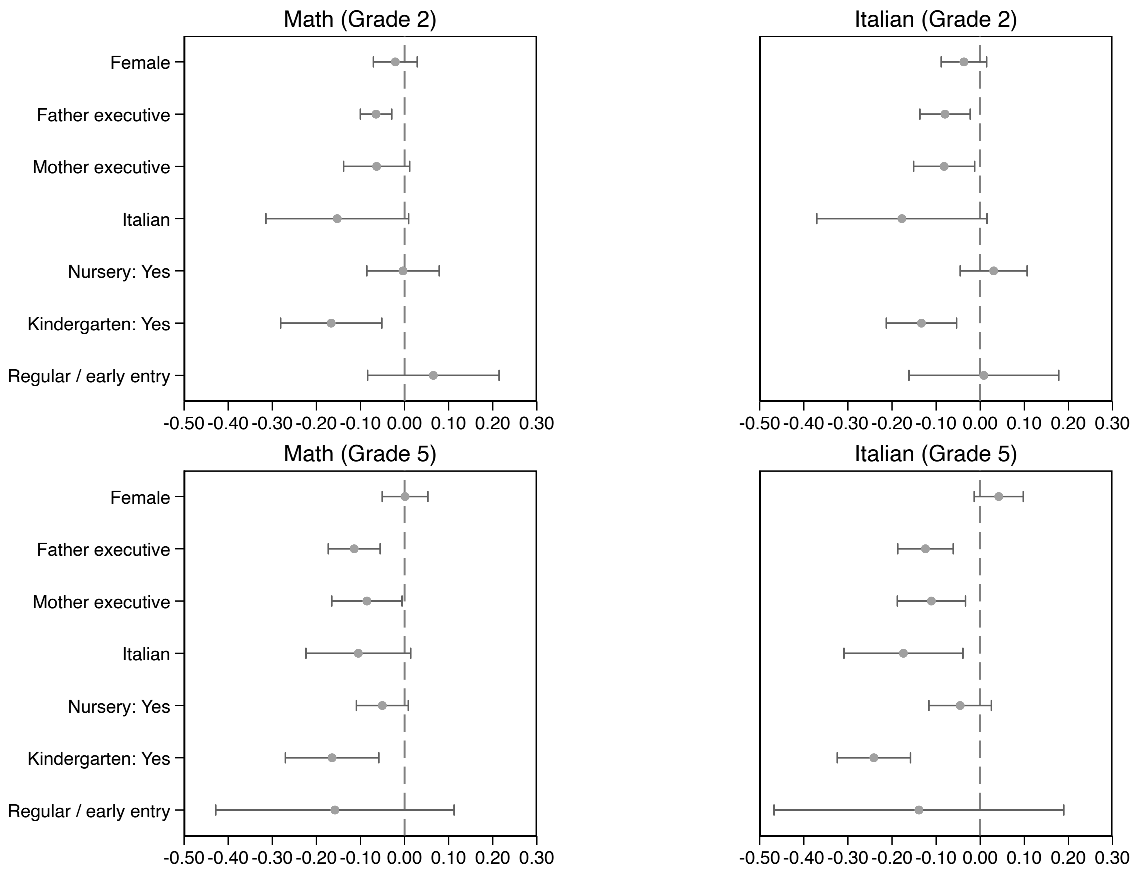

Figure 6: Treatment effect heterogeneity across student subgroups

Notes: Each panel reports estimated treatment effects for specific student subgroups, separately for Mathematics and Italian in Grades 2 and 5. Estimates are obtained from fully interacted specifications in which the Area C treatment indicator is interacted with subgroup membership. Whiskers denote 95% confidence intervals based on standard errors clustered at the school level.

