Air Pollution and Learning*

A. Patrick Behrer Joshua Goodman J. Parker Goyer R. Jisung Park

March 26, 2026

Abstract

We estimate the causal impact of air pollution on student learning using longitudinal PSAT data from over 16 million U.S. high school students who took the exam multiple times between 2007 and 2016. Our within-student design, combined with an instrumental variables strategy exploiting variation in wildfire smoke from distant fires, isolates plausibly exogenous changes in local PM2.5 concentrations. IV estimates are substantially larger than OLS, consistent with attenuation bias from measurement error. Exposure to moderately polluted days (8-12 $\mu g/m^3$) during the school year significantly reduces academic performance; scaling estimated effects by observed exposure implies an annual learning loss of approximately 0.055 standard deviations, or roughly 19% of typical yearly score growth. Effects are concentrated during the academic year rather than summer months, pointing toward instructional disruption and short-run cognitive interference rather than longer-run health channels. We document meaningful heterogeneity: students attending high-challenge schools and those with lower baseline achievement experience substantially larger adverse effects. Our results demonstrate harmful impacts at PM2.5 concentrations as low as 8 $\mu g/m^3$, below current and recently proposed EPA standards, with important implications for environmental regulation and educational equity.

Keywords: *Education; Pollution; Learning Outcomes;*
JEL Classification: *I21; I23; Q5.*

*A. Patrick Behrer, Development Research Group, World Bank, email: abehrer@worldbank.org; Joshua Goodman, Boston University, email: edecjosh@bu.edu; J. Parker Goyer, Policy Research, College Board, email: pgoyer@collegeboard.org; R. Jisung Park, University of Pennsylvania, email: rjpark@upenn.edu. The authors thank [**PEOPLE**] for their valuable comments and feedback. Funding from the Russell Sage Foundation is gratefully acknowledged. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1 Introduction

How does air pollution affect human capital accumulation? Determining whether, and the extent to which, air pollution impacts learning and human capital accumulation is of first-order importance to welfare for several reasons. First, the absolute magnitude of exposure is high: 99% of the world's population live in places that exceed the WHO's recommended levels of ambient PM2.5. Second, the distribution of exposure is highly uneven: globally, poorer individuals are far more likely to live in polluted areas (Rentschler and Leonova, 2023). Within countries more vulnerable communities are systematically more exposed to pollution (Colmer et al., 2020; Jbaily et al., 2022). Third, because local air pollution constitutes the classic negative externality, understanding the full breath of potential social costs associated with such pollution is critical to formulating efficient economic policies, including those aimed at reducing emissions as well as mitigating the harmful effects of pollution. Finally, accumulation of human capital is a pivotal input to economic growth at the macro level (Gethin, 2025) and economic mobility at the individual level (Heckman and Mosso, 2014), so that any effect of pollution on learning carries consequences that compound over a lifetime.

Our primary focus is on how pollution affects the cumulative rate of learning over time: specifically, how ongoing exposure to suboptimal air quality affects the total amount of human capital a student accumulates by the time they sit for important exams, continue to further schooling, or enter the labor force. This is a conceptually distinct question from whether pollution impairs performance on a specific exam day — a question the existing literature has addressed reasonably well. We are instead interested in whether sustained exposure to polluted air reduces the rate at which students actually learn.

Our data come from the College Board and comprise the universe of high school students who took the PSAT exam at least twice between 2007 and 2016. This student-level longitudinal data contains 16.7 million exam attempts by 7.8 million students from over 23,000 U.S. high schools. The focus on repeat test-takers is central to our identification strategy. Because many U.S. high school students take the PSAT more than once in grades 9 through 11, we can compare each student's performance across attempts that differ in the air quality they experienced in the preceding year, while holding constant fixed student-level factors such as underlying ability, motivation, household background,

and school quality. This within-student design is unusual in the pollution-and-learning literature, where outcomes are typically measured at the school or district level or students are observed only once. Our sample is substantially larger than any prior study of this question, more than sixteen times larger than the next largest examination of cumulative pollution effects on learning, giving us the statistical power to detect effects that may be modest on a per-day basis but consequential in aggregate.

A central challenge in estimating the effect of pollution on learning is that local air quality is not randomly assigned. Our within-student design addresses the concern that students in more polluted places differ systematically from those in less polluted places but does not fully address the possibility that year-to-year changes in local pollution are driven by changes in local economic conditions that independently affect learning. To isolate plausibly exogenous variation in pollution, we implement an instrumental variables strategy that exploits variation in wildfire smoke from distant fires (Childs et al., 2022). Smoke plumes from wildfires can travel hundreds of miles, raising local PM2.5 concentrations in ways that are largely orthogonal to local economic conditions, school operations, and household behavior (Zhang et al., 2025). By focusing on smoke from fires located at least 10 kilometers from a student's school, we isolate variation in local pollution that is driven by distant weather patterns and fire activity rather than by anything happening in the local community. Our first-stage F-statistics exceed 300 across all specifications, well above conventional thresholds for instrument strength. The instrumental variables strategy also mitigates attenuation bias driven by measurement error in our pollution measures, which are estimates from a machine learning model that predicts pollution concentrations from a variety of inputs. By instrumenting for total PM2.5 with smoke-derived PM2.5, which is subject to independent measurement error uncorrelated with the error in total PM2.5, the IV strategy purges this attenuation bias and recovers estimates closer to the true causal effect.

Our IV estimates indicate that particulate matter exposure reduces the rate of learning in economically meaningful ways. Scaling the estimated per-day effects by the exposure profile observed in our sample implies an average annual reduction in test-score growth of approximately -0.055 standard deviations, or roughly 19% of the learning that typically occurs over the course of a year, equivalent to about 2.3 months of learning. IV estimates are substantially larger than OLS

estimates, consistent both with attenuation bias arising from measurement error and with local economic fluctuations causing underestimation of effects. Our results are robust to a battery of falsification tests and alternative specifications, including controlling for the documented effects of temperature on learning both cumulatively and on exam days, as well as potential correlated trends in local economic activity and local air pollution exposure.

Notably, the learning impacts we estimate arise at PM2.5 concentrations that are close to typical ambient levels in the United States, rather than only at unusually high pollution levels. Observed effects are driven almost entirely by moderate pollution, days with PM2.5 between 8 and 12 $\mu g/m^3$, rather than by rarer heavy pollution events. Until recently, the EPA's National Ambient Air Quality Standards for Particulate Matter (PM NAAQS) used 12 $\mu g/m^3$ as the standard for a polluted day.[1] Our results indicate adverse effects at concentrations as low as 8 $\mu g/m^3$, underscoring that educational harms occur well below historical regulatory thresholds.

Importantly, these effects are concentrated during the school year rather than summer months. We find large and statistically significant effects of school-year pollution on test performance, but no comparable effect of summer pollution. This timing pattern points toward instructional disruption and short-run cognitive interference as the operative mechanisms, rather than longer-run health channels that would be expected to operate regardless of the academic calendar.

We also document meaningful heterogeneity in the learning impacts of pollution. Students attending high-challenge schools — those serving communities with greater socioeconomic disadvantage — experience substantially larger adverse effects of moderate pollution than students in lower-challenge schools, with IV estimates of the moderate-pollution effect roughly four times larger in high-challenge schools. Students who score below the national mean on their first PSAT attempt are similarly more vulnerable, experiencing effects nearly three times larger than their higher-achieving peers. These patterns suggest that pollution interacts with existing educational disadvantage: where resources, institutional capacity, and student supports are more constrained, the same environmental exposure does more harm. Taken together, our results imply that air pollution not only harms learning on average but also contributes to the widening of achievement gaps.

[1] On February 7, 2024, the US EPA lowered the National Ambient Air Quality Standards for Particulate Matter from $12\mu g/m^3$ to $9\mu g/m^3$. That decision is currently under reconsideration by the agency (`https://www.epa.gov/system/files/documents/2025-03/pm-naaqs_powering-the-great-american-comeback_fact-sheet.pdf`).

Our work contributes to four strands of the literature. First, our paper builds directly on a small but growing literature estimating the causal effect of cumulative pollution exposure on student learning. Wen and Burke (2022a) and Gilraine and Zheng (2022) combine nationwide school district-level longitudinal data with either smoke-attributable pollution or an instrument based on coal plant closures to show that higher PM2.5 during the school year reduces district-level test scores. Martin et al. (2025) use the construction of new bus rapid transit lines combined with wind direction to document learning losses from pollution in Bogotá. Dang et al. (2025) find adverse effects of pollution in the month and three months prior to an exam in Vietnam, though not over the full preceding year. A related literature examines the impact of pollution on student absenteeism, which is one channel through which cumulative exposure may affect learning (Komisarow and Pakhtigian, 2022; Chen et al., 2018; Conte Keivabu and Rüttenauer, 2022; Aziz and Elbakidze, 2025). Our study builds on this evidence by examining a sample more than sixteen times larger than any of these studies, using within-student rather than cross-sectional or district-level variation, and providing IV estimates based on a transparent and well-validated source of exogenous variation.

Second, our paper contributes to a larger and more established literature examining the effect of pollution on the day of an important exam (Marcotte, 2015; Ebenstein et al., 2016; Cavalcanti, 2024; Carneiro et al., 2021; Graff Zivin et al., 2020; Andersen et al., 2025; Koivuranta et al., 2024). This is a conceptually distinct question from the one we study. Acute impacts on exam day may reflect temporary physiological impairment or short-run changes in cognition at the moment of testing, consistent with evidence that pollution adversely affects cognitive performance on demanding tasks in non-academic settings rather than actual reductions in the stock of human capital (La Nauze and Severnini, 2021; Bedi et al., 2021; Archsmith et al., 2018; Huang et al., 2020; Dong et al., 2021; Qin et al., 2019). Our focus on pollution accumulated over the school year prior to an exam, combined with our finding that summer pollution has no comparable effect, helps to distinguish learning losses from test-day performance effects. Our results complement this literature by showing that the harms of pollution are not limited to the exam day itself but extend to the entire period of instruction leading up to it. This argument parallels the literature on heat and learning, where heat was first shown to have adverse acute impacts on student test performance (Park, 2022) and then subsequently to adversely impact the cumulative rate of learning (Garg et al., 2020; Park et al.,

2020, 2021).

Third, our results speak to a growing body of work showing that the physical conditions in which students learn matter for educational outcomes. Both Heissel et al. (2022) and Persico and Venator (2021) rely on wind direction combined with the location of major pollution sources, highways and industrial facilities respectively, to show that students in downwind schools suffer adverse learning impacts from plausibly elevated levels of pollution. Stafford (2015) finds that mold remediation programs that likely improve indoor air quality raise student performance. Most strikingly, Biasi et al. (2025) find that capital projects focused on HVAC improvements have the largest impact on test scores of any school infrastructure investment they study. There is also a separate but related literature showing that exposure to pollution in utero (Bharadwaj et al., 2017; Carneiro et al., 2024) or early in childhood Voorheis (2017) has adverse effects on learning later in life, with effects that may persist across generations (Colmer and Voorheis, 2020). Our results provide a specific, quantitative estimate of the learning cost of a marginal increase in PM2.5 that helps explain why improving indoor air quality through HVAC investment generates such large returns. Such estimates are also directly relevant to regulatory cost-benefit analysis, including the choice of whether to set PM NAAQS standards at 12, 9, or some lower threshold.

