Tech Boom or Bust?
The Effect of Data Center Expansion on Student Achievement

Samantha B. Kane, Matthew A. Kraft, and Christopher Cleveland

Abstract

Between 2022 and 2025, the United States experienced a nearly 200 percent growth in the construction of data centers, which can be substantial sources of air, noise, and water pollution. This paper estimates the effects on student achievement of attending a public elementary school within close proximity to a data center where the risk of potential exposure to pollution is most acute. We leverage the staggered timing of new data center openings within a differences-in-differences framework and find that data centers cause student achievement among third-graders to decline by an average of 0.22 school-level standard deviations in math (equivalent to a 0.10 student-level standard deviation decline) beginning six years after the initial air pollution permit is issued. Effects in ELA are similarly negative, but less precisely estimated. We explore the mechanisms behind this delayed onset of treatment after initial permitting including operational delays and slow ramp-ups, cumulative exposure effects, and exposure during sensitive periods of young children's development. We also find suggestive evidence of a dosage effect where students attending schools located near multiple data centers experience even larger declines.

1 Introduction

Data centers, the physical facilities that house computer systems and related hardware, have long been a core feature of the technological infrastructure that underlies the modern internet. In recent years, cloud computing and the advent of artificial intelligence (AI) have created intense demand for data centers, with construction across the United States increasing 190 percent since the launch of ChatGPT in 2022 (Pawlowski, 2025). Where companies site new data centers has become a contentious issue because of the perceived benefits and negative externalities they produce. Data

centers can be attractive to communities because these large capital projects create new construction activity, increase property-tax revenue, and offer specialized employment opportunities (JLARC, 2024). On the other hand, data centers can also have large environmental footprints, consuming immense amounts of energy and water, and generating considerable air, noise, and water pollution (Cassauwers, 2026; Han et al., 2024; Nicoletti, Ma, and Bass, 2025; Saul et al., 2025; Siddik, Shehabi, and Marston, 2021; Smith-Schoenwalder, 2026).

One of the primary sources of data center pollution stems from the industry's power demands. To maintain a "five nines" reliability standard (remaining online 99.999 percent of the time) developers must secure both primary power and backup systems. While connecting to the grid contributes to off-site pollution, on-site power sources affect data centers' local environmental impact (Martin, 2025). Facilities typically host multiple diesel generators for maintenance testing and emergency backup power. However, developers are increasingly using these generators and other technologies, such as mobile gas turbines, to meet peak demand or even serve as primary power sources (Ireland, 2026; Martin, 2025; Robb, 2026; Sadasivam and Bittle, 2026). This surge in demand is reflected in the market, with Rolls-Royce reporting that sales of power generation products for the data center segment grew nearly 50 percent in 2024 (Robb, 2026). These technologies are highly polluting. Diesel combustion, in particular, emits nitrogen oxides and other pollutants that jeopardize human health and the environment, while the constant operation of these units has triggered community concerns regarding noise pollution (Ireland, 2026; Yañez-Barnuevo, 2026).

In addition to staying online, data centers need to stay cool. Continuously operating servers inside data centers generate considerable amounts of heat as a byproduct. To prevent overheating, facilities rely on cooling systems and high-velocity fans that are sources of persistent noise pollution (Yañez-Barnuevo, 2026; JLARC, 2024). Data centers can also use a toxic family of chemicals known as PFAS (per- and polyfluoroalkyl substances)—referred to as "forever chemicals" because they do not naturally break down in the environment—throughout their infrastructure. PFAS chemicals are used across multiple states of matter: as gases for refrigerant cooling and fire suppression, as fluids in two-phase immersion cooling, and as solids within hardware components (Cassauwers, 2026; Hardin, 2026; Perkins, 2025). Although often circulated in closed-loop systems, PFAS gases can escape into the atmosphere, where they accelerate climate change, while liquid leakages and wastewater discharge can contaminate local water supplies (Cassauwers, 2026; Hardin, 2026). Beyond chemical contamination, the volume of water required for cooling further strains local communities, often leading to drought-like conditions (Copley, 2022; Skibell, 2026). Given these externalities, the specific location of data centers expansion is of crucial importance.

In this paper, we seek to inform understanding of the economic consequences of data centers as

well as data center siting policy by providing a fuller understanding of their localized consequences for students' human capital development. We focus on children because their developing bodies make them particularly susceptible to air, noise, and water pollution (Brumberg et al., 2021). A large literature documents the adverse effects that exposure to pollution has for children in utero and in early childhood. Negative effects include delays in cognitive development, asthma, elevated blood pressure as well as low academic performance, attendance, and behavior (Ames, Sharma, and Lyall, 2025; Brumberg et al., 2021; Chung, Persico, and Liu, 2025; Heissel, Persico, and Simon, 2022; Klatte, Bergström, and Lachmann, 2013; Persico and Venator, 2021; S. Stansfeld and Clark, 2015; S. A. Stansfeld et al., 2005). Emerging research on data centers has largely concentrated on measuring their effects on local labor and housing markets (Alvarez et al., 2026; Seagraves and Sirmans, 2026) and their aggregate monetized damages from pollution and greenhouse gases (Muller, 2026).

We estimate the effect of data center expansions on the academic achievement of young students who attend elementary schools located within close proximity of a new data center. We first draw on data and motivating descriptive evidence from Virginia—home to the world's largest concentration of data centers—to characterize the relationship between data centers and pollution. In Virginia, one-third of facilities are near residential areas (JLARC, 2024). We then use national data on data centers and third grade test scores to estimate the causal effect of data center expansion on achievement in nearby elementary schools. We use data from *Business Insider* on the location and timing of data centers built or approved for construction by the end of 2025, and combine these data with school-level math and ELA assessment results from the Education Data Center's State Assessment Data Repository.

The central identification challenge is that data center locations are not random: firms identify potential sites based on local regulation, power, land, taxes, and fiber connectivity, and communities make decisions based on their priorities and local contexts. We address this issue through a staggered differences-in-differences design leveraging the exogenous timing of a facility's first air permit, and analyze test score outcomes at schools within one mile of a newly permitted data center to those at comparison schools located between one and two miles away. This design serves to plausibly isolate the localized pollution effects of the facility from both the endogenous factors that drive where it is built as well as from any broad economic impacts of data centers on the communities where they are located.

We find that opening a new data center within one mile of a school causes an average decline of 0.22 school-level standard deviations in third-grade math scores beginning six years after the first permit is issued. This is equivalent to a 0.10 student-level standard deviation decline. Effects

on third-grade ELA scores are also negative, but only marginally significant at the 10 percent level (-0.14 and -0.06 standard deviations at the school and student levels, respectively). We also observe a dosage effect: negative impacts on achievement are larger for schools near multiple facilities, particularly in math. This timing could be explained by several potential mechanisms. Data centers often take one to three years to move from permit to operation, with operational capacity and emissions ramping up over several years and secondary facilities coming online in subsequent years. Delayed effects are also consistent with effects that emerge only after prolonged cumulative exposure to pollution over multiple years or during more sensitive periods of child development when students are enrolled in preschool and kindergarten.

We make three contributions through this project. First, this research builds on work on pollution and human capital formation. The negative impact of air pollution on student outcomes is well documented: exposure on test days affects day-of performance (Ebenstein, Lavy, and Roth, 2016; Marcotte, 2017), daily variation affects day-to-day academic and behavioral measures (Chung, Persico, and Liu, 2025), and prolonged exposure affects annual and longer-term educational outcomes (Aizer et al., 2018; Heissel, Persico, and Simon, 2022; Rau, Urzúa, and Reyes, 2015) and even criminal behavior (Grönqvist, Nilsson, and Robling, 2020). A parallel literature links noise to comprehension and test performance, including recent causal evidence (Aarons, 2026; Haines et al., 2001; Klatte, Bergström, and Lachmann, 2013; Matheson et al., 2010; S. A. Stansfeld et al., 2005). Closest in method, Ghafarianghadim and Heutel (2026) use thermal inversions to identify the effect of particulate pollution on preschool attendance. While prior literature has examined effects of air and noise pollution on educational performance, our paper is, to our knowledge, the first to examine the effects of pollution created by data centers on student achievement.

Second, our findings are related to the broader literature on the impact of pollution on children's health. Maternal exposure to air pollution during pregnancy is linked to negative birth outcomes, such as mortality and low birth weight (Alexander and Schwandt, 2022; Currie and Neidell, 2005; Knittel, Miller, and Sanders, 2016). Work by Baran et al. (2025) finds that the expansion of electric vehicle adoption improved air quality and measures of infant health. Beyond the prenatal period, research links outdoor air pollution to asthma attacks in children (Roy et al., 2011; Simeonova et al., 2021) and school bus emissions to student respiratory health (Austin, Heutel, and Kreisman, 2019; Beatty and Shimshack, 2011). Noise exposure during childhood is also detrimental, triggering stress responses and lowering well-being (S. Stansfeld and Clark, 2015). Early-life exposure to PFAS is linked to reduced cognitive, motor, and language development in infancy, as well as increased behavioral issues in childhood (Ames, Sharma, and Lyall, 2025). Our estimates capture the reduced-form effect of *exposure to data centers*. However, proximity to data centers plausibly

increases potential exposure to pollution such that negative effects on student achievement likely operate through these health channels.

Third, we add a human capital cost to the emerging economics of data centers and energy infrastructure externalities. The effects we estimate are local, but they sit atop a separate, global externality. Generating the electricity that powers U.S. data centers produces an estimated $25 billion in annual climate and air-pollution damages, geographically concentrated in a few states (Muller, 2026), and the PFAS used in cooling and fire suppression are long-lived, high-global-warming-potential gases when released (Cassauwers, 2026; Hardin, 2026; Perkins, 2025). Warming, in turn, depresses learning itself: hotter school years lower the rate of skill formation both across countries and within the United States (Park, Behrer, and Goodman, 2021), an effect concentrated among disadvantaged students that air conditioning only partly offsets (Park, Goodman, et al., 2020). The increase in data centers' climate footprint therefore compounds the local costs we estimate and ultimately impacts student achievement. This is a human capital cost that current aggregate damage accounting omits (Muller, 2026).

The rest of the paper proceeds as follows. In Section 3.1, we discuss the data, provide descriptive evidence on data center pollution in Virginia, and describe our national sample. In Section 3.2, we describe the empirical approach. Section 4 reports the main results, as well as results from tests of heterogeneity and threats to internal validity. Finally, Section 6 concludes.

2 Motivating Evidence

School proximity to data centers is not just a conceptual concern. Using data from our main analysis (see Section 3.1 and Appendix B), we estimate that there are 1,878 public schools across the United States within one mile of a data center that is built or approved for construction by the end of 2025. This figure translates to approximately 2 percent of all U.S. public schools and more than 980,000 students who are at heightened risk of potential exposure to the pollution generated by data centers.

Evidence from Virginia underscores the basis for pollution as the primary mechanism in our analysis. In 2026, the Virginia Department of Environmental Quality (DEQ) initiated an air monitoring project to measure ambient concentrations of carbon monoxide (CO), nitrogen dioxide (NO_2), and fine particulate matter ($PM_{2.5}$) within Loudoun County's "Data Center Alley" (*Data Center Air Monitoring Project* 2026). Between March and July 2026, the mean daily $PM_{2.5}$ concentrations at more than half of the monitoring sites exceeded the World Health Organization's (WHO) annual average health guideline of 5.0 $\mu g/m^3$ (*WHO global air quality guidelines* 2021). Furthermore, daily mean concentrations at every other monitoring site exceeded those recorded

at the Washington Dulles Airport, which is located upwind of "Data Center Alley", with some short-term hourly $PM_{2.5}$ concentrations exceeding the WHO's 24-hour threshold of 15 $\mu g/m^3$ on peak days. Similarly, mean hourly NO_2 concentrations across the data center corridor pushed up against the WHO annual mean target of 10 $\mu g/m^3$ ($\approx$ 5.3 parts per billion), while CO concentrations remained well within the guidelines. Notably, the EPA's National Ambient Air Quality Standards (NAAQS)—against which the DEQ benchmarks its findings—rely on significantly more lenient exposure thresholds than the WHO.