Table 1: Differences in annual mean NO_2 concentrations (weekdays, 7:00–20:00) relative to the pre-treatment period (2009–2011, mean = 62.272 µg/m^3)

Year	N (hours)	Mean NO_2	Difference from pre-period	p-value
2012	3452	56.606	-5.667	0.000
2013	3534	58.641	-3.631	0.000
2014	3428	51.608	-10.664	0.000
2015	3493	53.377	-8.895	0.000
2016	3603	52.816	-9.457	0.000
2017	3545	53.834	-8.438	0.000
2018	3217	51.333	-10.939	0.000
2019	3576	43.688	-18.585	0.000

Notes: The table reports differences in annual mean NO_2 concentrations (µg/m^3) calculated using hourly observations restricted to weekdays between 7:00 and 20:00. The reference period is the pooled 2009–2011 sample (mean = 62.272 µg/m^3). P-values are based on one-sided Welch two-sample t-tests testing whether annual means are lower than in the pre-period.

Table 2: Differences in annual mean PM_{10} concentrations (weekdays) relative to the pre-treatment period (2009–2011, mean = 46.099 µg/m^3)

Year	N (days)	Mean PM_{10}	Difference from pre-period	p-value
2012	260	42.823	-3.276	0.057
2013	254	34.827	-11.272	0.000
2014	255	33.788	-12.311	0.000
2015	237	41.671	-4.428	0.007
2016	249	33.980	-12.119	0.000
2017	250	39.828	-6.271	0.000
2018	243	34.049	-12.049	0.000
2019	240	29.921	-16.178	0.000

Notes: The table reports differences in annual mean PM_{10} concentrations (µg/m^3) computed from daily observations restricted to weekdays. The reference period is the pooled 2009–2011 sample (mean = 46.099 µg/m^3). P-values are based on one-sided Welch two-sample t-tests testing whether annual means are lower than in the pre-period.

Table 3: Covariate distribution before and after the implementation of Area C

	Area C (Before)	Area C (After)	p-value	Met. Area (Before)	Met. Area (After)	p-value
Gender						
Female	0.510	0.488	0.002	0.487	0.489	0.087
Male	0.487	0.490	0.637	0.508	0.506	0.124
Missing	0.003	0.022	0.000	0.005	0.004	0.203
Father's occupation						
Executive / senior official / university teacher / officer	0.210	0.194	0.005	0.058	0.051	0.000
Professional (employee or self-employed) / NCO	0.286	0.260	0.000	0.116	0.127	0.000
Clerical/teacher (graduate) / graduated military	0.040	0.069	0.000	0.107	0.211	0.000
Clerical/teacher (non-graduate) / enlisted	0.025	0.000	0.000	0.092	0.000	0.000
Self-employed (trader, artisan, farmer...)	0.073	0.065	0.024	0.132	0.118	0.000
Agricultural entrepreneur / farm owner	0.096	0.091	0.227	0.039	0.038	0.663
Manual worker / services / cooperative member	0.032	0.025	0.002	0.201	0.205	0.000
Retired	0.004	0.003	0.398	0.004	0.005	0.000
Unemployed	0.002	0.005	0.003	0.020	0.028	0.000
Homemaker	0.000	0.002	0.004	0.001	0.004	0.000
Missing	0.232	0.287	0.000	0.231	0.213	0.000
Mother's occupation						
Executive / senior official / university teacher / officer	0.080	0.086	0.137	0.016	0.018	0.000
Professional (employee or self-employed) / NCO	0.256	0.257	0.946	0.091	0.106	0.000
Clerical/teacher (graduate) / graduated military	0.000	0.147	0.000	0.000	0.312	0.000
Clerical/teacher (non-graduate) / enlisted	0.138	0.000	0.000	0.312	0.000	0.000
Self-employed (trader, artisan, farmer...)	0.060	0.053	0.028	0.049	0.047	0.000
Agricultural entrepreneur / farm owner	0.038	0.036	0.349	0.011	0.012	0.000
Manual worker / services / cooperative member	0.026	0.019	0.000	0.121	0.114	0.000
Retired	0.000	0.001	0.087	0.001	0.001	0.000
Unemployed	0.005	0.010	0.002	0.027	0.040	0.000
Homemaker	0.173	0.108	0.000	0.161	0.154	0.000
Missing	0.223	0.284	0.000	0.212	0.196	0.000
Origin						
Italian	0.927	0.881	0.000	0.852	0.805	0.000
Foreign (1st generation)	0.015	0.015	0.991	0.051	0.042	0.000
Foreign (2nd generation)	0.054	0.059	0.137	0.089	0.123	0.000
Missing	0.004	0.045	0.000	0.008	0.030	0.000
Nursery						
Yes	0.323	0.430	0.000	0.246	0.305	0.000
No	0.301	0.418	0.000	0.391	0.365	0.000
Missing	0.376	0.152	0.000	0.363	0.330	0.000
Kindergarten						
Yes	0.756	0.679	0.000	0.873	0.729	0.000
No	0.009	0.190	0.000	0.020	0.060	0.000
Missing	0.235	0.131	0.000	0.107	0.210	0.000
Track / regularity						
Regular / early entry	0.987	0.968	0.000	0.969	0.969	0.817
Delayed	0.012	0.010	0.059	0.027	0.026	0.009
Missing	0.000	0.022	0.000	0.004	0.005	0.000