Fourth and most broadly, our paper contributes to a literature on environmental inequality and its consequences for educational disparities. A well-documented literature shows that exposure to air pollution in the United States is unequally distributed, with communities of color and lower-income communities bearing disproportionate pollution burdens (Colmer et al., 2020; Jbaily et al., 2022), and that places with larger underrepresented minority populations and lower average incomes tend to experience higher ambient pollution concentrations. A separate literature documents the adverse health consequences of this unequal exposure, including effects on mortality, morbidity and dementia (Deryugina et al., 2019; Pope III et al., 2020; Bishop et al., 2022; Zhang et al., 2025). Our results connect these literatures by providing evidence that differential pollution exposure is one mechanism through which environmental inequality translates into educational inequality. Our decomposition analysis suggests that differential exposure to moderate pollution accounts for a small but non-trivial share of the Black-White achievement gap and the gap between high- and low-challenge schools. More broadly, our heterogeneity results imply that the students who are most exposed to pollution

are also the most harmed by it, a pattern of compounding disadvantage with important implications for both environmental and education policy.

2 Data and Descriptive Statistics

We combine three main data sources: nationwide student-level test score data from the College Board, high-resolution measures of local ambient PM2.5 concentrations, and weather data. We focus throughout on students who took the PSAT exam more than once, which allows us to compare each student's performance across attempts that differ in the air quality experienced in the preceding year. Below we describe each data source in turn and then present summary statistics for the analysis samples.

2.1 Test Score Data

Our test score data come from the College Board and comprise the universe of U.S. high school students who took the PSAT at least twice between 2007 and 2016. The PSAT is administered once per year, roughly during the third week of October, and most students first take it in tenth or eleventh grade, though some begin as early as ninth grade. Students take and retake the PSAT for a variety of reasons, including preparation for the SAT college entrance exam, qualification for National Merit Scholarships, and obtaining information about their college readiness.

The PSAT has several features that make it well-suited for studying the cumulative effects of pollution on learning. Because the exam is offered at a fixed time of year with advanced registration required, there is limited scope for endogenous taking or timing decisions. Proctoring is nationally harmonized and the test is centrally graded, limiting the potential for score manipulation. Most importantly, the test is designed to assess cumulative progress on skills learned during secondary school rather than generalized intelligence, making it well-suited for detecting the effects of environmental conditions on the rate of human capital accumulation rather than on test-taking ability.

We have test scores and dates from the universe of PSAT takers from high school classes between 2001 and 2022, though we restrict our primary analysis to exam years 2007 to 2016 due to the availability of pollution and wildfire smoke measures. We further restrict attention to students taking the PSAT

on the two most common national administration dates in each exam year, dropping students who take the exam on rarer makeup dates that may coincide with atypical testing conditions. After imposing this restriction, we recompute attempt numbers and drop students who become effectively one-time takers. We identify 9.8 million PSAT retakers in total, whose 18.8 million scores form the core of our analysis sample.

Our primary outcome is a student's combined math and reading score, standardized within each fall administration and scaled by a factor of 1,000 so that coefficients can be interpreted in thousandths of a standard deviation. The College Board data also contain demographic and geographic information on students, including gender, race, and parental education. We observe students' home addresses, which we use to assign census-tract-level mean incomes from the American Community Survey. High school identifiers allow us to assign students to specific PM2.5 exposure levels at the school's location and to merge school-level characteristics. We classify schools by their level of socioeconomic challenge using the College Board's Landscape index, a nationally normed percentile summarizing the educational and socioeconomic conditions in the communities served by each high school. We define high-challenge schools as those with a Landscape score at or above the 67th percentile. For our heterogeneity analysis by student achievement, we classify students as low- or high-achieving based on whether their first PSAT score falls below or above the national mean for that exam year.

2.2 Air Pollution Data

Our air pollution measures are derived from daily PM2.5 concentrations predicted using the machine learning approach of Di et al. (2019), which combines observed PM2.5 from ground-station monitoring data, satellite imagery, chemical transport models, and weather reanalysis products to produce daily predicted PM2.5 levels over the contiguous United States from 2000 to 2017 at a 1km x 1km grid resolution. For each high school in our data, we construct daily pollution measures by averaging all grid cells within 10 kilometers of the school's latitude and longitude, or if coordinates are unavailable, all grid cells within the school's ZIP code. This produces a dataset of school-specific daily PM2.5 levels for each day from 2000 to 2017.

We characterize students' pollution exposure as the total number of days falling into various concentration bins during the 365 days prior to the date they took the PSAT. Our primary measure

distinguishes three ranges: clean days with PM2.5 below 8 $\mu g/m^3$, moderately polluted days with PM2.5 between 8 and 12 $\mu g/m^3$, and heavily polluted days with PM2.5 at or above 12 $\mu g/m^3$. Days below 8 $\mu g/m^3$ constitute the reference category in all specifications. Using state-specific school calendars, we further distinguish pollution exposure on school-year days from pollution exposure during summer months, which allows us to test whether effects are concentrated during instructional time.

2.3 Wildfire Smoke Data

To construct our instrument for local PM2.5, we use data from Childs et al. (2022) on the daily contribution of wildfire smoke to ambient PM2.5 concentrations across the contiguous United States. These data are constructed by training a machine learning model on information from ground-based monitoring stations, satellite imagery, aerosol parcel transport models, and weather reanalysis products, and then applying the trained model to predict the level of PM2.5 attributable to wildfire smoke within 10km x 10km grid cells for each day from 2006 to 2023. The resulting dataset has been used extensively in both the economics and public health literatures to document the adverse impacts of wildfire smoke on mortality (Gould et al., 2024; Connolly et al., 2024), morbidity (Zhang et al., 2025), and the contribution of wildfires to PM2.5 trends across the United States (Burke et al., 2023).

We use a version of these data aggregated by the original authors to U.S. ZIP Code Tabulation Areas (ZCTAs), which provides predicted smoke-attributable PM2.5 levels for each ZCTA on each day of the year. We assign schools to smoke exposure levels based on the ZCTA corresponding to each school's ZIP code. The smoke data distinguish between days with no detected smoke plume, days with a detected plume but no predicted smoke-attributable PM2.5, and days with positive smoke-attributable PM2.5. We group the first two categories together as days with zero smoke contribution, and characterize remaining days by the level of smoke-attributable PM2.5 using the same bin boundaries as our total pollution measure: light smoke days with smoke PM2.5 between 0 and 8 $\mu g/m^3$, moderate smoke days between 8 and 12 $\mu g/m^3$, and heavy smoke days at or above 12 $\mu g/m^3$. The overlap between our PM2.5 data and the wildfire smoke data covers exam years 2007 to 2016, which defines the sample period for our IV estimates.

2.4 Summary Statistics

Table 1 presents summary statistics for our three analysis samples: all PSAT retakers with available test score data, the subsample with available PM2.5 data, and the IV-eligible subsample with both pollution and wildfire smoke data. The three samples are broadly similar along observable dimensions, which gives us confidence that the restrictions required for IV estimation do not introduce meaningful selection. We discuss each in turn.

The retaker sample is large and demographically diverse. Across all 17 million retakers, 54 percent are female, 53 percent identify as White, 13 percent as Black, 20 percent as Hispanic, and 9 percent as Asian. Average parental education falls between an Associate's and a Bachelor's degree, and the average high school Landscape challenge score is 40 on a nationally normed 100-point scale. Students take an average of 2.15 PSAT attempts, and first-attempt scores are centered near zero by construction, reflecting exam-year standardization. The subsample with available PM2.5 data is nearly identical along these dimensions. The IV-eligible subsample is slightly more Hispanic (23 percent versus 20 percent) and somewhat smaller — 9.8 million students across roughly 24,800 schools — reflecting the restriction to exam years 2007 to 2016, the requirement that both pollution and smoke data be available, and the exclusion of schools within 10 kilometers of wildfire perimeters.

Table 2 shows how student characteristics vary across racial groups and school challenge levels in the IV-eligible sample. White students have higher average first-attempt PSAT scores (0.27 standard deviations above the mean) than Hispanic students (-0.47) or Black students (-0.62), while Asian students score highest on average (0.49). Students in low-challenge schools score substantially higher on their first attempt (0.35 standard deviations above the mean) than students in high-challenge schools, who score well below the mean. These baseline differences motivate our heterogeneity analysis and underscore the importance of a within-student design that does not rely on cross-sectional comparisons across groups.

Pollution exposure varies substantially across students, schools, and years. In the full retaker sample, students experience on average 208 days per year with PM2.5 at or above 8 $\mu g/m^3$, split roughly evenly between moderate (101 days) and heavy (107 days) pollution. About 156 of these polluted days fall during the school year, with the remaining 52 occurring during summer months. Exposure

is somewhat lower in the IV-eligible sample — 190 total polluted days on average — reflecting the geographic distribution of locations with available smoke data. Figure 4 illustrates the considerable variation in pollution exposure across students: the distribution of total polluted days is spread widely, with moderate and heavy pollution days following quite different distributional shapes. Pollution exposure also varies meaningfully across demographic groups and school types in ways that directly motivate our heterogeneity analysis. As shown in Table 2, White students experience fewer polluted days on average (190 per year) than Black students (210) or Hispanic students (197), with moderate pollution days ranging from 99 for White students to 109 for Hispanic students. Students attending the highest-challenge schools experience on average 212 polluted days per year, compared with 188 for students in the lowest-challenge schools — a difference of roughly 24 days. These patterns are consistent with the well-documented relationship between socioeconomic disadvantage and pollution exposure (Colmer et al., 2020; Jbaily et al., 2022).

Smoke exposure is considerably lower in magnitude than total pollution exposure, as expected given that wildfire smoke is one of many contributors to ambient PM2.5. In the IV-eligible sample, the large majority of days have no detected smoke contribution. Days with moderate or heavy smoke-attributable PM2.5 are relatively rare but exhibit meaningful year-to-year variation across schools, which is the source of identifying variation we exploit in our IV strategy. As shown in Figure A1, smoke exposure is concentrated in the western United States and is substantially higher during summer months when wildfire activity peaks, though plumes regularly reach the Midwest and Northeast as well. The year-to-year variation in smoke exposure visible in Figure A6 reflects differences in fire activity and prevailing wind patterns across our sample period rather than systematic trends, providing the kind of quasi-random variation across years that our identification strategy requires.

3 Empirical Strategy

Our empirical strategy exploits within-student variation in prior-year air quality across repeated PSAT attempts to estimate the effect of cumulative pollution exposure on learning. Because many students take the PSAT more than once, we can compare the same student's performance across attempts that differ in the pollution they experienced in the preceding year, holding constant

everything that is fixed about that student. This within-student design addresses the concern that students in more polluted places differ systematically from those in less polluted places in ways that independently affect learning. It does not, however, fully address the possibility that year-to-year changes in local pollution are driven by changes in local economic conditions that also affect learning — for instance, a new manufacturing facility that simultaneously raises local PM2.5 and stimulates the local economy. To isolate variation in pollution that is plausibly free of this confound, we supplement the within-student design with an instrumental variables strategy that uses wildfire smoke from distant fires as a source of exogenous variation in local PM2.5. We first describe the OLS specification and then the IV strategy.