Publicly available data on emissions from data centers in Virginia provide further evidence that pollution is likely a primary mechanism. The DEQ issues air emissions permits—which data centers are required to obtain for construction, operation, and expansion—and provides annual emissions estimates reported by permitted facilities, which are indexed by a registration number. We link data center registration numbers—provided to us directly by the DEQ—to their annual emissions records from 2015 through 2025. An air permit associated with a registration number may contain multiple facility addresses but only one emissions record, and records are available for 90 percent of the data center registration numbers provided to us by the state.[1,2]

Figure 2A shows that nitrogen oxides (NO_x)—which are particularly harmful to respiratory health, especially for children (US EPA, 2016)—make up the majority of emissions across VA data centers, followed by carbon monoxide (CO) and several other pollutants, including volatile organic compounds (VOCs), particulate matter (PM_{10} and $PM_{2.5}$), sulfur dioxide (SO_2), and ammonia (NH_3). Emissions have increased over time as the number of data centers in the state has grown. In 2025, facilities in Virginia emitted more than 2,000 tons of pollutants, or an average of 13 tons per permit. Importantly, because NO_x, SO_2, and NH_3 are precursors to inorganic $PM_{2.5}$ (Holt, Selin, and Solomon, 2015), these direct emissions are a plausible contributor to the ambient $PM_{2.5}$ concentrations that the DEQ has documented.

We leverage these data to explore the timing and magnitude of emissions released from data centers relative to the date of their initial permit. We focus on a subset of air permits that correspond

[1] According to state permit guidelines, data centers typically use Tier II diesel engine driven generator sets for backup power, a type of engine technology that is the largest source of nitrogen oxides (NO_x), which includes nitrogen oxides and nitrogen dioxides (NO and NO_2). DEQ requires that data centers operate their diesel engines on a routine basis to ensure that they are in good operating condition and emit within the established permit limits (Environmental Quality, 2025). The total amount of hours an emergency generator can operate is limited to 500 hours per year, including testing and maintenance (*Air Permits* 2026).

[2] We focus on on the most direct source of emissions from data centers (Scope 1), which are emissions from the maintenance and operation of backup generators. There are two other types of emissions: Scope 2 and Scope 3. Scope 2 emissions are indirect emissions from a data center, such as emissions from the purchase of electricity. Scope 3 emissions are indirect emissions from a data center's value chain, such as emissions from the manufacturing of computer hardware.

with one facility address and were issued in or after 2015.[3] In Figure 2B, we plot a time series of aggregate total emissions per facility relative to initial permit. The figure shows an increase in mean emissions over time relative to the first year an air permit was issued. On average, a registration number reports between 10 and 16 tons of air emissions a year, starting one year after its first permit was issued. These air emissions are on par with those from facilities that are required to report to the EPA's Toxics Release Inventory (TRI) Program. In 2023, for example, over 21,000 TRI facilities reported 523 million pounds of air releases, which is an average of 24,000 pounds (12 tons) of hazardous chemicals per facility a year.[4]

Environmental impacts also extend beyond air quality. A 2024 report by the Joint Legislative Audit and Review Commission details Virginia resident accounts of a constant "drone" or "hum" from cooling systems that can be heard both in and outside of nearby residences, causing migraines, health problems from disrupted sleep, and an inability to concentrate on tasks. Residents also report noise disturbances from data center construction, including loud construction activities and vehicle traffic. In addition, some data centers use substantial amounts of water in their cooling systems, which involve the use of a toxic family of "forever chemicals" that other reporting suggests can contaminate local water supplies (Cassauwers, 2026; Hardin, 2026). These accounts lend support to the relevance of data centers' pollution pathways, which may significantly affect the well-being of local residents and students. Indeed, Han et al. (2024) estimate that data centers in Virginia could cause 14,000 asthma symptom cases, among other health outcomes, assuming that actual emissions from the backup generators are only 10 percent of the permitted level. The combined evidence from Virginia suggests that pollution from data centers is a nontrivial, negative environmental externality for nearby communities.

3 Methods

3.1 Data and Descriptive Statistics

We draw on a range of data sources to estimate the effect of data centers on student achievement at the national level. We start by using data on the location and timing of data centers from *Business Insider*, a news outlet that compiled a national accounting of data center air permits in 45 states and

[3] We identify facility addresses and air permit timing by manually linking the data center registration numbers provided by the VA DEQ to our data from *Business Insider*, which we detail in Section 3. We supplement this linking process using the VA DEQ's "Issued Air Permits for Data Centers" online database. We restrict the analysis to the data center registration numbers that appear in the annual emissions records, and whose first emissions record does not appear before the year of its initial air permit.

[4] These summary statistics were calculated using the TRI "Basic" data file from 2023. Air releases includes both fugitive air emissions and stack air emissions.

Washington, D.C. To create this dataset, *Business Insider* requested all air permits issued to data center facilities by state environmental agencies and local air districts using North American Industry Classification System (NAICS) and Standard Industrial Classification (SIC) codes associated with data centers. Permits may contain multiple addresses for data center facilities—observations are uniquely identified by permit and address. Figure 1 shows the significant growth in data centers over time—by 2025,1,416 facilities were built or approved for construction in the United States, with the greatest number in Virginia followed by California.[5]

We use public school data maintained by the NCES Common Core of Data (CCD) for three primary purposes in our analysis. First, we use the CCD to obtain a list of all U.S. public schools and their geocoordinates. Second, we use the CCD to obtain school-level covariates on student demographic composition, including grade-level enrollment by race/ethnicity and school-level enrollment by free-or-reduced-price lunch status. We also supplement these measures with zip-code-level median household income from the American Community Survey (ACS) 5-Year Estimates, matched to schools by zip code. Third, we use the CCD to exclude a small number of virtual schools from the sample and to retain only physical schools.[6]

Finally, we use national school-level assessment data from the Education Data Center's (EDC) State Assessment Data Repository, which aggregates publicly available state assessment data from all 50 states and the District of Columbia.[7] We obtain school-level average scale scores in English Language Arts (ELA) and math for schools with assessment data in grades three to five from the 2014–15 through 2024–25 school years. We standardize these scores by state, school year, grade level, and subject to account for assessment differences between states as well as within-state changes in assessments over time. The aggregate nature of these data create a possibility that any effects we find are driven by endogenous student mobility. We explore this potential threat in Section 5.2 and find little evidence of dynamic changes in student composition.

We focus our primary analyses on third-grade test scores because it is the earliest grade in which standardized testing results are widely available. Research documents how younger students are uniquely sensitive to environmental health hazards (Bearer, 1995) because of their developing physiological systems and higher air per body weight intake (Brumberg et al., 2021). Aizer et al. (2018), for instance, find that early exposure to lead in preschool-aged children has negative impacts

[5]This figure reflects the total number of observations reported by *Business Insider*. See Appendix B for more detail on these data.

[6]We exclude schools that NCES designates as "full virtual" (defined as exclusively virtual) or "face virtual" (defined as primarily virtual). NCES defines a "regular school" as a public elementary/secondary school that does not focus primarily on vocational, special, or alternative education.

[7]We use the EDC State Assessment Data Repository, Version 3.1.

third grade test scores. By focusing on this group, we capture the potential long-term academic consequences of environmental exposure during early elementary school highly malleable period of cognitive development. For completeness, we also show results for fourth and fifth grade tests scores in Section 5.4. We restrict the analytic sample to schools with at least nine years of third-grade assessment data in each subject to maintain a relatively balanced panel. While the 2019–2020 school year is missing for all schools due to COVID-19, we permit one additional year of missing data in either or both subjects. The majority of schools with nine years of assessment data are missing data from the 2020–2021 school year, and our main results are not sensitive to dropping assessment results from this year from the panel.

Using these data, we construct a school-by-year panel that identifies if and when a school is located within one or two miles of a data center, using the year of the initial permit as a proxy for the opening date. We conservatively assign potential exposure beginning with the school year whose spring term falls in the year of the permit. For example, a permit issued anywhere in calendar year 2016 is matched to the 2015–16 school year, ensuring that a school is coded as exposed in the earliest year consistent with the permit—capturing cases where, for instance, a permit was issued in the fall of 2016 after the end of the 2015-16 school year. We do this by first using the detailed geographic coordinates of the data center facilities from the *Business Insider* data and all U.S. public schools from the CCD to calculate the distance between each facility and school.[8] We then identify potential exposure to multiple vectors of pollution using radial distance given the flexibility this approach provides at capturing air, noise, and water pollution.

Based on prior research on air pollution, we expect that the concentration of pollutants will be greatest within a mile of a data center and decrease between one to two miles of a data center (Currie, Davis, et al., 2015; Anderson, 2020; Persico and Venator, 2021). We therefore designate a school as part of the treatment group if it is within one mile of the closest data center. In all cases, we define treatment timing based on the year the first permit is issued among those facilities.

Next, we identify a group of local schools to serve as a comparison group. A key challenge in this context is that data centers may generate economic externalities——such as from shorter-term construction employment and longer-term shifts in local electricity costs (Alvarez et al., 2026)——that could independently affect student outcomes. To isolate the specific impact of pollution from these macro-economic shocks, we follow past approaches and construct a comparison group of schools that are located between one and two miles of data centers where there is also a treated

[8]We use U.S. public school geocoordinates from the 2023–2024 school year to maximize location precision. The results are not meaningfully sensitive to excluding the small number of schools whose address changes prior to the 2023-24 school year affected their sample inclusion or treatment status.

school within one mile. In some cases, additional data centers open within a mile of a school already treated by exposure to an existing data center (see Figure 3). In these instances of multiple exposure, we use the first data center permits to identify the group of comparison schools and also ensure that no comparison schools are ever within a mile of any data center in our panel dataset.

The comparison schools in our sample consequently share the same local economic environment as the treated schools but remain outside the immediate radius of higher risk to pollution exposure (defined as within one mile). By comparing schools within one mile of a data center to schools within one to two miles of a data center, our design likely captures only differences in the risk of acute pollution exposure that dissipates with distance (Currie, Davis, et al., 2015) rather than any broader impacts the data center may have on communities. Descriptive findings in Table 2 show that the comparison schools are similar in their baseline observable characteristics relative to the treated schools; of the eight covariates tested, only the percent Hispanic shows a statistically significant difference at the 5 percent level, with schools within one mile having a 6 percentage point higher level of enrollment.

The final sample includes 740 schools across 12 states, consisting of 206 treated schools that are within one mile of a data center and 534 comparison schools that are within one to two miles of a data center (see Table 1). While Virginia has the largest number of data centers (379), followed by California (292), Texas (80), and New Jersey (61), the distribution of treated schools varies based on local proximity between schools and facilities. California contributes the largest share of treated schools to the sample (78), followed by New Jersey (49), Virginia (22), and Texas (20). Consistent with the expansion patterns shown in Figure 1, Virginia exhibits more data center growth than other states—its 22 treated schools are near 36 data centers in their initial treated years and 56 centers over the course of the panel. Treated schools in New Jersey make up a large share of the earliest treated cohorts, including 60 percent of all schools treated in 2016 and 2017, as shown in Figure A.1. The geographic footprint of later cohorts is more distributed, except for California's large share of treated schools in 2023 and 2024, and a smaller 2025 cohort overall. Broadly across the sample, three states—California, Virginia, and Texas—are represented in nearly every cohort.

Among treated schools, the number of data centers within a mile ranges from one to 11. Nearly 30 percent of treated schools (59 of 206 schools) are within a mile of more than one data center ("multi-site" schools). Some multi-site schools are exposed to more than one facility in the same year, others are exposed to additional nearby facilities in later years, and some experience both. As shown in Figure 3, more than half of the 59 multi-site schools (33 schools) are within one mile of multiple facilities in the same initial permit year. Of those 33 schools, 19 later have additional permits issued nearby, while the remaining 14 experience no further permits. The remaining 26

multi-site schools are near only one data center in the initial permit year, with additional nearby permits occurring in later years. The remaining 147 single-site treated schools are within a mile of only one data center. Importantly, we identify comparison schools—which are never within a mile of any data center—based on their proximity to the first data center(s) within a mile of treated schools. Furthermore, the temporal distribution of these initial permits is similar across schools in both the single- and multi-site groups, allaying concerns that differences in treatment intensity are confounded by differences in treatment timing.