Notes: The table reports the distribution of student covariates before and after the implementation of Area C, separately for schools located inside Area C and for schools in the wider Milan metropolitan area. "Before" refers to years prior to the introduction of Area C (2010–2011), and "After" refers to subsequent years (2012–2019). The table also reports p-values for tests of equality between before and after means within each area.

Table 4: Average standardized test scores before and after the implementation of Area C

	Area C (Before)	Area C (After)	p-value	Met. Area (Before)	Met. Area (After)	p-value
Grade 2 – Math	0.064	0.305	0.000	-0.002	-0.012	0.078
Grade 2 – Italian	0.213	0.321	0.000	-0.007	-0.012	0.349
Grade 5 – Math	0.133	0.249	0.000	-0.004	-0.010	0.292
Grade 5 – Italian	0.279	0.325	0.077	-0.009	-0.013	0.426

Notes: The table reports mean standardized test scores in Mathematics and Italian for Grades 2 and 5, separately for schools located inside Area C and for schools in the wider metropolitan area. "Before" refers to years prior to the introduction of Area C (2010–2011), and "After" refers to subsequent years (2012–2019). Standard errors are shown in parentheses.

Table 5: Effect of Milan's area C congestion policy on standardized test scores: baseline DID estimates

	Grade 2		Grade 5	
	Math	Italian	Math	Italian
Area C × Post-treatment period	0.256***	0.116*	0.142	0.073
	(0.085)	(0.063)	(0.115)	(0.080)
Year fixed effects	X	X	X	X
School fixed effects	X	X	X	X
Covariates	X	X	X	X
N	213,360	213,486	204,828	205,222

Notes: The table reports baseline difference-in-differences estimates of the effect of Area C on standardized test scores in Grades 2 and 5. All specifications include year and school fixed effects, as well as a rich set of student-level covariates: parental occupation (mother and father), gender, migration background, early childhood attendance (nursery and kindergarten), and grade regularity. Standard errors are clustered at the school level and reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Synthetic difference-in-differences estimates for the impact of Milan's area C congestion policy on standardized test scores

	Grade 2		Grade 5	
	Math	Italian	Math	Italian
Area C × Post-treatment period	0.219**	0.022	0.080	0.031
	(0.094)	(0.074)	(0.086)	(0.078)
Covariates	Yes	Yes	Yes	Yes
Balanced panel	Yes	Yes	Yes	Yes
Inference	Placebo	Placebo	Placebo	Placebo
Placebo reps	1,000	1,000	1,000	1,000
Schools	227	224	232	228
School-years	2,270	2,240	2,320	2,280