3.1 Within-Student OLS Design

The intuition behind our OLS design is that the same student, attending the same school, may experience meaningfully different levels of air pollution in the year preceding one PSAT attempt compared to another. By comparing a student's performance across these attempts, we can ask whether the attempt preceded by more polluted days yields lower scores — while holding constant the student's underlying ability, motivation, household background, and the long-run quality of their school.

Formally, let y_{istk} denote the standardized PSAT score for student i, attending school s, from graduating cohort c, taking attempt k on exam date t. We construct y_{istk} by standardizing the composite PSAT score within each fall administration and scaling by a factor of 1,000, so that coefficients can be interpreted in thousandths of a standard deviation. We characterize pollution exposure using bins that count the number of days in the 365 days prior to the exam on which daily average PM2.5 fell within each concentration range, with days below 8 $\mu g/m^3$ serving as the reference category in all specifications.

Our baseline specification uses a single measure of total polluted days, $D_{stk}^{\geq 8}$, which counts all days in the prior year with PM2.5 at or above 8 $\mu g/m^3$:

$$y_{istk} = \beta D_{stk}^{\geq 8} + X'_{stk}\gamma + \mu_i + \lambda_{ctk} + \varepsilon_{istk} \tag{1}$$

where μ_i is a student fixed effect and λ_{ctk} is a fixed effect for each combination of graduating cohort c, exam date t, and attempt number k. Student fixed effects absorb all time-invariant student characteristics, including ability, motivation, and household background. Cohort-date-attempt fixed effects absorb all variation tied to the timing of PSAT administrations, including differences across test dates, retake patterns, and state or district scheduling, so that identification is driven by idiosyncratic differences in pollution exposure across a given student's attempts.

Though we begin without additional controls beyond the above fixed effects, we successively add a vector X_{stk} that includes a rich set of time-varying controls: prior-year average maximum temperature and precipitation at the student's school, exam-day PM2.5, exam-day temperature and precipitation, and Census-division-specific linear time trends in exam year to account for region-specific changes in air quality, schooling, or test-taking behavior over time. In an augmented specification we also control for county-level unemployment in the exam year to address the possibility that local economic shocks simultaneously influence local air quality and educational achievement. Standard errors are clustered by high school, the level of the identifying variation.

To investigate whether effects differ across pollution intensity levels, we also decompose $D_{stk}^{\geq 8}$ into moderately polluted days (D_{stk}^{mod}, with PM2.5 between 8 and 12 $\mu g/m^3$) and heavily polluted days (D_{stk}^{heavy}, with PM2.5 at or above 12 $\mu g/m^3$) and estimate:

$$y_{istk} = \beta_2 D_{stk}^{mod} + \beta_3 D_{stk}^{heavy} + X'_{stk}\gamma + \mu_i + \lambda_{ctk} + \varepsilon_{istk} \tag{2}$$

This decomposition allows us to test whether the learning impacts of pollution are driven by the frequency of more common moderate pollution days or by heavier pollution events. The same fixed effects and control vector apply in both specifications, and both are estimated on the same sample of students and attempts.

Identification within this design comes from comparing the same student's PSAT performance across attempts preceded by different pollution histories, after removing the influence of fixed student characteristics, common shocks to students taking the exam at the same time, and the time-varying controls described above. Figure **??** illustrates the residual variation in pollution exposure that drives our estimates after netting out these fixed effects. There is substantial within-student variation in

prior-year pollution, with a standard deviation of roughly 13 days for total polluted days and 11 days for heavily polluted days.

3.2 Instrumental Variables Strategy

While the within-student design addresses selection on time-invariant characteristics, OLS estimates may still be biased if year-to-year changes in local pollution are correlated with year-to-year changes in local economic conditions that independently affect learning. A new factory that raises local PM2.5 while also creating jobs and raising household incomes would cause OLS to understate the true harm of pollution; a local recession that reduces both pollution and school resources would cause OLS to overstate it. The direction of this bias is theoretically ambiguous, which makes it difficult to draw confident conclusions from OLS estimates alone.

A further concern is attenuation bias from measurement error. Our school-level PM2.5 measures are constructed from a machine learning model that predicts pollution concentrations at 1km $\times$ 1km resolution, which introduces noise into the pollution exposure assigned to any given student. Classical measurement error of this kind biases OLS estimates toward zero, causing them to understate the true effect of pollution on learning. By instrumenting for total PM2.5 with smoke-derived PM2.5, which is subject to independent measurement error uncorrelated with the error in total PM2.5, the IV strategy purges this attenuation bias and recovers estimates closer to the true causal effect.

To address both concerns, we instrument for local PM2.5 using variation in wildfire smoke from distant fires. Wildfire smoke plumes can travel hundreds of miles from their source, raising local PM2.5 concentrations in communities far removed from any active fire. For schools located well away from wildfire perimeters, year-to-year variation in smoke exposure is driven primarily by distant weather patterns and fire activity rather than by local economic conditions or other community-level factors that might independently affect student learning. This makes smoke-derived PM2.5 a plausible instrument for total PM2.5. It is strongly correlated with local pollution levels, but its variation is largely orthogonal to the local economic conditions that confound OLS estimates.

We implement this strategy using the smoke data described above, which characterizes each school-year observation by the number of days with light (0-8 $\mu g/m^3$), moderate (8-12 $\mu g/m^3$), and heavy ($\geq$12 $\mu g/m^3$) smoke-attributable PM2.5 in the prior year. Our baseline first stage instruments for

total polluted days $D_{stk}^{\geq 8}$ using all three smoke bins:

$$D_{stk}^{\geq 8} = \pi_1 S_{stk}^{\ell} + \pi_2 S_{stk}^{m} + \pi_3 S_{stk}^{h} + X'_{stk}\rho + \mu_i + \lambda_{ctk} + \eta_{istk} \tag{3}$$

where S_{stk}^{ℓ}, S_{stk}^{m}, and S_{stk}^{h} denote the number of days with light, moderate, and heavy smoke-attributable PM2.5 respectively, and student and cohort-date-attempt fixed effects (as well as any additional controls) match those in the OLS specification.

The corresponding second stage is:

$$y_{istk} = \beta^{IV} \hat{D}_{stk}^{\geq 8} + X'_{stk}\gamma + \mu_i + \lambda_{ctk} + \varepsilon_{istk} \tag{4}$$

where $\hat{D}_{stk}^{\geq 8}$ is the fitted value from the first stage equation (3). In our preferred specification we decompose pollution into moderate and heavy days and estimate a system in which both D_{stk}^{mod} and D_{stk}^{heavy} are jointly instrumented by the three smoke bins, mirroring the intensity decomposition in equation (2).

The key identifying assumption is that wildfire smoke affects student performance only through its effect on local PM2.5 concentrations. That is, conditional on total PM2.5, smoke exposure has no independent effect on learning. The most obvious threat to this assumption is that schools located near an active wildfire may experience both elevated smoke and a range of other disruptions that affect learning through channels other than air quality. To address this concern, we exclude all school-year observations for schools located within 10 kilometers of a wildfire perimeter in the relevant exam year. We show in Appendix Table **??** that our results are robust to alternative distance thresholds. A secondary concern is that smoke exposure may be correlated with other weather conditions that independently affect learning. We address this by controlling flexibly for prior-year and exam-day temperature and precipitation in all specifications. The instrument is strong across all specifications: F-statistics exceed 300 in all cases, well above conventional thresholds for instrument strength (Lee et al., 2020).

4 Results

4.1 Main Results

We begin by presenting within-student OLS estimates of the effect of prior-year pollution on PSAT performance, and then show how these estimates change when we instrument for pollution using wildfire smoke. The comparison between OLS and IV estimates is itself informative: as we discuss below, the IV estimates are substantially larger, consistent with attenuation bias from measurement error in our school-level pollution measures.

Table 3 reports within-student OLS estimates of the effect of prior-year particulate exposure on PSAT performance. Panel A uses our single aggregate measure of pollution — the total number of days with PM2.5 at or above 8 $\mu g/m^3$ — while Panel B decomposes this into moderately polluted days (8–12 $\mu g/m^3$) and heavily polluted days ($\geq$12 $\mu g/m^3$). We discuss results from our preferred specification in Column 5, which includes the full set of controls: student fixed effects, CDT fixed effects, prior-year weather, exam-day pollution and weather, and Census-division-specific linear time trends.

In Panel A, the coefficient on total polluted days is consistently negative, around -.02 thousands of a standard deviation in magnitude, but statisically insignificant. Panel B reveals a clearer picture. Each additional moderately polluted day in the prior year reduces PSAT performance by -0.067 thousandths of a standard deviation, while the effect of heavy pollution days is small and statistically insignificant. These results establish that moderate particulate pollution is the key margin through which air quality affects academic performance in the OLS estimates.

The near-zero coefficient on heavy pollution days warrants careful interpretation. Because days with PM2.5 below 8 $\mu g/m^3$ constitute the reference category, this coefficient should not be read as evidence that heavy pollution is harmless in an absolute sense. Rather, it indicates that heavy pollution days do not have a differential impact relative to low pollution days. One explanation is that students and schools engage in greater adaptive behavior on the most severely polluted days — staying indoors, canceling outdoor activities, or otherwise reducing exposure — which attenuates the measured impact. A second possibility, which motivates our IV strategy, is that high pollution days are disproportionately driven by local economic activity that has offsetting positive effects on

student performance, causing OLS to understate the true harm of heavy pollution.

Table 4 presents first-stage results showing that wildfire smoke strongly predicts total PM2.5 exposure. In Column 5 of Panel A, additional light and moderate smoke days significantly increase the number of days with PM2.5 at or above 8 $\mu g/m^3$. The F-statistic in Column 5 is 383.32, far above conventional weak-instrument thresholds, confirming that our instrument is strong. Panel B shows equally strong predictive relationships for moderate and heavy polluted days separately.

Table 5 presents the corresponding IV estimates. In Column 5 of Panel A, the coefficient on total polluted days is -0.170 thousandths of a standard deviation, roughly four times larger in magnitude than the corresponding OLS estimate. In Panel B, the effect of moderately polluted days is -0.456, while the effect of heavily polluted days remains small and imprecisely estimated. The IV estimate for moderate days implies that a 10-day increase in moderately polluted days decreases scores by 4.56 thousandths of a standard deviation — an order of magnitude larger than the corresponding OLS estimate.