Notably, schools that are near more than one data center are exposed to substantially greater backup generator capacity. Each center's permit lists the total rated capacity (in kW) of all permitted generators covered by that permit. Average total capacity is roughly five times higher for schools near multiple facilities than for schools near a single facility (108,367 kW vs. 20,226 kW), suggesting that multi-site schools experience larger aggregate emissions.[9]

In addition to variation in the number of nearby data centers and potential pollution exposure, the number of matched comparison schools per treated school also varies. The number of comparison matches per treated school ranges from zero to 35, with 12 treated schools having no matches and one having the maximum. While our primary analyses pool across all treated and comparison schools, we also present results from specifications with state-by-year or county-by-year fixed effects to compare treated schools to comparison schools within the same region and year, absorbing local time-varying trends that might affect student outcomes and confound our treatment estimates.

3.2 Empirical Design

We apply a spatial difference-in-differences analysis strategy following prior studies (Persico and Venator, 2021) to estimate a reduced-form, partial-equilibrium effect of exposure to data centers. The two-way fixed effects (TWFE) event-study specification that motivates our analysis takes the form of:

$$Y_{st} = \beta_0 + \sum_{\substack{j=-9 \\ j\neq -1}}^{9} \beta_j \mathbb{1}[t - D_s = j] + \eta_s + \delta_t + \epsilon_{st}, \tag{1}$$

where Y_{st} is the average standardized ELA or math scale score for school s in the spring of year t. We focus on third grade outcomes because we hypothesize that the youngest students will be the

[9]For permits covering multiple facilities, we assigned each facility the average of the total rated capacity listed in the permit. For schools within a mile of multiple permits, we summed the permitted total rated capacity across facilities.

most sensitive to pollution and because this provides the longest panel of outcomes for students who enroll at treated schools in preschool or kindergarten after a data centered is issued its initial permit. The term D_s denotes the year a data center's first permit was issued. Our parameters of interest, β_j, represent the effect of a data center being within one mile of a school on mean test scores at event time j (where j is relative to the year the center's first permit was issued). School and year fixed effects, η_s and δ_t, remove time-invariant differences across schools and any common shocks in a given year. We weight the model by the number of tested third-grade students to identify the effect of being within one mile of a data center, on average, across all treated students in our sample.

The key identifying assumption is that treated schools would have followed parallel trends in academic achievement to comparison schools in the absence of the exogenous timing of a data center opening. Estimating long-run effects in our panel also requires the stronger assumption that parallel trends hold in every post-period (Baker et al., 2025). Given the delay between when a data center is issued its first air permit and when it is operational (i.e., 1–3 years) and our conservative matching of permit calendar years to academic school years, we anticipate dynamic treatment effects that materialize over time. We therefore use the Callaway and Sant'Anna (2021) estimator for our primary specification to account for these heterogeneous treatment effects and staggered treatment timings.

We also examine several dimensions of heterogeneity, including treatment intensity and student age. We construct a proxy for treatment intensity using a binary indicator that equals one for schools that are ever exposed to multiple data centers within one mile.

In addition, we report results from alternative specifications in Section 5 using a stacked event estimator, which allows for more flexible use of controls. In one set of specifications, we include state-by-year or county-by-year fixed effects. This approach ensures that treated schools are only compared to comparison schools within the same region and year, absorbing any broader state- or county-level geographic trends. The pattern of results remains unchanged.[10] In a second set of specifications, we control for potentially endogenous observable time-varying school and zip-code level demographic characteristics to assess the potential sensitivity of our result to changes in student composition. These covariates include the percent of students eligible for free or reduced-

[10]We do not present results that include zip-code-by-year fixed effects. Restricting the identifying variation strictly to within-zip comparisons mechanically splinters cross-boundary school matches, which reduces our sample size and leads to much noisier estimates. Sixty-nine percent of treated schools (142) share a zip code with at least one of their comparison school matches. Another 25 percent (52) do not share a zip code with any of their matches, while the remaining 6 percent do not have a comparison school match. Of the 142 schools that share a zip code with a match, 85 percent (120 schools) also have at least one comparison school located in a different zip code.

price lunch; the percent of third-grade students who are Black, White, Hispanic, or Asian; and the median household income of a school's zip code. While we exclude these controls from our primary estimation approach using the Callaway and Sant'Anna (2021) estimator, our results are not sensitive to their inclusion in doubly-robust applications of their estimator.

4 Results

4.1 Main Results

Table 3 reports the main findings for third-grade students. The estimates in the first column of Section 1 are from our preferred specification, which uses the Callaway and Sant'Anna (2021) estimator and is weighted by the number of tested third-grade students by subject. For all treated schools, the average effect across the full post period is −0.07 school-level standard deviations in math and −0.03 school-level standard deviations in ELA, suggesting small declines in test scores for schools within one mile of a data center that are not statistically significant. However, as shown in Figure 4, treatment effects are dynamic and emerge over time. In schools within one mile of a data center, math scores remain stable during the pre-period and early treatment years, and begin to decline six years after the initial permit was issued. This decline results in an average decrease of 0.22 school-level standard deviations between $T = 6$ and $T = 9$, an effect size equivalent to −0.10 standard deviations at the student level. ELA scores exhibit a similar pattern, decreasing by 0.14 school-level standard deviations between $T = 6$ and $T = 9$. This result is equivalent to –0.06 standard deviations at the student level and is marginally significant ($p < 0.10$).

In addition to being dynamic, effects vary by treatment intensity. As shown in Figure 5, the effects for schools near multiple facilities and those near a single facility diverge at $T = 6$. Accordingly, the first column of Section 2 in Table 3 reports that the average decline in math scores between $T = 6$ and $T = 9$ for schools near multiple facilities is −0.43 school-level standard deviations, which is equivalent to a 0.19 standard deviation decline at the student level. The corresponding effect in math for schools near a single facility (column 1 of Section 3 in Table 3) is −0.06 school-level standard deviations (or −0.03 standard deviations at the student level) and is not statistically significant. Effects for ELA are smaller in magnitude and not statistically significant for either group of treated schools.

4.2 What might explain the long delay in treatment effects?

Prolonged data center construction processes, along with matching permit calendar years to school years, clearly play a role in the delayed effects we find. Constructing a single data center building can take between one to three years of intensive site development——including tree clearing, land grading, construction, and equipment installation——meaning local environmental impacts from an operating facility likely appear years post-permitting (*Data Center Construction Timeline* 2026; JLARC, 2024). However, these factors alone are unlikely to cause such a long period of undetectable effects before negative impacts emerge after six years. We see three possible explanations for this pattern: intensity effects, cumulative dosage effects, and sensitive period effects.

First, effects may only arise because it takes data centers several years to ramp up to full operational capacity, finally reaching a level of pollution that is detrimental to young students. Descriptive evidence from Virginia data centers presented above in Figure 2B suggests that data centers do not reach peak emissions levels until four years after a permit is issued, although these estimates are imprecise. Intensity effects are also consistent with the larger effects for schools near multiple data centers. Exposure intensity is likely amplified for schools located near multiple data centers, whose permits are typically issued one to three years after the first in our main analytic sample. In Virginia, large data center campuses have taken up to seven years to fully complete, with work on additional buildings often beginning as soon as one is finished (JLARC, 2024).

A second plausible explanation for the delayed effects is cumulative dosage. The effects of potential exposure to pollution from data centers on students' academic achievement may only arise when cohorts are repeatedly exposed over many years. Point estimates from our event study first become notably negative, but not yet individually significant, at $T = 5$. This is the cohort of students who were first potentially exposed to pollution from data centers while attending school for four years (i.e., kindergarten in $T = 2$, first grade in $T = 3$, second grade in $T = 4$ and third grade in $T = 5$), assuming that data centers produce meaningful levels of pollution starting two years after a permit is issued. Slightly larger and significant effects arise for the following cohort of third graders in $T = 6$, the first cohort that could have been subject to five years of cumulative potential exposure if they attended preschool at a treated school. We find in the 2014–15 school year, the start of our panel, that 38 percent of the elementary schools in our treated sample enrolled students in prekindergarten while another 58 percent began in kindergarten.

A third candidate is that potential exposure to pollution from data centers is particularly harmful when students are early in their developmental stages. The effect we find only emerge after cohorts who could have attended preschool or kindergarten at treated schools age into third grade, starting

at $T = 5$ and $T = 6$. While it is difficult to distinguish between these three mechanisms given our aggregate measures of achievement, each of them are consistent with the pattern of results we find and offer plausible explanations for delayed effects.

5 Robustness

5.1 Alternative Specifications

We start by replicating our main Callaway and Sant'Anna (2021) model but removing weights. This approach identifies a different estimand, the average effect of a prototypical data center on student achievement (Solon, Haider, and Wooldridge, 2015). Here we find qualitatively similar but slightly smaller effects, with the pooled estimate for math across $T = 6$ to $T = 9$ at -0.14 school-level standard deviations (column 2 of Table 3). We then augment our main specification by using using the Callaway and Sant'Anna (2021) doubly robust estimator to control for school-level student demographic composition and zip-code-level household income. This approach combines inverse probability weighting with an outcome regression model to balance baseline covariates between the treated and comparison groups, and remains consistent if either component is correctly specified.[11] The estimates in column 3 of Table 3 are nearly identical to our main results.

For alternative estimates reported in columns 4 through 7, we use a stacked event study estimator, which is an intuitive and widely used alternative to more recent advances in the difference-in-differences literature. The stacked estimation approach builds a dataset that compares each year-specific cohort of treated schools to the full sample of never-treated schools. We then pool these cohort-specific effects across each stacked cohort in a single model by estimating equation (1) with interactions between indicators for each cohort with schools and also year fixed effects (Cengiz et al., 2022). Results from a stacked model that parallels our main specification, reported in column 4 of Table 3, are very similar in magnitude, sign, and significance to our main results.

Next, we assess the robustness of our baseline findings to more geographically granular controls by replacing the year fixed effects in the stacked event study with either state-by-year or county-by-year fixed effects to control for regional, year-specific shocks. This approach restricts our estimates to more localized identifying variation, effectively comparing treated schools to comparison schools within the same region and year. This tighter comparison absorbs any state- or county-specific time-varying trends, such as localized economic trends, that might affect student outcomes and

[11] We use the default doubly robust estimator, rather than the improved doubly robust estimator, because the default is the more stable estimator.

confound our treatment estimates. The resulting estimates in columns 5 and 6 of Table 3 are stable and consistent with our main findings.

5.2 Tests for student sorting

Given our school-level aggregate data do not allow us to track individual students over time, it is possible that dynamic student sorting across schools could be driving the results. For example, we would expect to see a similar pattern of results if the siting, construction, and operation of a data center decreased property values and caused wealthier families to exit and lower-income families to enter a neighborhood. We perform two tests to rule out this alternative explanation. Together, these results suggest that endogenous student mobility cannot account for the effects we find.

Following Pei, Pischke, and Schwandt (2019), we present placebo tests in Figure A.2 examining whether proximity to a data center affects school demographic composition and neighborhood income. Several patterns emerge. The proportion of White and Asian third-grade students are stable over time. There is a small decrease in the proportion of third-grade Black students (an average of 3 percentage points between $T = 6$ and $T = 9$) and a small increase in the proportion of third-grade Hispanic students (an average of 3 percentage points between $T = 6$ and $T = 9$) (see Table 4). The share of students eligible for free or reduced-price lunch—our most direct proxy for household economic disadvantage—is stable throughout the post-treatment period. Median household income at the zip-code level is similarly flat. If income-based sorting were driving the decline in test scores, we would expect the percent of students eligible for free or reduced-price lunch to rise and median household incomes to fall with the onset of effects; they do not.

We also directly test the sensitivity of our results to these small changes in student demographic composition by controlling for these potentially endogenous, time-varying "bad controls" in stacked event studies. Here, we utilize the stacked event study design because the Callaway and Sant'Anna (2021) estimator uses base-period values for estimation when covariates are time-varying. The estimates in column 7 of Table 3 are consistent with our main findings. This stability provides further evidence that the observed decline in test scores is not driven by treatment-induced shifts in the demographic composition of schools or neighborhoods.