Notes: This table reports SDID estimates of the effect of Area C on standardized test scores in Grades 2 and 5. All specifications include the same rich set of student-level covariates as in the baseline DID models: parental occupation (mother and father), gender, migration background, early childhood attendance (nursery and kindergarten), and grade regularity. The SDID estimator is implemented on a balanced panel at the school level, and inference is based on placebo (permutation) tests with 1,000 random reassignments of treatment status. Standard errors are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Robustness of difference-in-differences estimates – Grade 2 Mathematics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Area C × Post-treatment period	0.253***	0.256***	0.249***	0.256***	0.252***	0.265***	0.215**	0.282***	0.276***	0.148**	0.232***
	(0.086)	(0.085)	(0.068)	(0.085)	(0.094)	(0.085)	(0.087)	(0.086)	(0.084)	(0.073)	(0.076)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class FE	No	No	Yes	No	No	No	No	No	No	No	No
Covariates	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balanced panel	No	No	No	No	Yes	No	No	No	No	No	No
Wild bootstrap	No	No	No	Yes	No	No	No	No	No	No	No
MICE	No	No	No	No	No	Yes	No	No	No	No	No
Excl. outbound	No	No	No	No	No	No	Yes	No	No	No	No
Excl. surrounding	No	No	No	No	No	No	No	Yes	No	No	No
De-std. outcome	No	No	No	No	No	No	No	No	Yes	No	No
School trends	No	No	No	No	No	No	No	No	No	Yes	No
PSM-DID	No	No	No	No	No	No	No	No	No	No	Yes
Observations	213,360	213,360	213,360	213,360	171,247	211,923	84,904	136,129	213,360	213,360	212,719

Notes: The dependent variable is the standardized Grade 2 Mathematics test score. Each column reports the coefficient on the interaction between Area C and the post-treatment period from a separate difference-in-differences specification. Column (1) presents the baseline model with school and year fixed effects and no individual covariates. Column (2) adds the full set of student-level covariates. Column (3) replaces school fixed effects with class fixed effects. Column (4) reports estimates from a specification identical to column (2), but inference is based on wild cluster bootstrap-t with 1,000 repetitions and clustering at the school level. Column (5) restricts the sample to a balanced panel of schools observed in all ten years. Column (6) uses multiple imputation by chained equations with 20 imputations for missing background variables, and re-estimates the baseline model with covariates. Column (7) excludes schools located in the outbound zone. Column (8) excludes schools located in the area immediately surrounding Area C. Column (9) uses a de-standardized outcome obtained by dropping the year dimension from the grade–subject–year standardization. Column (10) augments the baseline specification with school-specific linear time trends. Column (11) reports a propensity-score-matched DID, where treated schools are matched to controls using kernel matching with trimming of extreme propensity scores. Robust standard errors clustered at the school level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Difference-in-Differences Estimates – Grade 2 Italian

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Area C × Post-treatment period	0.112*	0.116*	0.114**	0.116*	0.089	0.132**	0.085	0.135**	0.122*	0.064	0.115**
	(0.058)	(0.063)	(0.056)	(0.063)	(0.068)	(0.061)	(0.065)	(0.064)	(0.065)	(0.068)	(0.057)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class FE	No	No	Yes	No	No	No	No	No	No	No	No
Covariates	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balanced panel	No	No	No	No	Yes	No	No	No	No	No	No
Wild bootstrap	No	No	No	Yes	No	No	No	No	No	No	No
MICE	No	No	No	No	No	Yes	No	No	No	No	No
Excl. outbound	No	No	No	No	No	No	Yes	No	No	No	No
Excl. surrounding	No	No	No	No	No	No	No	Yes	No	No	No
De-std. outcome	No	No	No	No	No	No	No	No	Yes	No	No
School trends	No	No	No	No	No	No	No	No	No	Yes	No
PSM-DID	No	No	No	No	No	No	No	No	No	No	Yes
Observations	213,486	213,486	213,486	213,486	170,139	212,138	84,956	136,295	213,486	213,486	212,837