The substantially larger IV estimates are consistent with two sources of downward bias in OLS. First, classical measurement error in our school-level PM2.5 measures attenuates OLS coefficients toward zero; the IV purges this bias by using smoke-derived PM2.5 as an instrument, which is subject to independent measurement error uncorrelated with the error in total PM2.5. Second, the positive correlation between local economic activity and pollution may cause OLS to understate pollution's true harm, particularly for heavily polluted days that are more likely to be driven by industrial activity. The IV estimates, which isolate variation in pollution driven by distant wildfire smoke, are free of this confound. Together these results provide strong causal evidence that particulate matter meaningfully reduces academic performance, with effects concentrated in the moderate pollution range that students experience on a routine basis.

4.2 School-Year versus Summer Exposure

The main results establish that prior-year pollution reduces learning, but they do not tell us through what mechanism. A key distinction is between mechanisms that operate specifically during instructional time — such as cognitive interference on polluted school days, disruption to classroom instruction, or increased absenteeism — and longer-run health channels that would be

expected to accumulate regardless of whether exposure occurs during the school year or summer. To distinguish between these possibilities, we decompose prior-year pollution into school-year and summer components, using state-specific school calendars to classify each day. If effects operate through instructional disruption or short-run cognitive impairment, we would expect the school-year component to drive our results. If effects instead reflect cumulative health impacts, summer pollution should matter as well.

Formally, we replace the full-year pollution measures in our main specifications with separate counts of polluted days during the school year and during summer months. In the IV models, school-year and summer pollution are instrumented using corresponding measures of smoke-derived PM2.5 from each season separately, with light, moderate, and heavy smoke bins linked separately to school-year and summer periods.

Table 6 presents IV estimates of the school-year versus summer decomposition. The results are striking. In Column 3 of Panel A — our preferred specification — an additional school-year polluted day reduces PSAT performance by -0.186 thousandths of a standard deviation, while the effect of summer pollution days is small and statistically insignificant. Panel B sharpens this result: each additional school-year moderately polluted day reduces scores by -0.702 thousandths of a standard deviation, while the corresponding summer exposure coefficients are small and statistically insignificant when school-year pollution is controlled for.

These timing estimates strongly suggest that learning disruptions arising from polluted school days — rather than cumulative long-run health exposure — drive the relationship between particulate matter and academic performance. A 10-day increase in school-year moderate pollution reduces scores by 7.02 thousandths of a standard deviation in our preferred specification. The absence of any comparable effect during summer months, when students are not engaged in formal instruction, is difficult to reconcile with a purely health-based mechanism and instead points toward instructional disruption and short-run cognitive interference as the primary channels through which pollution reduces learning.

4.3 Heterogeneity by High School Challenge

A central question in understanding the educational impacts of air pollution is whether the same environmental exposure generates different learning consequences across school contexts. If pollution affects learning primarily through biological or mechanical channels, its marginal impact should be similar across schools facing comparable exposure. In contrast, if learning losses are shaped by institutional and behavioral responses — such as schools' and families' ability to mitigate exposure or adjust instructional environments — then the educational effects of pollution may vary systematically with schools' capacity to adapt.

To assess whether effects differ across schools serving communities with different socioeconomic conditions, we stratify high schools into terciles of the College Board Landscape challenge index, which is a nationally normed percentile summarizing community educational and socioeconomic conditions constructed using all SAT, PSAT, and AP assessment takers across five consecutive graduating cohorts. We define high-challenge schools as those with a Landscape score at or above the 67th percentile, corresponding to the top third of schools nationally. We then re-estimate our preferred OLS and IV specifications separately within each group, focusing on school-year pollution to concentrate the analysis on the period when institutional and behavioral responses to pollution are most likely to shape learning outcomes.

Table 7 presents results. In the OLS estimates (Columns 1 and 2), an additional moderately polluted school-year day reduces scores by -0.162 thousandths of a standard deviation in high-challenge schools, while the corresponding effect in low-challenge schools is small and statistically indistinguishable from zero. The IV estimates sharpen this pattern considerably. In low-challenge schools, an additional moderately polluted school-year day reduces scores by -0.221 thousandths of a standard deviation. In high-challenge schools, the corresponding effect is -0.838, nearly four times larger, with an implied high-low challenge difference of -0.617. Heavy pollution days show a similar pattern of larger effects in high-challenge schools, though the differences are smaller in magnitude and only marginally significant.

These results imply that particulate pollution disproportionately harms students in more socioeconomically challenged school environments. The concentration of impacts in the moderate pollution

range suggests that the mechanism operates through the accumulation of many routine exposure days rather than through more extreme events. One interpretation is that high-challenge schools and the families they serve have fewer resources to mitigate exposure on moderately polluted days — through improved HVAC systems, scheduling adjustments, or other protective measures — while extreme pollution events may trigger non-discretionary responses such as school closures that partially equalize outcomes across school types. This pattern of greater vulnerability in disadvantaged settings has important implications for understanding how pollution contributes to educational inequality, a question we return to below.

4.4 Heterogeneity by Baseline Achievement

We next examine whether the effects of pollution differ for students who begin above versus below the national mean on their first PSAT attempt. Students who enter high school with lower baseline achievement may have fewer academic buffers against environmental shocks — less capacity to compensate for disrupted instruction, less access to supplementary academic support, or greater sensitivity to factors that impair concentration and effort. If so, we would expect pollution to generate larger learning losses among lower-achieving students, further widening achievement gaps over time.

To test this, we divide the retaker sample into two groups based on each student's standardized PSAT score on their first attempt relative to the national mean among all test takers in the same exam year. We then re-estimate our preferred OLS and IV specifications separately within each group, using full-year pollution measures. The baseline achievement indicator is defined using the first PSAT attempt and held fixed for all subsequent attempts by the same student, so that it is not contaminated by the pollution histories we are studying.

Table 8 presents results. The OLS estimates in Panel B suggest that harm from moderate pollution is concentrated among lower-achieving students: for students below the national mean, an additional moderately polluted day reduces scores by -0.128 thousandths of a standard deviation, while the corresponding coefficient for higher-achieving students is small and statistically insignificant. The IV estimates sharpen and extend this pattern. For students starting at or above the national mean, an additional moderately polluted day reduces scores by -0.25 thousandths of a standard deviation.

For students starting below the national mean, the estimated impact is nearly three times larger at -0.738. These coefficients imply that a 10-day increase in moderately polluted days lowers scores by approximately 2.50 thousandths of a standard deviation for higher-achieving students and 7.38 thousandths for lower-achieving students.

The IV estimates for heavy pollution days show an interesting reversal: effects are negative and significant for higher-achieving students (-0.194, s.e. 0.057) but small and imprecise for lower-achieving students (0.046, s.e. 0.068). We are cautious about overinterpreting this pattern given the relative rarity of heavily polluted days and the resulting limited within-student variation; we place more weight on the moderate-pollution coefficients as the primary margin of interest.

Taken together, these results suggest that moderate pollution reduces test performance across the achievement distribution, but that the impacts are substantially larger for students who begin below the national mean. This pattern is consistent with the broader heterogeneity findings in Section 5.3: just as high-challenge schools experience larger effects than low-challenge schools, lower-achieving students experience larger effects than their higher-achieving peers. Both patterns point toward the same conclusion — that pollution interacts with existing disadvantage to amplify learning losses, with implications for achievement gaps that we quantify in Section 6.

5 Magnitudes

To translate our IV estimates into economically interpretable magnitudes, we scale the estimated per-day effects by the observed exposure profile of students in the IV-eligible sample and express the resulting achievement costs in standard deviation units, percentile shifts, and time-based learning equivalents. Throughout, we interpret these quantities as local average treatment effects for the subpopulation of students whose pollution exposure is shifted by distant wildfire smoke.

5.1 Total achievement cost of observed exposure

The total annual achievement cost of pollution for the average student is the sum of the estimated per-day effects, weighted by average exposure in each concentration bin:

$$\text{Total cost} = \hat{\beta}_{8\text{–}12} \cdot \mathbb{E}[D_{8\text{–}12}] + \hat{\beta}_{\geq 12} \cdot \mathbb{E}[D_{\geq 12}] \tag{5}$$

Plugging in the IV coefficients from our preferred specification (Column 5, Table **??**, Panel B) and the sample mean exposure levels from the IV-eligible sample (Table **??**) yields an implied annual reduction in PSAT performance of -0.055 standard deviations. Under a normal approximation near the median, this corresponds to a percentile shift of approximately -2.2 percentage points. As a robustness check, we repeat the exercise collapsing moderate and heavy days into a single 8+ measure; that specification implies -0.040 standard deviations, confirming that the binned results are not an artifact of the two-category parameterization.

5.2 Benchmarking against typical learning gains

To express these effects in time-based units, we estimate an empirical learning rate from within-student score growth between consecutive PSAT attempts. The mean gain between the first and second attempt is 294 standardized-score points (where 1,000 points equal one standard deviation), implying approximately 24 points of growth per month. Dividing the average-student scaled effect of -55 points by this monthly rate yields an implied annual learning loss of approximately 2.3 months, or roughly 19% of typical yearly score growth. For comparison, this effect is also equivalent to about 19% of the mean attempt-to-attempt score gain and 13% of the mean absolute change between attempts—benchmarks that are directly observable in the data without any functional-form assumptions about the learning rate.

Two features of this estimate are worth emphasizing. First, it arises almost entirely from routine moderate pollution exposure (PM2.5 between 8 and 12 μg/m^3) rather than from extreme events: the IV coefficient on heavily polluted days is small and imprecisely estimated (Table **??**, Panel B), so the scaling exercise is driven almost entirely by the 103 moderate-pollution days the average student experiences. Second, as we discuss in the following paragraph, the effects are concentrated during the school year rather than the summer, which points toward instructional disruption as the operative mechanism rather than cumulative health degradation.

Contribution to achievement gaps

Beyond the average effect, our estimates shed light on whether differential pollution exposure amplifies existing achievement gaps. We construct a mechanical decomposition in the spirit of Blinder-Oaxaca: for any two groups $g \in \{A, B\}$, the contribution of differential moderate-pollution exposure to the observed score gap is

$$\text{Contribution}_{AB} = \hat{\beta}_{8\text{–}12} \left(\mathbb{E}[D^{A}_{8\text{–}12}] - \mathbb{E}[D^{B}_{8\text{–}12}] \right) \tag{6}$$

where $\hat{\beta}_{8\text{–}12}$ is taken from the pooled IV specification. This is an accounting exercise that holds the estimated pollution–score relationship fixed; it isolates the portion of the observed gap associated with differential exposure rather than providing a causal decomposition of broader disparities.

Applied to the Black–White gap, Black students experience on average 7.4 more moderate-pollution days per year than White students (106 versus 99, Table **??**), while scoring 888 standardized-score points below White students at their first PSAT attempt. Scaling the exposure difference by the IV per-day coefficient implies that differential moderate-pollution exposure mechanically accounts for approximately 3.8 points, or 0.4%, of the observed Black–White gap at first attempt. Though modest in relative terms, this contribution represents a quantifiable, policy-addressable share of a gap that is itself the product of many compounding disadvantages.