5.3 Isolating localized pollution effect

We rely on prior evidence that the concentration and effects of pollution fade out as distance from the source increases as a motivation for identifying effects within a one-mile radius in our primary specification. However, it is possible that our reduced form estimates are capturing other hyper-

localized effects of data centers, such as local economic effects. We explore this possibility by estimating the impact of data center proximity on schools located within a one-to-two mile radius, using schools that are two-to-three miles away as the comparison group. This approach excludes all schools that are within one mile of data centers (the treated schools in our main specification). Economic effects on local neighborhood, such as an increase in local consumers due to new data center employees, are less likely to be strictly bound by radial distance and detectable within one-to-two miles. Figure A.3 illustrates that for this outer ring, the coefficients remain near zero and statistically insignificant across both the pre- and post-treatment periods. This lack of detectable effects beyond the one-mile threshold suggests that effects we find are hyperlocalized and consistent with pollution being the primary mechanism.

5.4 Alternative outcomes

Our preferred approach focuses on the first tested grade in elementary school, third graders, because research suggests the negative health effects of pollution are most acute for younger children due to their early stages of physical and neurological development. We explore whether the negative effects we find for third graders persist as they age by estimating effects on fourth- and fifth-grade cohorts.[12] Extending our analyses to these older grades comes with potential tradeoffs; if effects emerge in later post-treatment years, there is the most limited power to detect these effects.

Results from models that test for effects on fourth grade achievement in math shown in Figure A.4 find that effects first emerge in $T = 7$, exactly one year after the initial effects appear for third graders. This one-year lag is consistent with a delayed cohort-specific effect driven by exposure during students' earliest years in school. If cumulative exposure were the primary driver, we might expect to see effects on fourth graders also emerge in $T = 6$, given their total number of years of exposure were the same as third graders. What differs is that they were not exposed at quite an early age as the third grade cohort in $T = 6$. The estimated treatment effects between $T = 7$ and $T = 9$ on fourth-grade math (-0.21σ) and ELA (-0.14σ) in Table 5 are also similar in magnitude to the effects on third-grade test scores. This also suggests a sustained effect of exposure during a sensitive period rather than a growing, cumulative effect over time. In contrast, results for fifth-grade math are less detectable. However, given the staggered timing of data center openings, the specific fifth-grade cohort that was treated in their early elementary years only enter the sample toward the end of our panel, where a reduced number of treated units limits our statistical power to detect similar

[12] Of the 740 schools in our main sample, which we use to estimate effects on third-grade test scores, 733 schools have available fourth-grade test score data and 718 schools have available fifth-grade test score data. For these two subsamples, 99.6 percent of schools have at least 9 years of fourth-grade test scores and approximately 98 percent of schools have at least 9 years of fifth-grade test scores in ELA or math.

shifts. Overall, results from fourth grade test scores suggest that cohorts of students exposed to data centers starting in early elementary school may experience sustained negative effects on their human capital development, although we cannot confirm if this pattern persists into fifth grade and beyond.

6 Conclusion

The rapid expansion of data centers has become a major driver of economic growth due to both the infrastructure these centers require and the technologies they support. Evidence also suggests these data centers produce substantial negative externalities due to their expansive energy consumption and environmental footprint. Questions about continued investments in these facilities and where they are located require detailed evidence on their broader costs and benefits. Our analyses shed new light on the negative effects of data centers on students who attend schools within very close proximity of these centers. Importantly, our estimates only capture localized partial-equilibrium effects of data centers. It is possible data centers produce other broader externalities on students and schools not reflected in our findings.

The magnitude of effects we find are substantial and economically meaningful. A decline of 0.10 student-level standard deviations in third grade math scores represents the equivalent of between a month to two months of learning (Thum and Kuhfeld, 2020). Data centers reduce student achievement to a greater degree, in absolute magnitude, than over half of 700 plus educational interventions evaluated in randomized control trials (Kraft, 2020). We follow the approach used in (Chetty, Friedman, and Rockoff, 2014) to put this effect into economic terms. First, we calculate the mean present value of lifetime earnings at age 8—the typical age of a third-grader—to be approximately $\$973,000$.[13] If a 0.10 student-level standard deviation decline in test scores—the main finding in our study—is associated with a 1.2 percent decrease in earnings (Chetty, Friedman, and Rockoff, 2014), then the financial loss of attending school near a data center is 1.2 percent $\times$ $\$973,000 \simeq \$12,000$ per student and a total loss of $\$876,000$ across the typical third grade cohort in our sample.[14] As a result, local economic gains from data center development may be partially offset by future earnings losses.

[13] We calculate this number using wage and salary income from the 2025 Current Population Survey Annual Social and Economic Supplement to create a cross-sectional earnings profile for individuals between ages 18 and 65 (Ruggles et al., 2025). We discount each age's mean wages using a 3 percent real rate and then sum across all ages in the career range.

[14] The 1.2 percent figure is based on the finding by Chetty, Friedman, and Rockoff (2014) that a one standard deviation increase in test scores is associated with a 12 percent increase in earnings at age 28. The calculation assumes that the percentage loss in earnings remains constant over the life cycle, and that there is a parallel structure between the effect of an increase and a decrease in test scores on earnings.

The pattern of results we find is consistent with air, water, and noise pollution being the primary driver of the large localized negative effects on student achievement. These findings point to the importance of closely considering the proximity of daycare centers and pre-K-12 schools to the proposed sites for future data centers. The results also suggest that efforts to reduce pollution from data centers and mitigate its effects through investments in modern school infrastructure such as through HVAC systems, soundproofing, and water filtration are important directions for minimizing these negative environmental externalities on young children.

7 Tables

Table 1: Distribution of Data Centers and School Treatment Status by State

	Total Data Center Count	Data Centers Near Schools	First Set of Data Centers	Treated Schools	Treated: Single-Site	Treated: Multi-Site	Comparison Schools
CA	292	49	46	78	68	10	235
CO	44	7	6	9	8	1	22
MA	18	4	4	7	5	2	19
MI	7	2	2	2	2	0	5
MN	13	2	2	2	2	0	6
NC	8	2	2	3	3	0	0
NE	14	2	2	6	6	0	7
NJ	61	32	23	49	27	22	123
NY	19	5	3	7	1	6	33
TX	80	21	19	20	17	3	48
VA	379	56	36	22	7	15	35
WI	6	1	1	1	1	0	1
Total	941	183	146	206	147	59	534

Notes: This table reports the number of data centers and schools across states. The first column reports the number of data centers built or approved for construction in each state. The second column reports the number of data centers that are within one mile of schools in the sample. The third column reports the number of data centers that are within one mile of schools in the sample that define the treatment timing. The fourth column reports the number of treated schools in the sample, and the fifth and sixth columns disaggregate treated schools into those near one or multiple data centers. The last column reports the number of comparison schools in the sample.

Table 2: Summary Statistics

	Treated Schools	Comparison Schools	Difference	p-value
Percent White (Grade 3)	0.30	0.26	0.04	0.094
Percent Black (Grade 3)	0.14	0.11	0.03	0.064
Percent Hispanic (Grade 3)	0.41	0.47	-0.06	0.018
Percent Asian (Grade 3)	0.11	0.12	-0.01	0.493
Percent FRPL (School)	0.53	0.56	-0.04	0.155
Median HH Income (Zip Code)	70176	68469	1706	0.536
Average Scale Score, Math	1391	1494	-103	0.149
Average Scale Score, ELA	1381	1485	-104	0.145

Notes: This table presents school-level summary statistics measured in each school's baseline year (its first year in the panel). The Difference column reports the treated-group mean minus the comparison-group mean; the p-value column reports the result of a two-sample t-test of equality of means between the two groups.

Table 3: The Effect of Data Center Expansion on Third Grade Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Section 1: All Schools							
Panel A: Math							
ATT	-0.07	-0.03	-0.11	-0.05	-0.06	-0.01	-0.07
	(0.052)	(0.042)	(0.066)	(0.042)	(0.041)	(0.040)	(0.046)
$T \geq 6$	-0.22**	-0.14**	-0.23**	-0.22***	-0.23***	-0.16**	-0.22***
	(0.092)	(0.068)	(0.106)	(0.078)	(0.076)	(0.069)	(0.080)
Observations	6,594	6,594	5,164	51,398	51,317	50,545	40,707
Panel B: ELA							
ATT	-0.03	-0.01	-0.05	-0.02	-0.02	0.01	-0.03
	(0.041)	(0.039)	(0.056)	(0.036)	(0.036)	(0.039)	(0.039)
$T \geq 6$	-0.14*	-0.06	-0.17*	-0.13*	-0.15*	-0.08	-0.15*
	(0.076)	(0.067)	(0.097)	(0.080)	(0.079)	(0.079)	(0.078)
Observations	6,639	6,639	5,191	51,596	51,515	50,724	40,802
Section 2: Schools Near Multiple Data Centers							
Panel A: Math							
ATT	-0.17*	-0.10	-0.20*	-0.11	-0.12	-0.04	-0.13
	(0.100)	(0.068)	(0.119)	(0.082)	(0.081)	(0.079)	(0.088)
$T \geq 6$	-0.43***	-0.28**	-0.44***	-0.35***	-0.37***	-0.24*	-0.35***
	(0.149)	(0.111)	(0.167)	(0.133)	(0.133)	(0.127)	(0.138)
Observations	5,472	5,472	4,313	50,034	49,944	49,133	39,582
Panel B: ELA							
ATT	-0.06	-0.02	-0.07	-0.02	-0.03	0.04	-0.02
	(0.076)	(0.061)	(0.104)	(0.066)	(0.068)	(0.075)	(0.071)
$T \geq 6$	-0.20	-0.08	-0.26	-0.16	-0.19	-0.07	-0.19
	(0.138)	(0.106)	(0.164)	(0.145)	(0.146)	(0.142)	(0.139)
Observations	5,491	5,491	4,322	50,224	50,134	49,303	39,671
Section 3: Schools Near One Data Center							
Panel A: Math							
ATT	-0.02	0.01	-0.04	-0.03	-0.03	0.01	-0.04
	(0.052)	(0.052)	(0.064)	(0.046)	(0.045)	(0.043)	(0.051)
$T \geq 6$	-0.06	-0.04	-0.04	-0.11*	-0.11*	-0.08	-0.11*
	(0.073)	(0.077)	(0.084)	(0.060)	(0.060)	(0.063)	(0.061)
Observations	6,070	6,070	4,763	50,844	50,763	49,982	40,255
Panel B: ELA							
ATT	-0.02	-0.00	-0.02	-0.01	-0.02	-0.00	-0.03
	(0.044)	(0.049)	(0.055)	(0.042)	(0.041)	(0.045)	(0.047)
$T \geq 6$	-0.08	-0.04	-0.07	-0.12*	-0.13*	-0.08	-0.13**
	(0.068)	(0.081)	(0.077)	(0.066)	(0.068)	(0.077)	(0.063)
Observations	6,115	6,115	4,790	51,042	50,961	50,161	40,351

Notes: Column definitions: (1) Callaway and Sant'Anna (2021) estimator, weighted; (2) Callaway and Sant'Anna (2021) estimator, unweighted; (3) Callaway and Sant'Anna (2021) doubly robust estimator, weighted, covariates; (4) Stacked event study, weighted; (5) Stacked event study, state x year FE, weighted; (6) Stacked event study, county x year FE, weighted; (7) Stacked event study, endogenous covariates, weighted. In the Callaway and Sant'Anna (2021) specifications, the overall ATT is the simple aggregation of group-time average treatment effects, and the dynamic estimate is the aggregated ATT for event times $T = 6$ through $T = 9$. In the stacked event study specifications, the overall estimate is obtained by replacing the full set of post-treatment event-time indicators with a single post-treatment indicator, and the dynamic estimate is the unweighted average of the event study coefficients for event times $T = 6$ through $T = 9$.