Notes: The dependent variable is the standardized Grade 2 Italian test score. Each column reports the coefficient on the interaction between Area C and the post-treatment period from a separate difference-in-differences specification. Column (1) presents the baseline model with school and year fixed effects and no individual covariates. Column (2) adds the full set of student-level covariates. Column (3) replaces school fixed effects with class fixed effects. Column (4) reports estimates from a specification identical to column (2), but inference is based on wild cluster bootstrap-t with 1,000 repetitions and clustering at the school level. Column (5) restricts the sample to a balanced panel of schools observed in all ten years. Column (6) uses multiple imputation by chained equations with 20 imputations for missing background variables, and re-estimates the baseline model with covariates. Column (7) excludes schools located in the outbound zone. Column (8) excludes schools located in the area immediately surrounding Area C. Column (9) uses a de-standardized outcome obtained by dropping the year dimension from the grade–subject–year standardization. Column (10) augments the baseline specification with school-specific linear time trends. Column (11) reports a propensity-score-matched DID, where treated schools are matched to controls using kernel matching with trimming of extreme propensity scores. Robust standard errors clustered at the school level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Difference-in-Differences Estimates – Grade 5 Mathematics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Area C × Post-treatment period	0.120	0.142	0.154	0.142	0.155	0.135	0.111	0.158	0.165	0.128	0.108
	(0.111)	(0.115)	(0.116)	(0.115)	(0.124)	(0.110)	(0.117)	(0.115)	(0.102)	(0.100)	(0.109)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class FE	No	No	Yes	No	No	No	No	No	No	No	No
Covariates	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balanced panel	No	No	No	No	Yes	No	No	No	No	No	No
Wild bootstrap	No	No	No	Yes	No	No	No	No	No	No	No
MICE	No	No	No	No	No	Yes	No	No	No	No	No
Excl. outbound	No	No	No	No	No	No	Yes	No	No	No	No
Excl. surrounding	No	No	No	No	No	No	No	Yes	No	No	No
De-std. outcome	No	No	No	No	No	No	No	No	Yes	No	No
School trends	No	No	No	No	No	No	No	No	No	Yes	No
PSM-DID	No	No	No	No	No	No	No	No	No	No	Yes
Observations	204,828	204,828	204,828	204,828	166,494	204,141	81,873	130,455	204,828	204,828	204,097

Notes: The dependent variable is the standardized Grade 5 Mathematics test score. Each column reports the coefficient on the interaction between Area C and the post-treatment period from a separate difference-in-differences specification. Column (1) presents the baseline model with school and year fixed effects and no individual covariates. Column (2) adds the full set of student-level covariates. Column (3) replaces school fixed effects with class fixed effects. Column (4) reports estimates from a specification identical to column (2), but inference is based on wild cluster bootstrap-t with 1,000 repetitions and clustering at the school level. Column (5) restricts the sample to a balanced panel of schools observed in all ten years. Column (6) uses multiple imputation by chained equations with 20 imputations for missing background variables, and re-estimates the baseline model with covariates. Column (7) excludes schools located in the outbound zone. Column (8) excludes schools located in the area immediately surrounding Area C. Column (9) uses a de-standardized outcome obtained by dropping the year dimension from the grade–subject–year standardization. Column (10) augments the baseline specification with school-specific linear time trends. Column (11) reports a propensity-score-matched DID, where treated schools are matched to controls using kernel matching with trimming of extreme propensity scores. Robust standard errors clustered at the school level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Difference-in-Differences Estimates – Grade 5 Italian

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Area C × Post-treatment period	0.054	0.073	0.084	0.073	0.088	0.078	0.041	0.087	0.111	0.067	0.040
	(0.073)	(0.080)	(0.085)	(0.080)	(0.085)	(0.075)	(0.081)	(0.081)	(0.069)	(0.070)	(0.077)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class FE	No	No	Yes	No	No	No	No	No	No	No	No
Covariates	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balanced panel	No	No	No	No	Yes	No	No	No	No	No	No
Wild bootstrap	No	No	No	Yes	No	No	No	No	No	No	No
MICE	No	No	No	No	No	Yes	No	No	No	No	No
Excl. outbound	No	No	No	No	No	No	Yes	No	No	No	No
Excl. surrounding	No	No	No	No	No	No	No	Yes	No	No	No
De-std. outcome	No	No	No	No	No	No	No	No	Yes	No	No
School trends	No	No	No	No	No	No	No	No	No	Yes	No
PSM-DID	No	No	No	No	No	No	No	No	No	No	Yes
Observations	205,222	205,222	205,222	205,222	165,201	204,514	81,968	130,888	205,222	205,222	204,484