The analogous decomposition by school challenge level yields a somewhat larger share. Students in high-challenge schools experience 9.6 more moderate-pollution days per year than those in low-challenge schools (Table **??**) and score 729 points lower at first attempt. Applying the pooled IV coefficient implies that differential exposure accounts for approximately 4.9 points, or 0.7%, of the high–low challenge gap. Importantly, the pooled decomposition understates the true contribution to this gap, because the IV estimates themselves show that high-challenge students are roughly four times more sensitive to moderate pollution than low-challenge students (Table **??**: -0.838 versus -0.221 thousandths of a standard deviation per day). Allowing for heterogeneous slopes—using group-specific coefficients from Table **??** rather than the pooled estimate—amplifies the implied contribution substantially, since the same exposure difference is multiplied by a much larger effect size for the more disadvantaged group. Taken together, these decompositions suggest that pollution

not only reduces average achievement but does so in a way that compounds with existing educational disadvantage: the students who are most exposed to pollution are also the most harmed by it.

5.3 Implications of declining pollution over time

A complementary way to assess the aggregate importance of our estimates is to ask how much of the change in student achievement over our sample period can be accounted for by the secular decline in pollution exposure. Mean exposure to days with PM2.5 at or above 8 μg/m^3 fell by approximately 102 days over the estimation sample (Figure **??**), reflecting broad improvements in air quality driven by Clean Air Act regulation and the ongoing shift away from coal-fired power generation. Applying the IV per-day coefficient from the 8+ specification to this reduction implies an improvement of approximately 21 standardized-score points, or 0.021 standard deviations, attributable to declining pollution exposure over this period—equivalent to roughly 0.9 months of learning.

Two caveats are worth stating clearly. First, this calculation is descriptive: it scales a causal per-day estimate by an observed time-series change in exposure, but does not itself identify the causal effect of the pollution trend on achievement, since other factors affecting learning evolved concurrently. Second, the estimate holds the pollution–score relationship fixed over time, which may not be appropriate if the composition of pollution sources or the adaptive capacity of schools changed over the sample period. Subject to these qualifications, the calculation suggests that the long-run decline in ambient particulate matter has meaningfully supported student achievement, and conversely that policy reversals which raise PM2.5 concentrations back toward historical levels could impose achievement costs of similar magnitude.

6 Conclusion

This paper examines the relationship between ambient air pollution and student learning. Using longitudinal data on more than sixteen million U.S. high school students and a within-student research design, we assess how variation in PM2.5 exposure over the school year influences changes in academic performance across repeated test attempts. An instrumental variables strategy based on distant wildfire smoke isolates plausibly exogenous variation in local pollution, and suggests that pollution exposure causally reduces learning over the course of the academic year, rather than solely

influencing performance on the day of an exam.

The estimated effects are driven primarily by exposure to moderately elevated pollution levels and appear to be concentrated during periods when school is in session. We find little evidence that summer exposure plays a comparable role, which is consistent with mechanisms operating through instructional disruption or short-run cognitive effects rather than through longer-run health channels. We also document heterogeneity across students and schools: adverse learning effects are larger among students attending more economically challenged schools and among those with lower baseline achievement. While these patterns do not identify specific mechanisms, they suggest that pollution may interact with other constraints faced by students and schools.

Taken together, the results indicate that ambient air pollution – even at levels below prevailing official standards – is a relevant input into the learning process. Because even moderate pollution levels are common, small per-day effects may accumulate over time. More broadly, the findings suggest that analyses of educational outcomes and human capital formation may benefit from accounting for environmental conditions alongside more traditional inputs.

References

Andersen, C. M., J. Brandt, J. H. Christensen, L. M. Frohn, C. Geels, T. Hener, M. Simonsen, and L. Skipper (2025). Air pollution and cognition in children: Evidence from national tests in denmark. *Journal of Human Resources*.

Archsmith, J., A. G. Heyes, and S. Saberian (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.

Aziz, M. and L. Elbakidze (2025). Invisible barrier: The impact of air quality on chronic school absenteeism in the us. *Environmental and Resource Economics 88*(3), 655–680.

Bedi, A. S., M. Y. Nakaguma, B. J. Restrepo, and M. Rieger (2021). Particle pollution and cognition: Evidence from sensitive cognitive tests in Brazil. *Journal of the Association of Environmental and Resource Economists 8*(3), 443–474.

Bharadwaj, P., M. Gibson, J. G. Zivin, and C. Neilson (2017). Gray Matters: Fetal Pollution Exposure and Human Capital Formation. *Journal of the Association of Environmental and Resource Economists 4*(2), 505–542.

Biasi, B., J. Lafortune, and D. Schönholzer (2025). What works and for whom? effectiveness and efficiency of school capital investments across the us. *The Quarterly Journal of Economics*, qjaf013.

Bishop, K. C., J. D. Ketcham, and N. V. Kuminoff (2022, 12). Hazed and Confused: The Effect of Air Pollution On Dementia. *The Review of Economic Studies*. rdac078.

Burke, M., M. L. Childs, B. de la Cuesta, M. Qiu, J. Li, C. F. Gould, S. Heft-Neal, and M. Wara (2023). The contribution of wildfire to pm2. 5 trends in the usa. *Nature 622*(7984), 761–766.

Carneiro, J., M. A. Cole, and E. Strobl (2021). The effects of air pollution on students' cognitive performance: evidence from brazilian university entrance tests. *Journal of the Association of Environmental and Resource Economists 8*(6), 1051–1077.

Carneiro, J., M. A. Cole, and E. Strobl (2024). Foetal exposure to air pollution and students' cognitive performance: Evidence from agricultural fires in brazil. *Oxford Bulletin of Economics and Statistics 86*(1), 156–186.

Cavalcanti, P. H. (2024). Agricultural fires and student performance: Evidence from brazil.

Chen, S., C. Guo, and X. Huang (2018). Air pollution, student health, and school absences: Evidence from China. *Journal of Environmental Economics and Management 92*, 465–497.

Childs, M. L., J. Li, J. Wen, S. Heft-Neal, A. Driscoll, S. Wang, C. F. Gould, M. Qiu, J. Burney, and M. Burke (2022). Daily local-level estimates of ambient wildfire smoke pm2. 5 for the contiguous us. *Environmental Science & Technology 56*(19), 13607–13621.

Colmer, J., I. Hardman, J. Shimshack, and J. Voorheis (2020). Disparities in pm2. 5 air pollution in the United States. *Science 369*(6503), 575–578.

Colmer, J. and J. Voorheis (2020, November). The grandkids aren't alright: The intergenerational effects of prenatal pollution exposure. Working paper, Center for Economic Studies, U.S. Census Bureau.

Connolly, R., M. E. Marlier, D. A. Garcia-Gonzales, J. Wilkins, J. Su, C. Bekker, J. Jung, E. Bonilla, R. T. Burnett, Y. Zhu, et al. (2024). Mortality attributable to pm2. 5 from wildland fires in california from 2008 to 2018. *Science advances 10*(23), eadl1252.

Conte Keivabu, R. and T. Rüttenauer (2022). London congestion charge: the impact on air pollution and school attendance by socioeconomic status. *Population and Environment 43*(4), 576–596.

Dang, H.-A., D. Minh, and C. V. Nguyen (2024). The impacts of climate change and air pollution on children's education outcomes: Evidence from viet nam. Technical report, ADBI Working Paper.

Dang, H.-A. H., M. N. Do, and C. V. Nguyen (2025). The impacts of climate change and air pollution on children's education outcomes: Evidence from vietnam. *Economics of Education Review 106*, 102656.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review 109*(12), 4178–4219.

Di, Q., H. Amini, L. Shi, I. Kloog, R. Silvern, J. Kelly, M. B. Sabath, C. Choirat, P. Koutrakis, A. Lyapustin, et al. (2019). An ensemble-based model of pm2. 5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment international 130*, 104909.

Dong, R., R. Fisman, Y. Wang, and N. Xu (2021). Air pollution, affect, and forecasting bias: Evidence from Chinese financial analysts. *Journal of Financial Economics 139*(3), 971–984.

Ebenstein, A., V. Lavy, and S. 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.

Feng, Y., E. Castro, Y. Wei, T. Jin, X. Qiu, F. Dominici, and J. Schwartz (2024). Long-term exposure to ambient pm2. 5, particulate constituents and hospital admissions from non-respiratory infection. *Nature Communications 15*(1), 1518.

Garg, T., M. Jagnani, and V. Taraz (2020). Temperature and human capital in India. *Journal of the Association of Environmental and Resource Economists 7*(6), 1113–1150.

Gethin, A. (2025). Distributional growth accounting: Education and the reduction of global poverty, 1980–2019. *The Quarterly Journal of Economics 140*(4), 2571–2618.

Gilraine, M. and A. Zheng (2022, May). Air Pollution and Student Performance in the U.S. Working Paper 30061, National Bureau of Economic Research.

Gould, C. F., S. Heft-Neal, M. Johnson, J. Aguilera, M. Burke, and K. Nadeau (2024). Health effects of wildfire smoke exposure. *Annual Review of Medicine 75*(1), 277–292.

Graff Zivin, J., Y. Song, Q. Tang, and P. Zhang (2020). Temperature and high-stakes cognitive performance: Evidence from the national college entrance examination in China. *Journal of Environmental Economics and Management 104*, 102365.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annu. Rev. Econ. 6*(1), 689–733.

Heissel, J. A., C. Persico, and D. Simon (2022). Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance. *Journal of Human Resources 57*(3), 747–776.

Huang, J., N. Xu, and H. Yu (2020, October). Pollution and Performance: Do Investors Make Worse Trades on Hazy Days? *Management Science 66*(10), 4455–4476.

Jbaily, A., X. Zhou, J. Liu, T.-H. Lee, L. Kamareddine, S. Verguet, and F. Dominici (2022). Air pollution exposure disparities across us population and income groups. *Nature 601*(7892), 228–233.

Koivuranta, M., M. Korhonen, and J. Lehto (2024). The impact of ambient pm2. 5 air pollution on student performance: Evidence from finnish matriculation examinations. *Economics Letters 235*, 111542.

Komisarow, S. and E. L. Pakhtigian (2022). Are power plant closures a breath of fresh air? local air quality and school absences. *Journal of Environmental Economics and Management*, 102569.

La Nauze, A. and E. R. Severnini (2021). Air pollution and adult cognition: Evidence from brain training. Technical report, National Bureau of Economic Research.

Lee, D. S., J. McCrary, M. J. Moreira, and J. Porter (2020). Valid t-ratio inference for iv. *arXiv preprint arXiv:2010.05058*.

Marcotte, D. E. (2015). Allergy test: Seasonal allergens and performance in school. *Journal of health economics 40*, 132–140.

Martin, D., D. A. Romero, and D. Salcedo (2025). The hidden cost of mobility: Public transit pollution and academic achievement.

Park, R. J. (2022). Hot Temperature and High-Stakes Performance. *Journal of Human Resources 57*(2), 400–434.

Park, R. J., A. P. Behrer, and J. Goodman (2021, January). Learning is inhibited by heat exposure, both internationally and within the United States. *Nature Human Behaviour 5*(1), 19–27.

Park, R. J., J. Goodman, M. Hurwitz, and J. Smith (2020, May). Heat and learning. *American Economic Journal: Economic Policy 12*(2), 306–39.