Table 4: Placebo Tests

	White	Black	Hispanic	Asian	FRPL	HH Income
ATT	0.00	-0.01**	0.02**	-0.00	-0.00	431.51
	(0.006)	(0.006)	(0.010)	(0.004)	(0.006)	(1132.513)
$T \geq 6$	0.01	-0.03***	0.03***	0.00	-0.01	-649.99
	(0.011)	(0.010)	(0.012)	(0.007)	(0.014)	(2666.714)
Observations	6,189	6,020	6,502	5,938	6,361	6,641

Notes: Estimates use the Callaway and Sant'Anna (2021) estimator, weighted by the number of tested third-grade students in math. Results are similar if we use the number of tested third-grade students ELA. The overall ATT is the simple aggregation of group-time average treatment effects, and the dynamic estimate is the aggregated ATT for event times $T = 6$ through $T = 9$. The percent of White, Black, Hispanic, and Asian students is measured at the third-grade level. The percent of students eligible for free-or-reduced price lunch is measured at the school level. Median household income is measured at the zip code level.

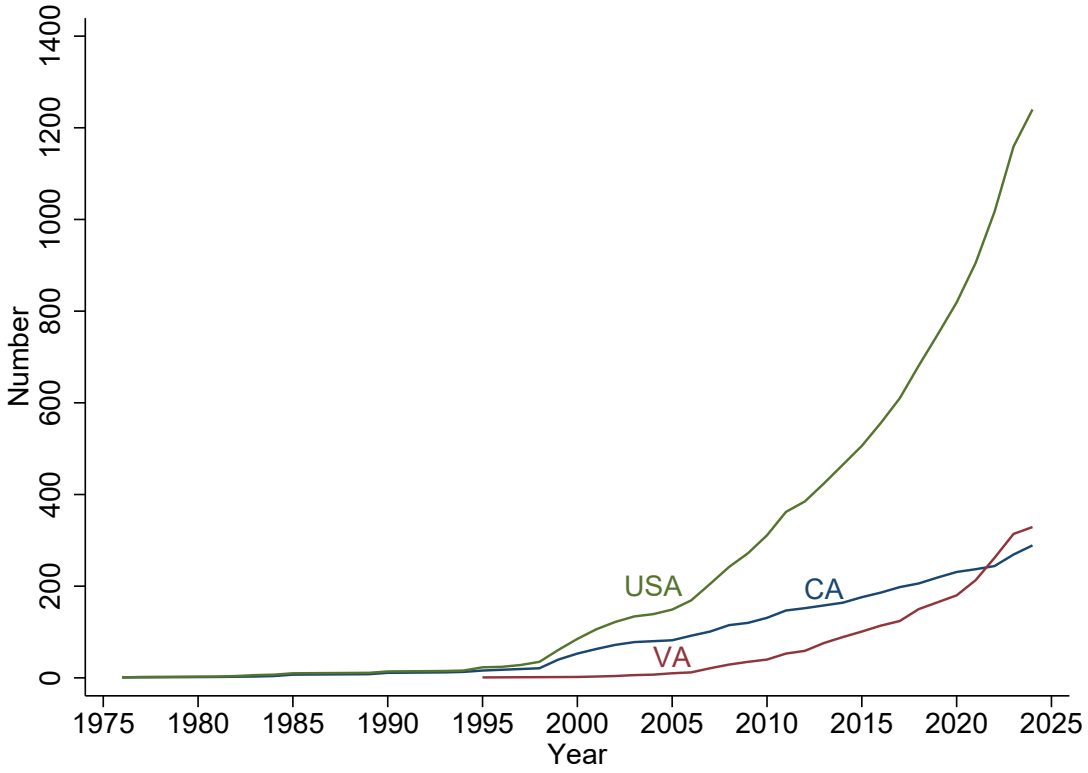
Table 5: Results for Fourth- and Fifth-Grade Test Scores

	(1) Grade 4	(2) Grade 5
Panel A: Math		
ATT	-0.06	-0.05
	(0.046)	(0.047)
Dynamic	-0.21**	-0.14
	(0.097)	(0.095)
Observations	6,569	6,306
Panel B: ELA		
ATT	-0.03	0.07**
	(0.039)	(0.034)
Dynamic	-0.13	0.02
	(0.086)	(0.069)
Observations	6,523	6,356

Notes: Estimates use the Callaway and Sant'Anna (2021) estimator, weighted by the number of tested students by grade and subject. The overall ATT is the simple aggregation of group-time average treatment effects. The dynamic estimate is the aggregated ATT for event times $T = 7$ through $T = 9$ for fourth grade students and $T = 8$ through $T = 9$ for fifth grade students.

8 Figures

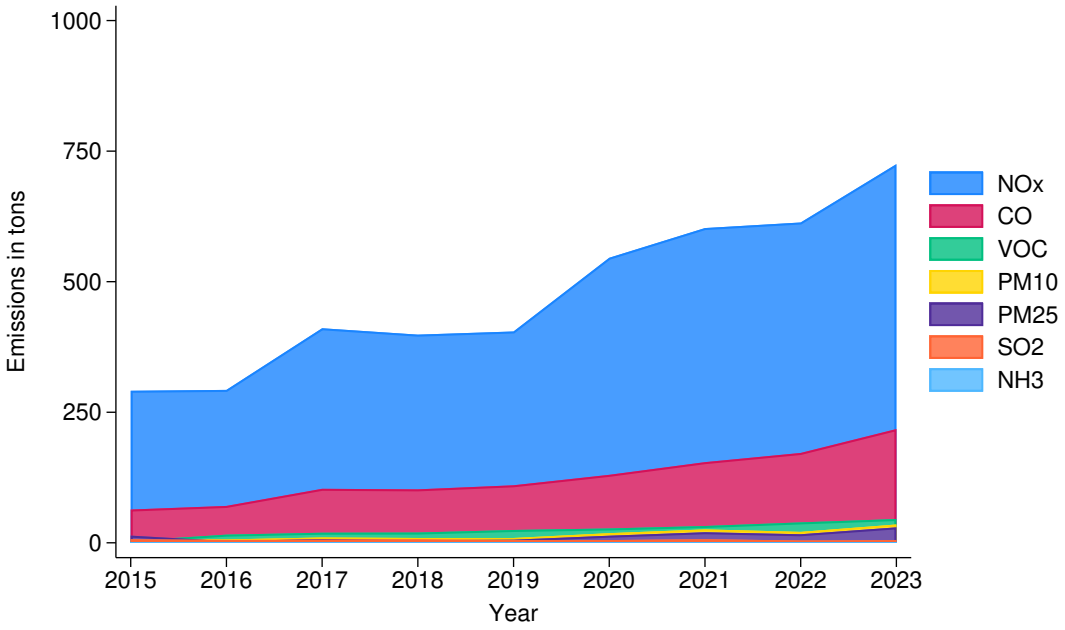
Figure 1: Data Center Expansion



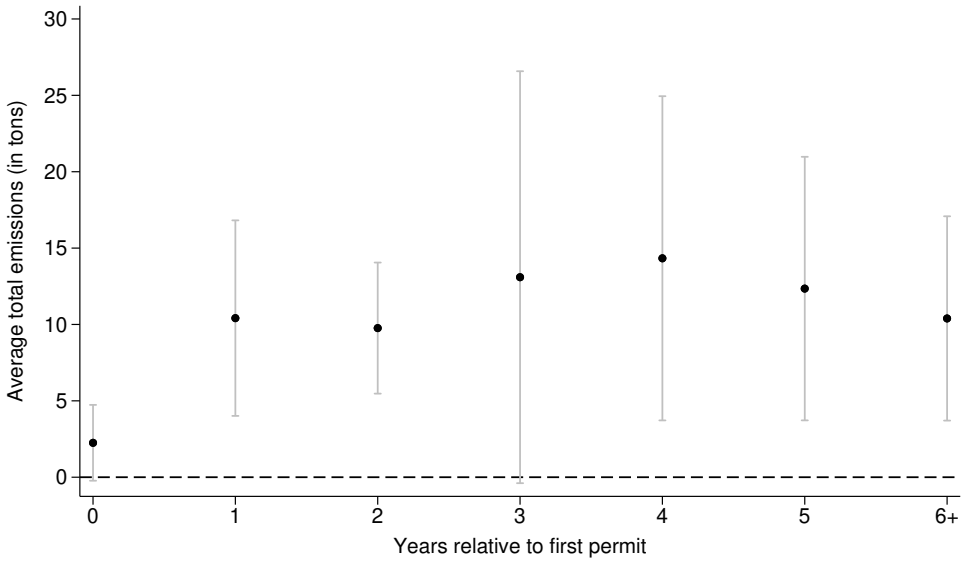
Notes: This figure plots the number of data centers built or approved for construction by 2025—uniquely identified by permit and address—using data from *Business Insider*.

Figure 2: Data Center Emissions in Virginia

(A) VA Emissions in Tons by Type

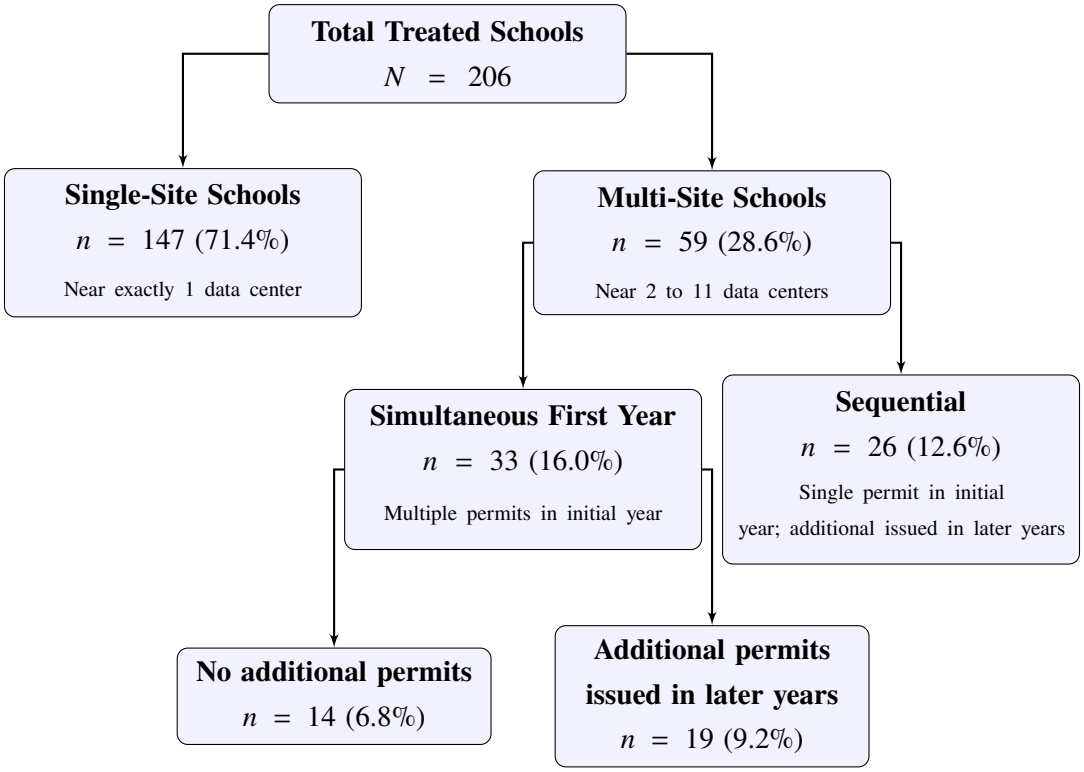


(B) VA Emissions Relative to Year of First Permit



Notes: The first figure plots the annual self-reported emissions from 2015 through 2023 that are linked to unique data center registration numbers in Virginia. The second figure shows mean within-center changes in emissions for the subset of Virginia facilities with air permits that were issued in or after 2015. Emissions are reported in tons.