Notes: The dependent variable is the standardized Grade 5 Italian test score. Each column reports the coefficient on the interaction between Area C and the post-treatment period from a separate difference-in-differences specification. Column (1) presents the baseline model with school and year fixed effects and no individual covariates. Column (2) adds the full set of student-level covariates. Column (3) replaces school fixed effects with class fixed effects. Column (4) reports estimates from a specification identical to column (2), but inference is based on wild cluster bootstrap-t with 1,000 repetitions and clustering at the school level. Column (5) restricts the sample to a balanced panel of schools observed in all ten years. Column (6) uses multiple imputation by chained equations with 20 imputations for missing background variables, and re-estimates the baseline model with covariates. Column (7) excludes schools located in the outbound zone. Column (8) excludes schools located in the area immediately surrounding Area C. Column (9) uses a de-standardized outcome obtained by dropping the year dimension from the grade–subject–year standardization. Column (10) augments the baseline specification with school-specific linear time trends. Column (11) reports a propensity-score-matched DID, where treated schools are matched to controls using kernel matching with trimming of extreme propensity scores. Robust standard errors clustered at the school level are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

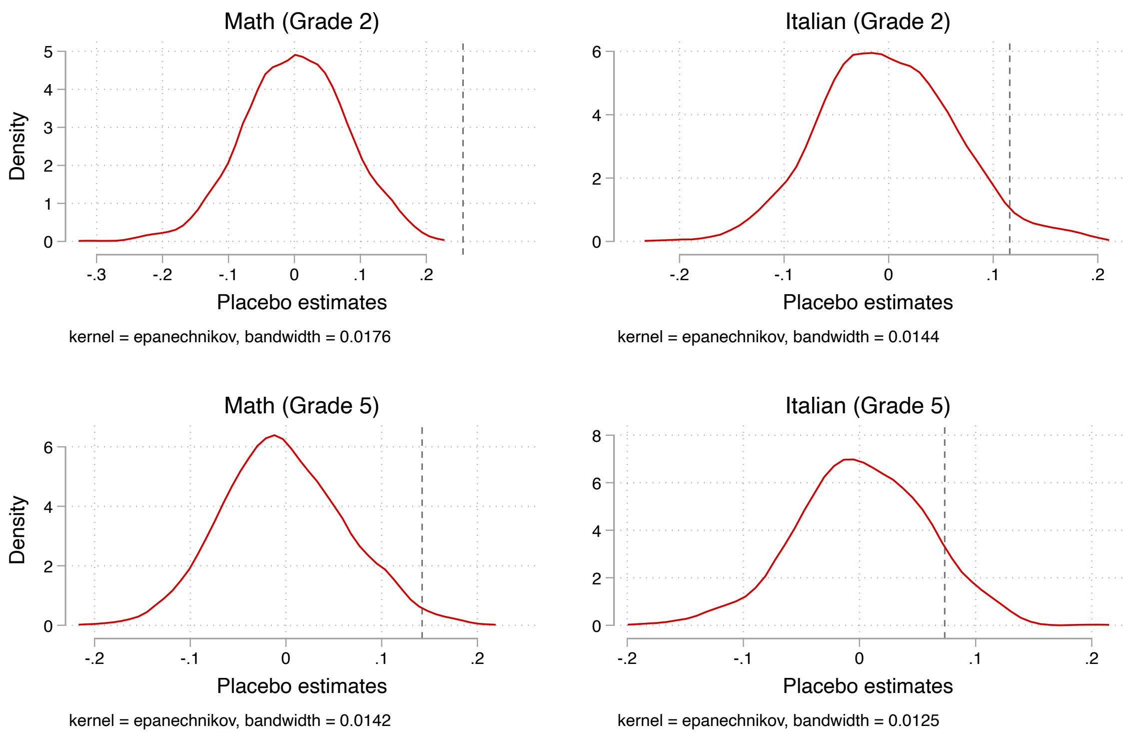

Figure A.1: Randomization inference distributions of treatment effects

Notes: Each panel reports the randomization inference distribution of estimates for the corresponding subject–grade combination (Mathematics and Italian in Grades 2 and 5). The curve shows the density of the test statistic under the null hypothesis generated by repeated reassignments of the treatment indicator. The vertical dashed line marks the estimate obtained under the actual treatment assignment.

Table A.1: Placebo estimates

	Grade 2		Grade 5	
	Math	Italian	Math	Italian
Panel A. Placebo on treated units				
Placebo treated schools × Post-treatment period	-0.057	0.048	-0.064	-0.052
	(0.062)	(0.059)	(0.071)	(0.060)
Year fixed effects	X	X	X	X
School fixed effects	X	X	X	X
Covariates	X	X	X	X
N	205,687	205,721	197,328	197,588
Panel B. Placebo on timing				
Area C × Placebo post-treatment period	0.086	-0.102	0.033	-0.076
	(0.124)	(0.088)	(0.147)	(0.128)
Year fixed effects	X	X	X	X
School fixed effects	X	X	X	X
Covariates	X	X	X	X
N	46,874	47,468	45,182	45,813