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

Pope III, C. A., N. Coleman, Z. A. Pond, and R. T. Burnett (2020). Fine particulate air pollution and human mortality: 25+ years of cohort studies. *Environmental research 183*, 108924.

Qin, Y., J. Wu, and J. Yan (2019). Negotiating housing deal on a polluted day: Consequences and possible explanations. *Journal of Environmental Economics and Management 94*, 161–187.

Rentschler, J. and N. Leonova (2023). Global air pollution exposure and poverty. *Nature communications 14*(1), 4432.

Shi, S., M. Qin, B. Shen, Y. Cai, T. Liu, F. Yang, W. Gong, X. Liu, J. Liang, Q. Zhao, et al. (2020). Association of cardiac injury with mortality in hospitalized patients with covid-19 in wuhan, China. *JAMA cardiology*.

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

Voorheis, J. (2017, May). Air quality, human capital formation and the long-term effects of environmental inequality at birth. Working paper, Center for Administrative Records Research and Applications, U.S. Census Bureau.

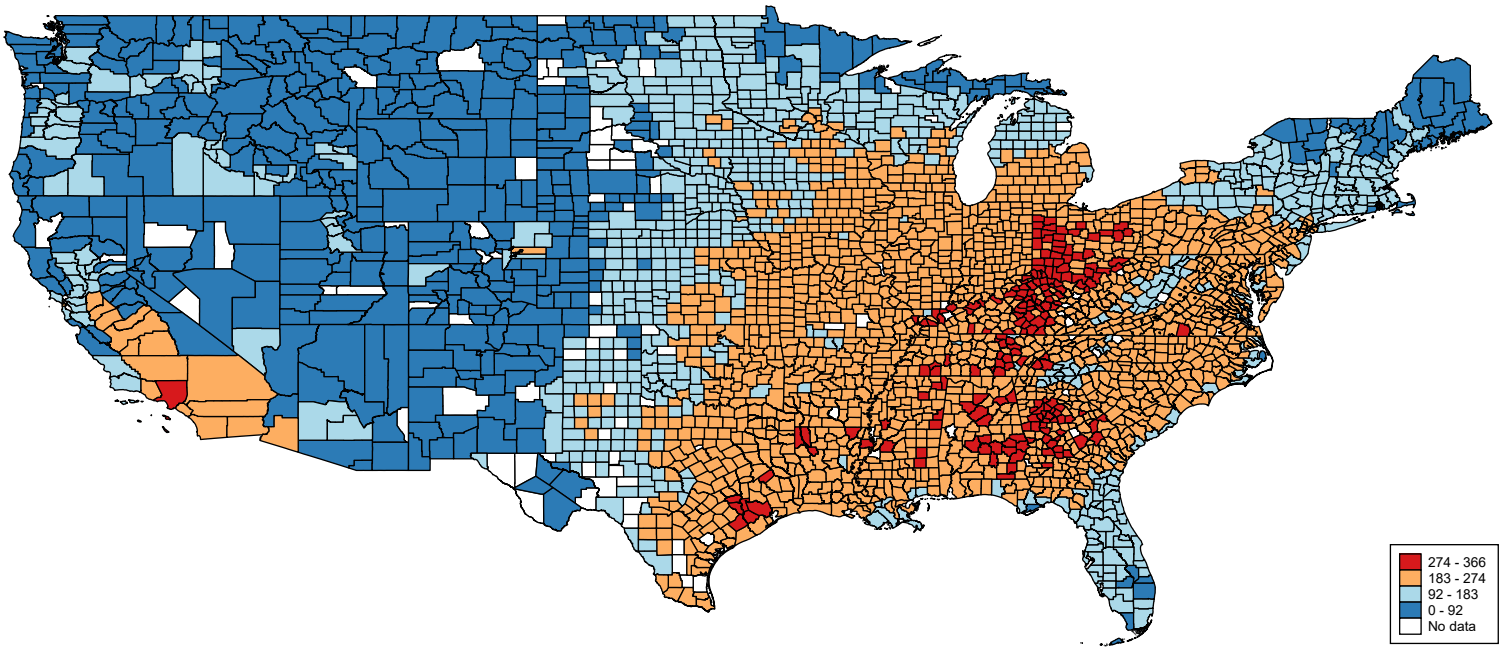
Wen, J. and M. Burke (2022a). Lower test scores from wildfire smoke exposure. *Nature Sustainability 5*(11), 947–955.

Wen, J. and M. Burke (2022b). Lower test scores from wildfire smoke exposure. *Nature Sustainability 5*(11), 947–955.

Zhang, Q., Y. Wang, Q. Xiao, G. Geng, S. J. Davis, X. Liu, J. Yang, J. Liu, W. Huang, C. He, et al. (2025). Long-range pm2. 5 pollution and health impacts from the 2023 canadian wildfires. *Nature*, 1–7.

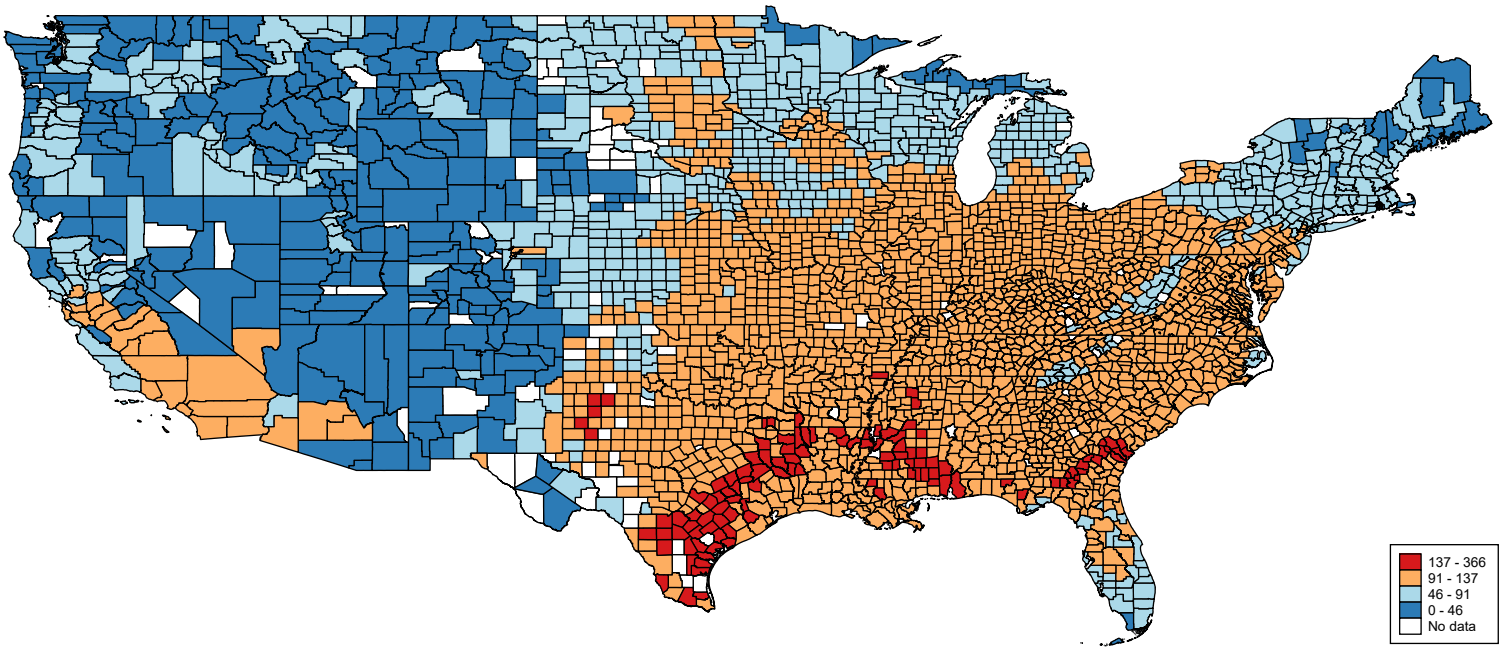
Zhang, X., H. Liu, X. Wu, L. Jia, K. Gadhave, L. Wang, K. Zhang, H. Li, R. Chen, R. Kumbhar, et al. (2025). Lewy body dementia promotion by air pollutants. *Science 389*(6764), eadu4132.

Figure 1: Spatial Distribution of Pollution



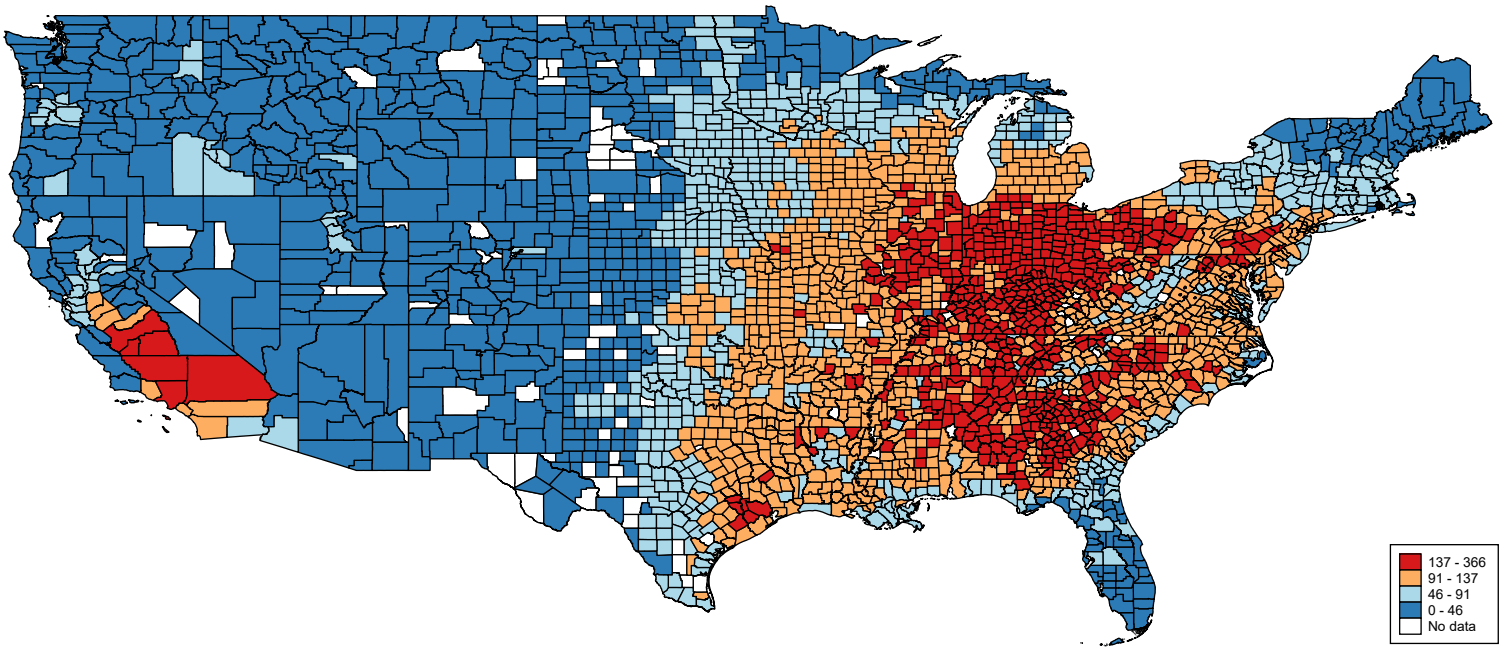
Notes: Shown above is the county-level average annual number of polluted days for PSAT takers in our sample. Polluted days are those with PM2.5 of at least 8 $\mu g/m^3$.