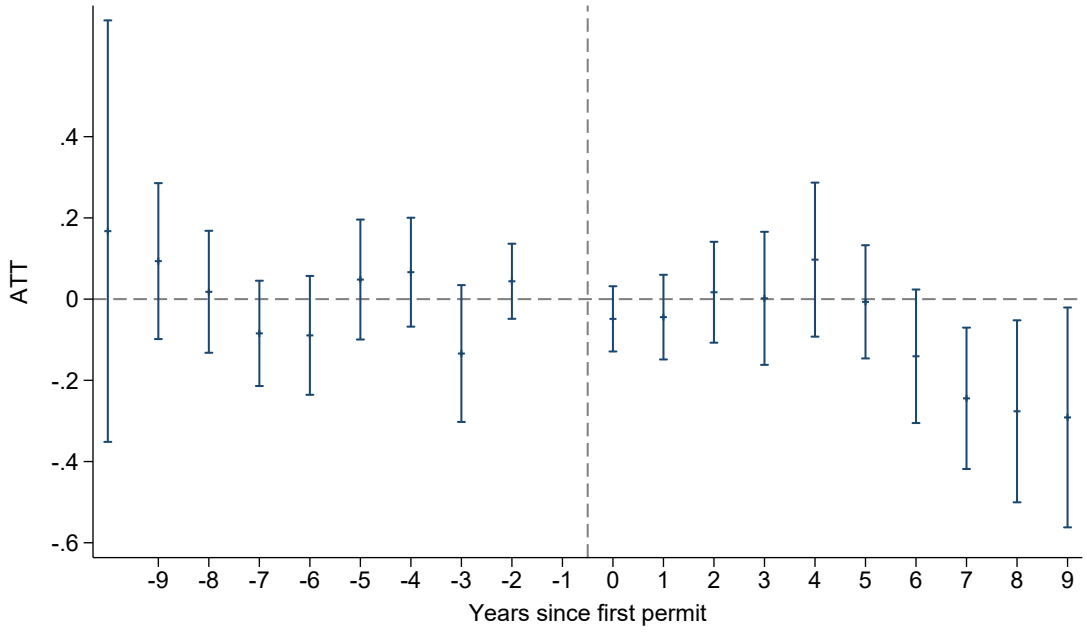
Figure 3: Treatment Status Sample Breakdown



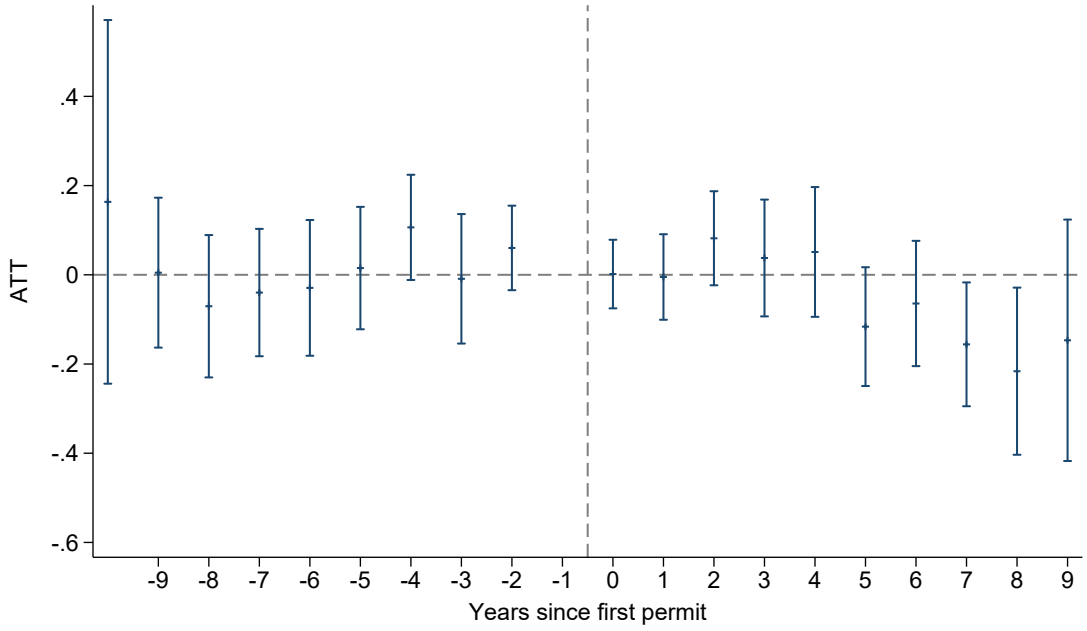
Notes: This figure shows the distribution of treated schools in the sample. See text for additional details.

Figure 4: Event Studies of the Effect of Data Center Expansion on Test Scores

(A) Math



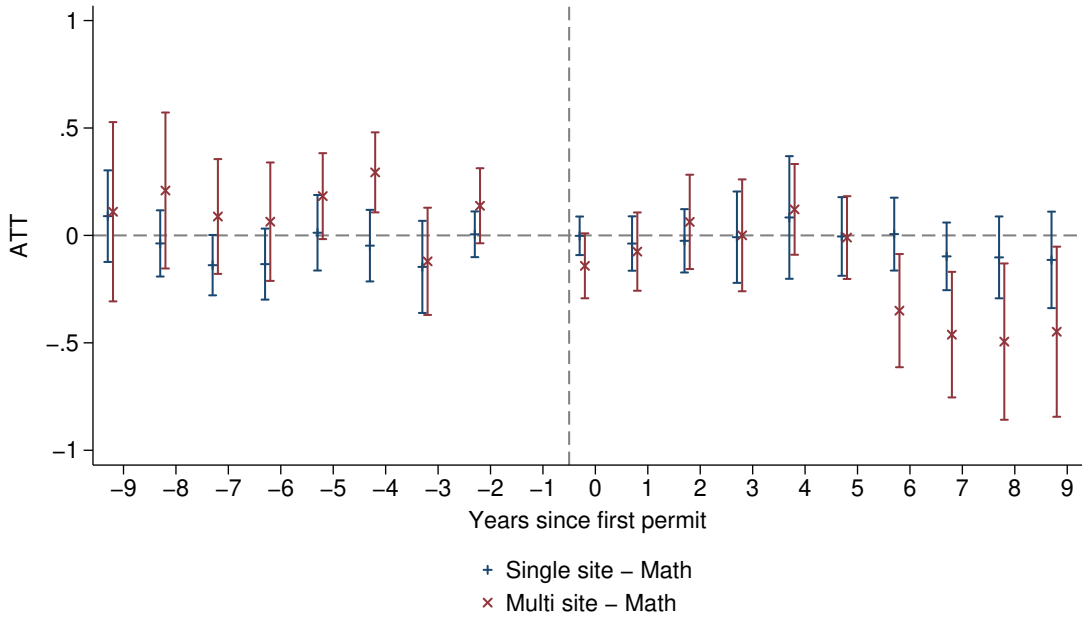
(B) ELA



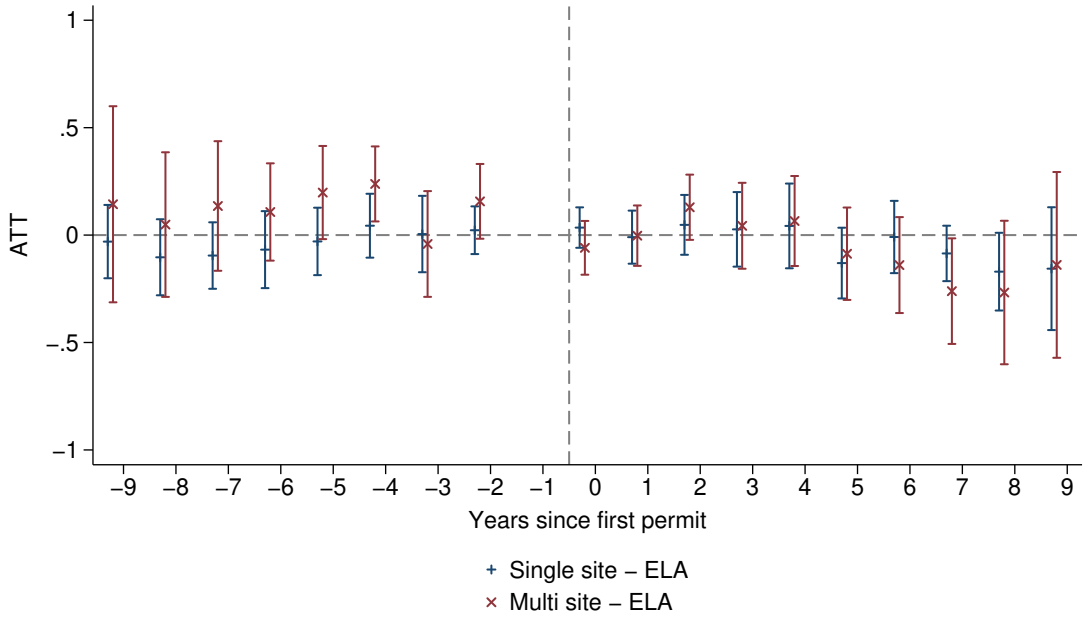
Notes: This figure shows the coefficients and 95 percent confidence intervals from event study regressions estimated using the long difference Callaway and Sant'Anna (2021) estimator. We omit $T = -10$ because of a loss of precision.

Figure 5: Effect of Data Center Expansion on Test Scores: Heterogeneity by Treatment Intensity

(A) Math



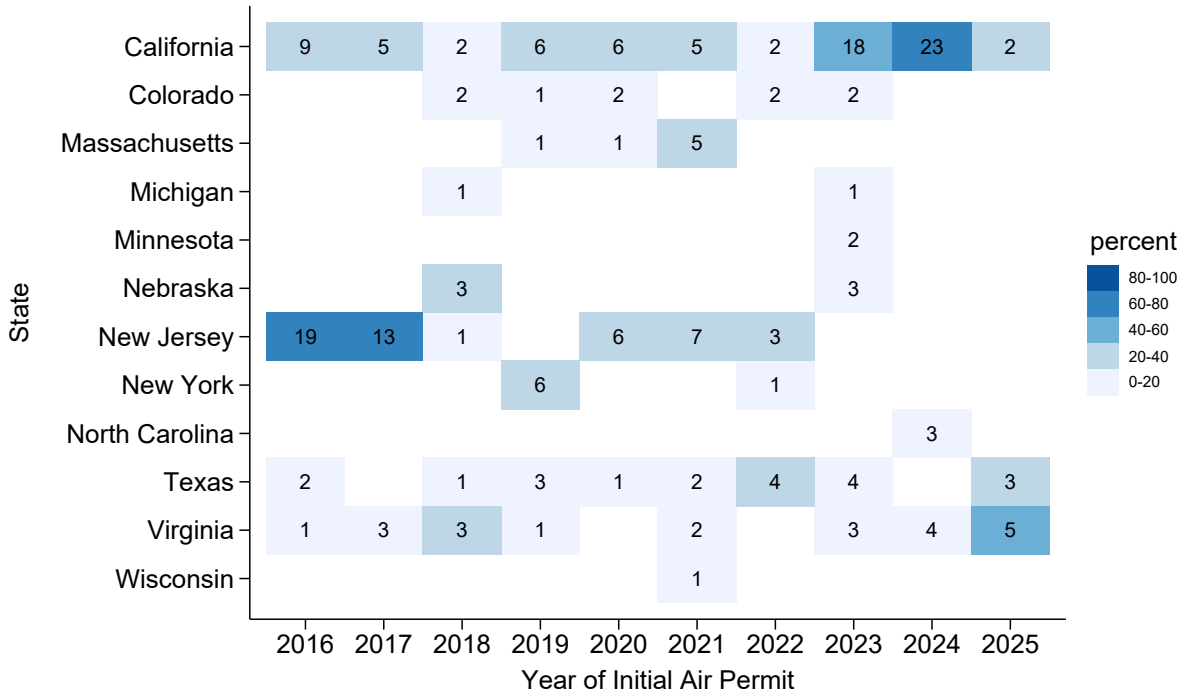
(B) ELA



Notes: This figure shows the coefficients and 95 percent confidence intervals from event study regressions estimated using the long difference Callaway and Sant'Anna (2021) estimator. We omit $T = -10$ because of a loss of precision.