Notes: This table reports placebo estimates used to assess the validity of the difference-in-differences identification strategy. Panel A presents estimates from a placebo test on treated units, in which treatment status is randomly reassigned among schools in the sample. Panel B presents estimates from a placebo test on timing, in which the post-treatment period is shifted to a placebo date prior to the actual introduction of Area C in January 2012. All specifications include year and school fixed effects, as well as the same rich set of student-level covariates as in the baseline DID models: parental occupation (mother and father), gender, migration background, early childhood attendance (nursery and kindergarten), and grade regularity. Standard errors are clustered at the school level and reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Effect of Milan's Area C congestion policy on standardized test scores: robustness check excluding year 2019

	Grade 2		Grade 5	
	Math	Italian	Math	Italian
Area C × Post-treatment period	0.257***	0.120*	0.150	0.085
	(0.086)	(0.065)	(0.115)	(0.080)
Year fixed effects	X	X	X	X
School fixed effects	X	X	X	X
Covariates	X	X	X	X
N	193,982	194,193	184,672	185,222

Notes: The table replicates the baseline difference-in-differences estimates excluding the year 2019, in which Area B — the broader LEZ encompassing Area C — was introduced (February 2019). Since INVALSI tests for the 2018/2019 academic year were administered in spring 2019, the effective exposure window to Area B is limited to a few weeks. All specifications include year and school fixed effects, as well as a rich set of student-level covariates: parental occupation (mother and father), gender, migration background, early childhood attendance (nursery and kindergarten), and grade regularity. Standard errors are clustered at the school level and reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Synthetic difference-in-differences estimates for the impact of Milan's Area C congestion policy on standardized test scores: jackknife robustness check

	Grade 2		Grade 5	
	Math	Italian	Math	Italian
Area C $\times$ Post-treatment period	0.219**	0.022	0.080	0.031
	(0.094)	(0.072)	(0.123)	(0.108)
Covariates	Yes	Yes	Yes	Yes
Balanced panel	Yes	Yes	Yes	Yes
Inference	Jackknife	Jackknife	Jackknife	Jackknife
Schools	227	224	232	228
School-years	2,270	2,240	2,320	2,280

Notes: This table replicates the SDID estimates reported in Table 6 using jackknife inference in place of placebo-based standard errors. All specifications include the same rich set of student-level covariates as in the baseline DID models: parental occupation (mother and father), gender, migration background, early childhood attendance (nursery and kindergarten), and grade regularity. Standard errors are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.