Figure 2: Spatial Distribution of Moderately Polluted Days



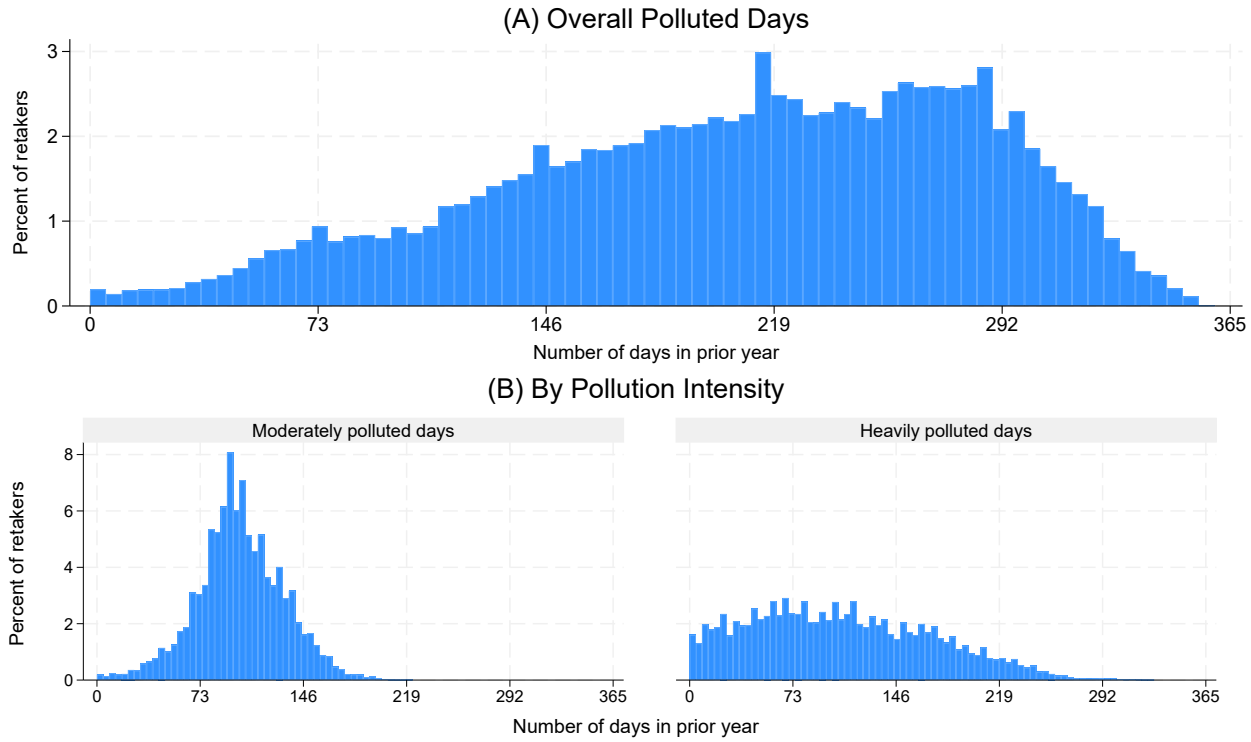
Notes: The above figure shows county-level average number of polluted days in PSAT exam years for all students from the high school classes of 2001-22 who took the PSAT at least once in grades 9-11. We defined a moderately polluted day as one PM2.5 was $\geq$8 and <12. County refers to the physical location of the high school the student last attended. Categories 3 and 2 represent the county-level mean +/- the county-level standard deviation. Pollution in PSAT years 1997-2000 and 2017-2019 was not available. COVID-impacted PSAT years 2020 and 2021 were also excluded.

Figure 3: Spatial Distribution of Heavily Polluted Days



Notes: The above figure shows county-level average number of polluted days in PSAT exam years for all students from the high school classes of 2001-22 who took the PSAT at least once in grades 9-11. We defined a heavily polluted day as one PM2.5 was $\geq$12. County refers to the physical location of the high school the student last attended. Categories 3 and 2 represent the county-level mean +/- the county-level standard deviation. Pollution in PSAT years 1997-2000 and 2017-2019 was not available. COVID-impacted PSAT years 2020 and 2021 were also excluded.

Figure 4: Student-Level Variation in Pollution Exposure



Notes: Shown above is the distribution of students' average number of polluted days experienced at their school location in the year prior to their first PSAT take. School location is defined using a 2 km zone around the school's latitude and longitude, if available, or else using the school's ZIP code. The graphs include all students from the high school classes of 2001-22 who took the PSAT more than once in grades 9-11 and had available pollution data. In panel A, polluted days are those with PM2.5 of at least 8 μg/m^3. Panel B distinguishes between moderately polluted days (PM2.5 of 8-12 μg/m^3) and heavily polluted days (PM2.5 $>$ 12 μg/m^3).

Table 1: Summary Statistics

		Eligible for Analysis	
	All Retakers	Retakers with PM2.5 Data	Retakers with PM2.5 and Smoke Data
Panel A. Demographics			
Female	0.54	0.54	0.53
White	0.53	0.54	0.51
Asian	0.09	0.09	0.09
Black	0.13	0.14	0.14
Hispanic	0.20	0.20	0.23
Highest Parent Education	6.38	6.40	6.33
High School Challenge	40.21	40.39	41.15
Panel B. PSAT scores			
PSAT Attempts	2.15	2.16	2.18
First PSAT z-score	-0.00	-0.01	-0.02
Panel C. Pollution			
Prior-year 8+ pollution days	207.8	208.1	190.3
Moderate [8,12) pollution days	101.0	101.2	103.3
Heavy [12,inf) pollution days	106.7	107.0	87.0
Schoolyear 8+ pollution days	156.1	156.4	141.5
Summer 8+ pollution days	51.6	51.8	48.8
Panel D. Other weather variables			
Prior-year mean temperature	59.4	59.5	59.8
Prior-year mean precipitation	0.1	0.1	0.1
N(attempts)	36,483,054	27,673,932	18,793,469
N(students)	17,008,108	13,661,615	9,821,498
N(schools)	28,257	26,695	24,838

Notes: Means of key variables are shown. Column 1 contains all students in our sample who took the PSAT at least twice, excluding COVID-affected years of 2020 and 2021. Column 2 contains students who took the PSAT at least twice in years with available pollution data. Column 3 contains students who took the PSAT at least twice in years with available pollution and smoke data, excluding school-years with distance to fires $\leq$10 km (the sample analyzed with instrumental variables regression). High school challenge is a percentile relative to all U.S. high schools, with higher numbers representing more disadvantaged communities. Polluted days are those with PM2.5 of at least 8 μg/m^3. These are also segmented into pollution days during the previous school year versus summer. Pollution days in the previous year are also segmented into bins: moderately polluted ($\leq$8 PM2.5 <12 μg/m^3) and heavily polluted (PM2.5 $\geq$ 12 μg/m^3) days. All weather variables were measured in the year prior to each student's first PSAT attempt year.

Table 2: Summary Statistics by Subgroups of Retakers

	Student Race				Challenge of High School		
	White	Asian	Black	Hispanic	Low	Medium	High
Panel B. PSAT Scores.							
PSAT Attempts	2.17	2.19	2.20	2.18	2.17	2.19	2.18
First PSAT z-score	0.27	0.49	-0.62	-0.47	0.35	-0.07	-0.60
Panel C. Pollution.							
Prior-year 8+ pollution days	189.8	196.4	210.4	197.1	188.3	191.1	212.2
Moderate [8,12) pollution days	99.3	103.4	106.1	109.4	98.4	102.8	111.6
Heavy [12,inf) pollution days	90.5	93.0	104.3	87.8	89.9	88.3	100.6
Schoolyear 8+ pollution days	140.7	148.6	156.5	147.2	139.3	142.5	159.4
Summer 8+ pollution days	49.1	47.8	53.9	50.0	49.0	48.6	52.9
Panel D. Other Weather Variables.							
Prior-year mean temperature	58.1	59.2	60.5	63.5	57.1	60.8	63.1
Prior-year mean precipitation	0.1	0.1	0.1	0.1	0.1	0.1	0.1
N(attempts)	9,415,522	1,625,764	2,600,039	4,107,552	7,393,692	7,440,489	3,860,008
N(students)	4,901,415	856,583	1,331,871	2,180,268	3,851,276	3,896,702	2,017,837
N(schools)	15,185	1,496	3,288	3,526	5,604	10,731	5,591

Notes: Means of key variables are shown. The table contains subgroups of students who took the PSAT at least twice in years with available pollution and smoke data, excluding school-years with distance to fires $\leq$10 km (the sample analyzed with instrumental variables regression). High school challenge is a percentile relative to all U.S. high schools, with higher numbers representing more disadvantaged communities. Low challenge was defined as <25, medium challenge as 25-74, high challenge as $\geq$75. Polluted days are those with $PM_{2.5}$ of at least 8 $\mu g/m^3$. These are also segmented into pollution days during the previous school year versus summer. Pollution days in the previous year are also segmented into bins: moderately polluted ($\leq$8 $PM_{2.5}$ <12 $\mu g/m^3$) and heavily polluted ($PM_{2.5} \geq 12$ $\mu g/m^3$) days. All weather variables were measured in the year prior to each student's first PSAT attempt year.

Table 3: OLS - Impact of Prior Year Pollution on Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)
(A) Overall pollution						
Pollution days	-0.006	-0.021	-0.021	-0.022	-0.017	-0.017
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
(B) By pollution intensity						
Moderate pollution days	-0.053**	-0.057**	-0.056**	-0.057**	-0.067***	-0.065***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Heavy pollution days	0.029	0.007	0.006	0.006	0.023	0.021
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
N	16,742,701	16,742,701	16,742,701	16,742,701	16,742,701	16,742,701
Prior year weather	No	Yes	Yes	Yes	Yes	Yes
Exam day pollution	No	No	Yes	Yes	Yes	Yes
Exam day weather	No	No	No	Yes	Yes	Yes
Division-specific trends	No	No	No	No	Yes	Yes
County unemployment rate	No	No	No	No	No	Yes

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measures shown. Pollution is measured as days in designated bin ranges in the 365 days preceding the PSAT take. In panel A, polluted days are those with $PM_{2.5}$ of at least 8 $\mu g/m^3$. Panel B distinguishes between moderately polluted days ($\geq$8 $PM_{2.5}$ <12 $\mu g/m^3$) and heavily polluted days ($PM_{2.5} \geq 12$ $\mu g/m^3$). All regressions include student fixed effects and fixed effects for each combination of cohort, test date and take number. Column 2 adds controls for prior year temperature and precipitation. Column 3 adds controls for test day pollution levels. Column 4 adds controls for test day temperature and precipitation. Column 5 adds 9 Census division-specific linear time trends. Column 6 adds controls for the county-level unemployment rate in the PSAT take year. The sample comprises all students from the high school classes of 2007-16 who took the PSAT more than once and had available pollution and smoke data (the sample that could be analyzed using instrumental variables).