A Appendix

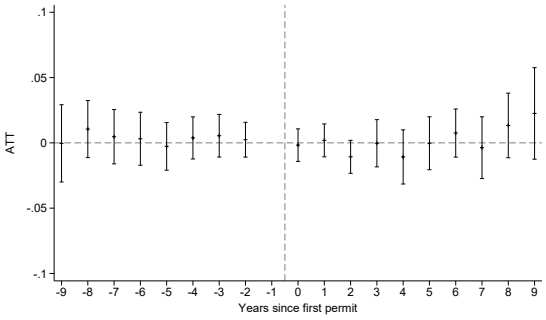
Figure A.1: Distribution of Treated Schools Across States and Cohorts



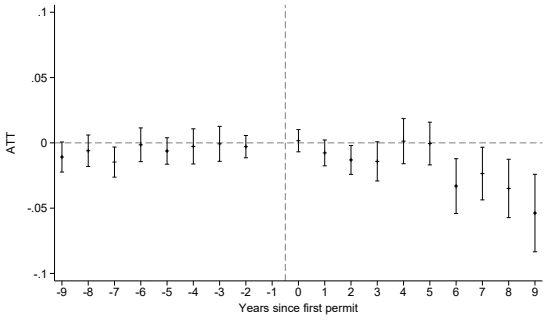
Notes: This figure shows the distribution of treated schools in the sample. See text for additional details.

Figure A.2: Placebo Tests with School and Zip Code-Level Covariates

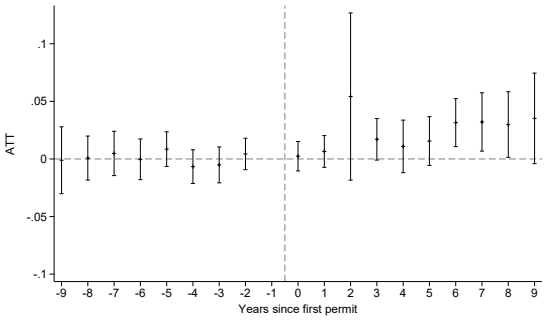
(A) White third graders



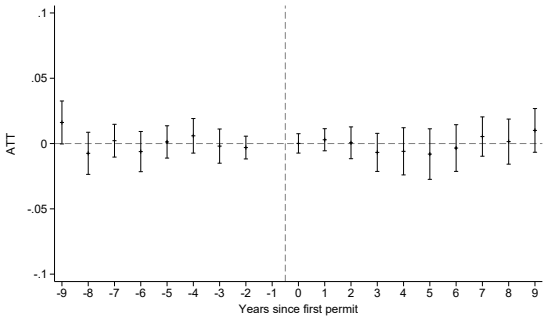
(B) Black third graders



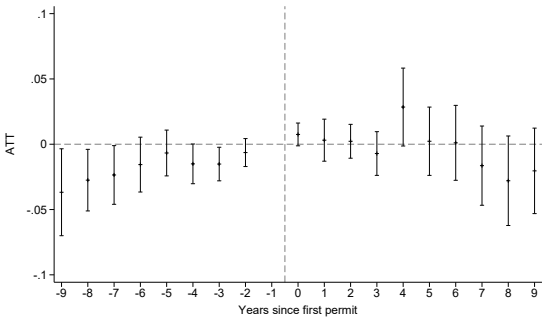
(C) Hispanic third graders



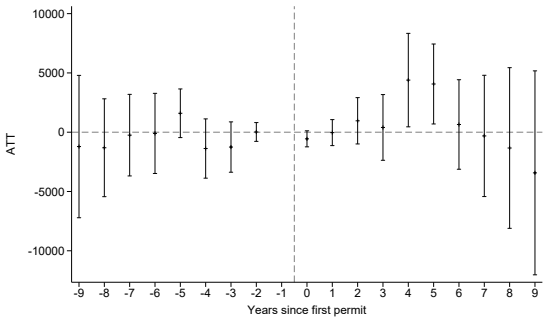
(D) Asian third graders



(E) Free or reduced-price lunch (all grades)



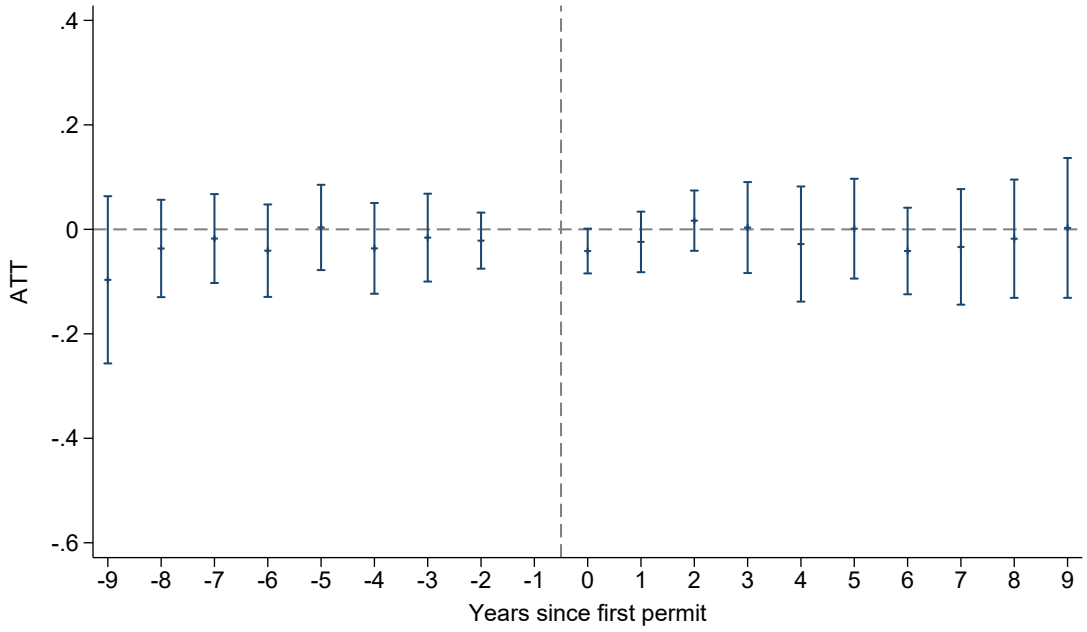
(F) Median household income



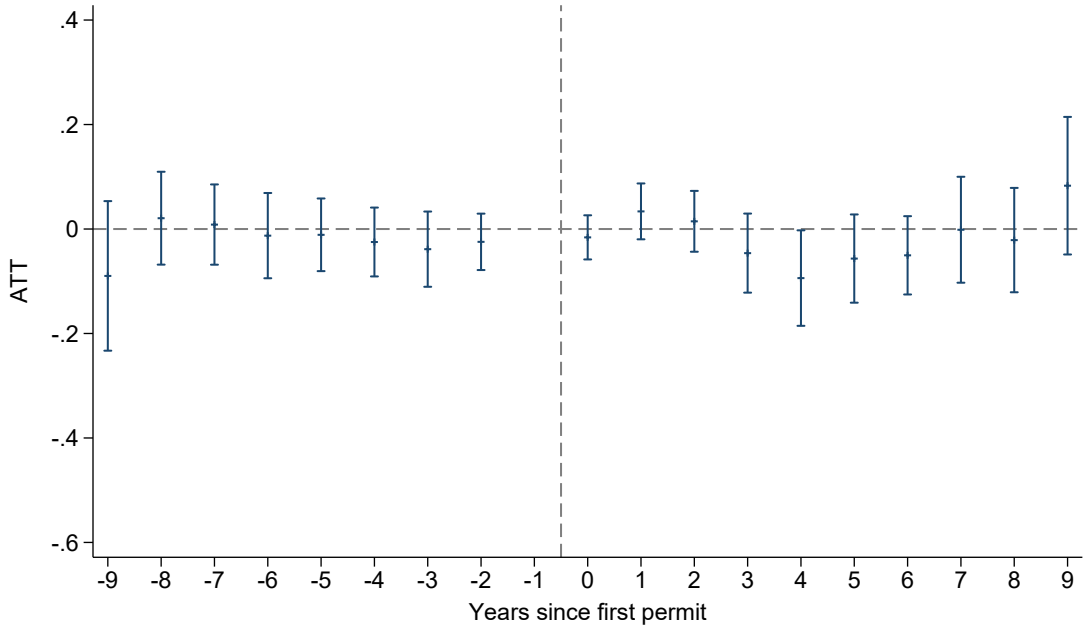
Notes: This figure shows the coefficients and 95 percent confidence intervals from event study regressions estimated using the long difference Callaway and Sant'Anna (2021) estimator. Each specification is weighted by the number of tested students in math; results are similar if we use the number of tested third-grade students ELA. We omit $T = -10$ because of a loss of precision.

Figure A.3: Fade Out Effects

(A) Math



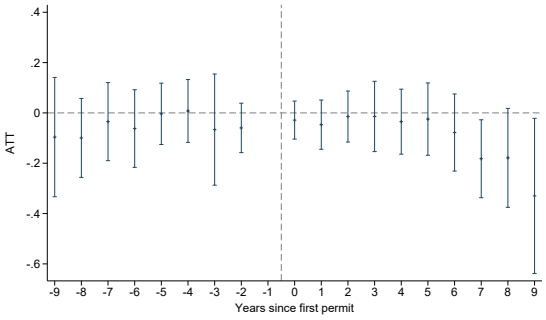
(B) ELA



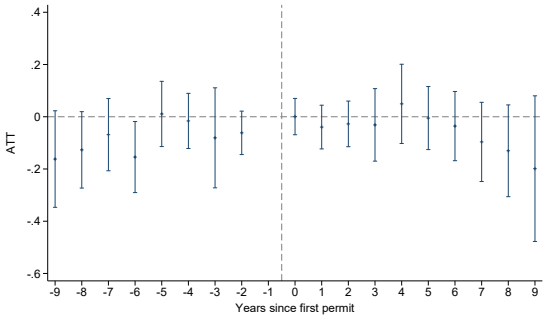
Notes: This figure shows the coefficients and 95 percent confidence intervals from event study regressions estimated using the long difference Callaway and Sant'Anna (2021) estimator. We omit $T = -10$ because of a loss of precision.

Figure A.4: Effects for Fourth- and Fifth-Grade Students

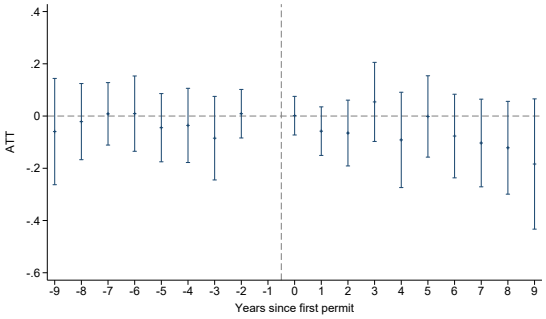
(A) Fourth grade math



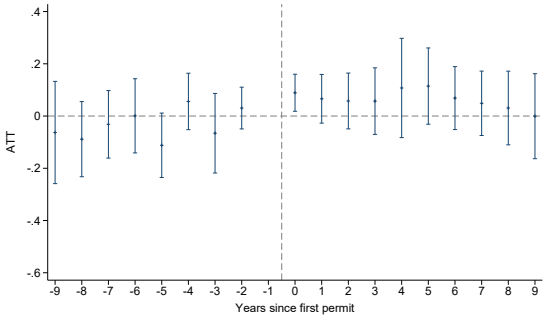
(B) Fourth grade ELA



(C) Fifth grade math



(D) Fifth grade ELA



Notes: This figure shows the coefficients and 95 percent confidence intervals from event study regressions estimated using the long difference Callaway and Sant'Anna (2021) estimator. We omit $T = -10$ because of a loss of precision.

B Appendix

B.1 *Business Insider* Data

Business Insider requested air permits and applications from 50 states and Washington, D.C., asking agencies to search permits using federal industry codes associated with data centers. To confirm state searches were complete, *Business Insider* cross-referenced permits against company disclosures, news articles, and data from datacentermap.com, Baxtel, and other trade websites. The subsequent data and reporting by *Business Insider* report 1,416 data centers that were built or approved for construction. Some permits contain multiple addresses for data center facilities. When clearly distinguishable by individual address, *Business Insider* counted each data center in the permit and added to the dataset, but assigned each data center the same ID. In other words, duplicate IDs indicate multiple facilities contained under one permit, and each facility is uniquely identified by ID and address. In the data, there are two facilities with the same ID and address, two facilities that are missing an address, and 6 permits/15 facilities with different addresses but the same geocoordinates. All of these observations—excluding one facility that is missing the year its first permit was issued—are included in Figure 1. None of these facilities with discrepancies are used in creating our main analytic sample.

Data reporting structures differ by state based on variations in their regulatory environments. For example, *Business Insider* reports that California issues distinct permits for each individual building within a data center campus, yielding highly disaggregated tracking. On the other hand, states like Arizona and Texas document data centers at the aggregate campus level rather than by individual building. Crucially, because these buildings are co-located in close physical proximity, this variation in geographic aggregation is unlikely to systematically alter a school's treatment classification or introduce meaningful spatial measurement error.

B.2 School-level σ to student-level σ conversion

Our analysis uses annual school-level data from the Education Data Center's State Assessment Data Repository and reports effect sizes in school-level standard deviations. To scale our estimates to the student level, we multiply the school-level mean effect sizes by the square root of the between-school intraclass correlation (ICC) (Lipsey et al., 2012). We use the between-school ICCs from Shear, Taylor, and Fahle (2026), which are calculating using average math and reading/language arts test scores in grades 3-8 for schools from all 50 states and D.C. from 2009 to 2019 from the Stanford Education Data Archive (SEDA) Version 5.0. These ICCs measure the proportion of variation in test scores that is between schools within states. The average ICC for third-grade math is 0.190

standard deviations and the average ICC for third-grade reading/language arts is 0.172 standard deviations.

References

Aarons, Joshua (2026). *Takeoffs and Test Scores: The Impact of Noise Pollution on Standardized Test Performance*. SSRN Scholarly Paper. Rochester, NY.

Air Permits (2026).

Aizer, Anna et al. (2018). "Do Low Levels of Blood Lead Reduce Children's Future Test Scores?" *American Economic Journal: Applied Economics* 10.1, 307–341.

Alexander, Diane and Hannes Schwandt (2022). "The Impact of Car Pollution on Infant and Child Health: Evidence from Emissions Cheating". *The Review of Economic Studies* 89.6, 2872–2910.

Alvarez, Fernando E. et al. (2026). *Data Centers and Local Economies in the Age of AI: A Shift–Share Approach*. Working Paper.

Ames, Jennifer L, Vanshika Sharma, and Kristen Lyall (2025). "Effects of Early-life PFAS Exposure on Child Neurodevelopment: A Review of the Evidence and Research gaps". *Current Environmental Health Reports* 12.1, 9.

Anderson, Michael L (2020). "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality". *Journal of the European Economic Association* 18.4, 1886–1927.

Austin, Wes, Garth Heutel, and Daniel Kreisman (2019). "School bus emissions, student health and academic performance". *Economics of Education Review* 70, 109–126.

Baker, Andrew et al. (2025). "Difference-in-Differences Designs: A Practitioner's Guide".

Baran, Cavit et al. (2025). *Clean Rides, Healthy Lives: The Impact of Electric Vehicle Adoption on Air Quality and Infant Health*. Working Paper.

Bearer, Cynthia F. (1995). "Environmental Health Hazards: How Children Are Different from Adults". *The Future of Children* 5.2, 11–26.

Beatty, Timothy K.M. and Jay P. Shimshack (2011). "School buses, diesel emissions, and respiratory health". *Journal of Health Economics* 30.5, 987–999.

Brumberg, Heather L. et al. (2021). "Ambient Air Pollution: Health Hazards to Children". *Pediatrics* 147.6, e2021051484.

Callaway, Brantly and Pedro H.C. Sant'Anna (2021). "Difference-in-Differences with multiple time periods". *Journal of Econometrics* 225.2, 200–230.

Cassauwers, Tom (2026). "Data Centers Have a PFAS Problem". *Sierra*. URL: https://www.sierraclub.org/sierra/data-centers-have-pfas-problem.

Cengiz, Doruk et al. (2022). "Seeing beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes". *Journal of Labor Economics* 40.S1, S203–S247.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014). "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood". *American Economic Review* 104.9, 2633–2679.

Chung, Sarah, Claudia Persico, and Jing Liu (2025). *The Effects of Daily Air Pollution on Students and Teachers*. Working Paper.

Copley, Michael (2022). "Data centers, backbone of the digital economy, face water scarcity and climate risk". *NPR*. URL: https://www.npr.org/2022/08/30/1119938708/data-centers-backbone-of-the-digital-economy-face-water-scarcity-and-climate-ris.

Currie, Janet, Lucas Davis, et al. (2015). "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings". *The American economic review* 105.2, 678–709.

Currie, Janet and Matthew Neidell (2005). "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" *The Quarterly Journal of Economics* 120.3, 1003–1030.

Data Center Air Monitoring Project (2026).

Data Center Construction Timeline (2026). *Data Center Construction Timeline: How Long It Takes + What Happens in Each Stage*.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth (2016). "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution". *American Economic Journal: Applied Economics* 8.4, 36–65.

Environmental Quality, Virginia Department of (2025). *Data Center Air Permits Guidelines*.

Ghafarianghadim, Mahta and Garth Heutel (2026). *Air Pollution and Early Childhood Outcomes: Evidence from Georgia Childcare Providers*. Working Paper 35367. National Bureau of Economic Research.

Grönqvist, Hans, J. Peter Nilsson, and Per-Olof Robling (2020). "Understanding How Low Levels of Early Lead Exposure Affect Children's Life Trajectories". *Journal of Political Economy* 128.9, 3376–3433.

Haines, M. M. et al. (2001). "Chronic aircraft noise exposure, stress responses, mental health and cognitive performance in school children". *Psychological Medicine* 31.2, 265–277.

Han, Yuelin et al. (2024). *The Unpaid Toll: Quantifying the Public Health Impact of AI*.

Hardin, Andie May (2026). "Data Centers Are Contributing to PFAS Forever Chemical Pollution". *EESI*. URL: https://www.eesi.org/articles/view/data-centers-are-contributing-to-pfas-forever-chemical-pollution.

Heissel, Jennifer A., Claudia Persico, and David Simon (2022). "Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance". *Journal of Human Resources*.

Holt, Jareth, Noelle E. Selin, and Susan Solomon (2015). "Changes in Inorganic Fine Particulate Matter Sensitivities to Precursors Due to Large-Scale US Emissions Reductions". *Environmental Science & Technology* 49.8, 4834–4841.

Ireland, Illan (2026). *A battle over data centers heats up along the Mississippi-Tennessee state line*. Tennessee Lookout. URL: https://tennesseelookout.com/2026/03/18/a-battle-over-data-centers-heats-up-along-the-mississippi-tennessee-state-line/.

JLARC (2024). *Data Centers in Virginia*. Tech. rep. 598. Virginia General Assembly.

Klatte, Maria, Kirstin Bergström, and Thomas Lachmann (2013). "Does noise affect learning? A short review on noise effects on cognitive performance in children". *Frontiers in Psychology* 4, 578.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders (2016). "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health". *The Review of Economics and Statistics* 98.2, 350–366.

Lipsey, Mark W. et al. (2012). *Translating the Statistical Representation of the Effects of Education Interventions Into More Readily Interpretable Forms*. Tech. rep. NCSER 2013-3000. Institute of Education Sciences.

Marcotte, Dave E. (2017). "Something in the air? Air quality and children's educational outcomes". *Economics of Education Review* 56, 141–151.

Martin, Arcelia (2025). *Data Centers' Use of Diesel Generators for Backup Power Is Commonplace—and Problematic*. Inside Climate News. URL: https://insideclimatenews.org/news/12112025/data-center-diesel-generators-noise-pollution/.

Matheson, Mark et al. (2010). "The effects of road traffic and aircraft noise exposure on children's episodic memory: the RANCH project". *Noise & Health* 12.49, 244–254.

Muller, Nicholas Z. (2026). *Measuring the Impact of Data Centers in the United States Economy: Monetary Damage from Air Pollution and Greenhouse Gas Emissions*. Working Paper 35100. National Bureau of Economic Research.

Nicoletti, Leonardo, Michelle Ma, and Dina Bass (2025). "How AI Demand Is Draining Local Water Supplies". *Bloomberg*.

Park, R. Jisung, A. Patrick Behrer, and Joshua Goodman (2021). "Learning Is Inhibited by Heat Exposure, Both Internationally and Within the United States". *Nature Human Behaviour* 5.1, 19–27.

Park, R. Jisung, Joshua Goodman, et al. (2020). "Heat and Learning". *American Economic Journal: Economic Policy* 12.2, 306–339.

Pawlowski, Alex (2025). *Data Center Boom 2025: AI Fuels 190% Surge in US Builds*. Substack newsletter.

Pei, Zhuan, Jörn-Steffen Pischke, and Hannes Schwandt (2019). "Poorly Measured Confounders are More Useful on the Left than on the Right". *Journal of Business & Economic Statistics* 37.2, 205–216.

Perkins, Tom (2025). "Advocates raise alarm over Pfas pollution from datacenters amid AI boom". *The Guardian*. URL: https://www.theguardian.com/environment/2025/oct/04/pfas-pollution-data-centers-ai.

Persico, Claudia and Joanna Venator (2021). "The Effects of Local Industrial Pollution on Students and Schools". *Journal of Human Resources* 56.2, 406–445.

Rau, Tomás, Sergio Urzúa, and Loreto Reyes (2015). "Early Exposure to Hazardous Waste and Academic Achievement: Evidence from a Case of Environmental Negligence". *Journal of the Association of Environmental and Resource Economists* 2.4, 527–563.

Robb, Drew (2026). "From Backup to Prime Power: How AI Data Centers Are Bypassing the Grid". *POWER Magazine*. URL: https://www.powermag.com/from-backup-to-prime-power-how-ai-data-centers-are-bypassing-the-grid/.

Roy, Angkana et al. (2011). "The Effects of Outdoor Air Pollutants on the Costs of Pediatric Asthma Hospitalizations in the United States, 1999 to 2007". *Medical Care* 49.9, 810–817.

Ruggles, Steven et al. (2025). *IPUMS USA: Version 16.0*. Minneapolis, MN.

Sadasivam, Naveena and Jake Bittle (2026). *Data centers are scrambling to power the AI boom with natural gas*. Grist. URL: https://grist.org/energy/data-centers-natural-gas-methane-behind-the-meter/.

Saul, Josh et al. (2025). "How AI Data Centers Are Sending Your Power Bill Soaring". *Bloomberg*.

Seagraves, Cayman and Stace Sirmans (2026). *AI Exposure and Housing Markets*. Working Paper. Social Science Research Network (SSRN).

Shear, Benjamin R., Joseph A. Taylor, and Erin M. Fahle (2026). "Determinants of Between-School Variation in Student Achievement: Results from US Population Data". *Journal of Research on Educational Effectiveness*, 1–25.

Siddik, Md Abu Bakar, Arman Shehabi, and Landon Marston (2021). "The environmental footprint of data centers in the United States". *Environmental Research Letters* 16.6, 064017.

Simeonova, Emilia et al. (2021). "Congestion Pricing, Air Pollution, and Children's Health". *Journal of Human Resources* 56.4, 971–996.

Skibell, Arianna (2026). "A data center drained 30M gallons of water unnoticed — until residents complained about low water pressure". *POLITICO*. URL: https://www.politico.com/news/2026/05/08/georgia-data-centers-water-00909988.

Smith-Schoenwalder, Cecelia (2026). "'Living in Hell': Data Center Neighbors Grapple With Noise, Air Pollution". *U.S. News & World Report*.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge (2015). "What Are We Weighting For?" *Journal of Human Resources* 50.2, 301–316.

Stansfeld, S. A. et al. (2005). "Aircraft and road traffic noise and children's cognition and health: a cross-national study". *Lancet* 365.9475, 1942–1949.

Stansfeld, Stephen and Charlotte Clark (2015). "Health Effects of Noise Exposure in Children". *Current Environmental Health Reports* 2.2, 171–178.

Thum, Yeow Meng and Megan Kuhfeld (2020). *NWEA 2020 Map Growth Achievement Status and Growth Norms for Students and Schools*. Tech. rep. Portland, OR: NWEA.

US EPA, OAR (2016). *Basic Information about NO2*. Overviews and Factsheets.

WHO global air quality guidelines (2021). *WHO global air quality guidelines*. Tech. rep. Geneva: World Health Organization.

Yañez-Barnuevo, Miguel (2026). *Communities Are Raising Noise Pollution Concerns About Data Centers*. EESI. URL: https://www.eesi.org/articles/view/communities-are-raising-noise-pollution-concernsabout-data-centers.