Table 4: Instrumental Variables - Impact of Prior Year Smoke Pollution on PSAT Scores (First Stage)

	(1)	(2)	(3)	(4)	(5)	(6)
(A) First stage: Moderate pollution days						
Light smoke pollution days	0.412***	0.453***	0.453***	0.456***	0.459***	0.460***
	(0.014)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)
Moderate smoke pollution days	0.353***	0.407***	0.407***	0.403***	0.408***	0.385***
	(0.054)	(0.055)	(0.055)	(0.056)	(0.056)	(0.056)
Heavy smoke pollution days	-1.180***	-1.135***	-1.135***	-1.138***	-1.111***	-1.089***
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
(B) First stage: Heavy pollution days						
Light smoke pollution days	-0.201***	-0.154***	-0.133***	-0.146***	-0.146***	-0.148***
	(0.016)	(0.017)	(0.016)	(0.017)	(0.017)	(0.016)
Moderate smoke pollution days	1.510***	1.493***	1.522***	1.644***	1.619***	1.651***
	(0.057)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)
Heavy smoke pollution days	1.465***	1.451***	1.457***	1.485***	1.471***	1.439***
	(0.050)	(0.050)	(0.049)	(0.050)	(0.050)	(0.049)
F-statistic:	331.62	354.98	368.04	384.52	384.24	384.73
Prior year temperature and precipitation	No	Yes	Yes	Yes	Yes	Yes
Exam day pollution	No	No	Yes	Yes	Yes	Yes
Exam day temperature and precipitation	No	No	No	Yes	Yes	Yes
Region-specific linear trends	No	No	No	No	Yes	Yes
County-level unemployment rate	No	No	No	No	No	Yes

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01), for the first stage of the instrumental variables regression where moderate and heavy pollution days are predicted at the second stage. Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measure shown, where pollution is instrumented by pollution from smoke. Pollution is measured as days in designated bin ranges in the 365 days preceding the PSAT take. The reference category for the first-stage regression is the bin representing number of days with no smoke or 0 $PM_{2.5}$ from smoke. Light smoke days are days with smoke >0 $PM_{2.5}$ <8. Moderate smoke days are days with smoke ≥8 $PM_{2.5}$ <12. Heavy smoke days are days with smoke $PM_{2.5}$ ≥12 $\mu g/m^3$. All regressions include student fixed effects and fixed effects for each combination of cohort, test date and take number. Column 2 adds controls for prior year temperature and precipitation. Column 3 adds controls for test day pollution levels. Column 4 adds controls for test day temperature and precipitation. Column 5 adds 9 Census division-specific linear time trends. Column 6 adds controls for the county-level unemployment rate in the PSAT take year. Attempt-student rows for schools ≤10 km from a fire in a given year were excluded. The sample comprises all students from the high school classes of 2007-16 who took the PSAT more than once and had available pollution and smoke data.

Table 5: Instrumental Variables - Impact of Prior Year Smoke Pollution on PSAT Scores (Second Stage)

	(1)	(2)	(3)	(4)	(5)	(6)
(A) Second Stage: Overall pollution						
Pollution days	-0.276***	-0.371***	-0.359***	-0.245***	-0.208***	-0.200***
	(0.053)	(0.054)	(0.054)	(0.052)	(0.051)	(0.051)
(B) Second stage: By pollution intensity						
Moderate pollution days	-0.482***	-0.545***	-0.534***	-0.498***	-0.508***	-0.506***
	(0.078)	(0.078)	(0.078)	(0.077)	(0.077)	(0.077)
Heavy pollution days	-0.174***	-0.264***	-0.256***	-0.106*	-0.022	-0.025
	(0.054)	(0.057)	(0.057)	(0.055)	(0.054)	(0.054)
N	16,742,701	16,742,701	16,742,701	16,742,701	16,742,701	16,742,701
Prior year temperature and precipitation	No	Yes	Yes	Yes	Yes	Yes
Exam day pollution	No	No	Yes	Yes	Yes	Yes
Exam day temperature and precipitation	No	No	No	Yes	Yes	Yes
Region-specific linear trends	No	No	No	No	Yes	Yes
County-level unemployment rate	No	No	No	No	No	Yes

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01), for the second stage of the instrumental variables regression(s). Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measure shown, where pollution is instrumented by pollution from smoke. Pollution is measured as days in designated bin ranges in the 365 days preceding the PSAT take. In Panel A, pollution from smoke is instrumenting for pollution days ≥8. In Panel B, pollution from smoke is instrumenting for moderately polluted days (8≤ $PM_{2.5}$ <12 $\mu g/m^3$) and heavily polluted days ($PM_{2.5}$ ≥12 $\mu g/m^3$). The reference category for the first-stage regression for both panels is the bin representing number of days with no smoke or 0 $PM_{2.5}$ from smoke, using the instruments shown in Table 4. All regressions include student fixed effects and fixed effects for each combination of cohort, test date and take number. Column 2 adds controls for prior year temperature and precipitation. Column 3 adds controls for test day pollution levels. Column 4 adds controls for test day temperature and precipitation. Column 5 adds 9 Census division-specific linear time trends. Column 6 adds controls for the county-level unemployment rate in the PSAT take year. Attempt-student rows for schools ≤10 km from a fire in a given year were excluded. The sample comprises all students from the high school classes of 2007-16 who took the PSAT more than once and had available pollution and smoke data.

Table 6: Instrumental Variables - Impact of School Year and Summer Pollution

	(1)	(2)	(3)
(A) Overall pollution			
School year pollution days	-0.097		-0.186**
	(0.068)		(0.078)
Summer pollution days		-0.039	0.329
		(0.193)	(0.234)
(B) By pollution intensity			
School year moderate pollution days	-0.347***		-0.702**
	(0.105)		(0.325)
School year heavy pollution days	-0.032		-0.152
	(0.067)		(0.137)
Summer moderate pollution days		-2.228***	0.587
		(0.411)	(1.009)
Summer heavy pollution days		-0.603***	0.551
		(0.218)	(0.486)
N	16,742,701	16,742,701	16,742,701

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measure shown, where pollution is instrumented by pollution from smoke. Pollution is measured as days in the year preceding the PSAT take, separating the calendar into days during the school year and days during the summer break. In panel A, polluted days are those with PM2.5 of at least 8 $\mu g/m^3$. Panel B distinguishes between moderately polluted days (PM2.5 of 8-12 $\mu g/m^3$) and heavily polluted days (PM2.5 $>$ 12 $\mu g/m^3$). The reference category for the instrument in the first-stage regression is the bin representing number of days with no smoke or 0 PM2.5 from smoke. Light smoke days are days with smoke 0$<$ PM2.5 $<$8. In Panel A, smoke days are classified as days with smoke PM2.5 $\geq$8. In Panel B, moderate and heavy smoke days are also distinguished. All regressions instrument pollution using pollution from smoke from distant fires and include student fixed effects and fixed effects for each combination of cohort, test date and take number, as well as controls for prior year temperature and precipitation, test day pollution levels, test day temperature and precipitation, and Census division-specific linear time trends. Attempt-student rows for schools $\leq$10 km from a fire in a given year were excluded. The sample comprises all students from the high school classes of 2001-22 who took the PSAT more than once and had available pollution and smoke pollution data.

Table 7: Heterogeneity by School Challenge

	OLS		Instrumental Variables	
	Low	High	Low	High
	(1)	(2)	(3)	(4)
(A) Overall pollution				
School year pollution days	0.011	-0.028	-0.160**	-0.208*
	(0.021)	(0.033)	(0.076)	(0.112)
(B) By pollution intensity				
School year moderate pollution days	-0.023	-0.162***	-0.221**	-0.838***
	(0.029)	(0.046)	(0.102)	(0.230)
School year heavy pollution days	0.042	0.071*	-0.020	-0.250**
	(0.029)	(0.043)	(0.080)	(0.111)
N	12,325,159	4,333,762	12,325,159	4,333,762

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measures shown. In columns 1-2, regression is estimated using OLS. In columns 3-4, regression is estimated using two-stage least squares instrumental variables. Pollution is measured as days in the year preceding the PSAT take, counting only days during the school year. In panel A, polluted days are those with $PM_{2.5}$ of at least 8 μ/m^3. Panel B distinguishes between moderately polluted days ($PM_{2.5}$ of 8-12 μ/m^3) and heavily polluted days ($PM_{2.5} > 12$ μ/m^3). All regressions include student fixed effects and fixed effects for each combination of cohort, test date and take number, as well as controls for prior year temperature and precipitation, test day pollution levels, test day temperature and precipitation, and Census division-specific linear time trends. The sample comprises all students from the high school classes of 2007-16 who took the PSAT more than once and had available pollution data (took the PSAT at least twice in exam years 2007-2016), and whose school in a given attempt year was more than 10 km from the nearest fire. Challenge is a nationally normed percentile developed by the College Board to measure how challenging it is for a given high school's students to access educational and socioeconomic opportunities. It incorporates data on college attendance, household structure, family income, and other measures from a variety of sources. Estimates are presented by a fixed national Challenge cutoff: high-challenge schools have Challenge ≥ 67 (national percentile), and low-challenge schools have Challenge < 67.

Table 8: Heterogeneity by Initial Test Score Relative to National Mean

	OLS		Instrumental Variables	
	At or Above (1)	Below (2)	At or Above (3)	Below (4)
(A) Overall pollution				
Pollution days	0.028*	-0.043**	-0.224***	-0.056
	(0.015)	(0.020)	(0.056)	(0.067)
(B) By pollution intensity				
Moderate pollution days	0.026	-0.128***	-0.250***	-0.738***
	(0.021)	(0.027)	(0.073)	(0.091)
Heavy pollution days	0.030	0.023	-0.194***	0.046
	(0.020)	(0.027)	(0.057)	(0.068)
N	7,744,674	8,647,039	7,744,674	8,647,039

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measures shown. In columns 1-2, regression is estimated using OLS. In columns 3-4, regression is estimated using two-stage least squares instrumental variables. Pollution is measured as days in the year preceding the PSAT take. All regressions include student fixed effects and fixed effects for each combination of cohort, test date and take number, as well as controls for prior year temperature and precipitation, test day pollution levels, test day temperature and precipitation, and Census division-specific linear time trends. The sample comprises all students from the high school classes of 2001-22 who took the PSAT more than once and had available pollution data (took the PSAT at least twice in exam years 2007-2016), and whose school in a given attempt year was more than 10 km from the nearest fire. Each pair of columns divides students by their initial score relative to a given year's mean score among all test takers in the nation.

Appendix – For Online Publication

Table of Contents

A1 Appendix Exhibits **41**

A1.1 Appendix figures 41

A1.2 Potential maps for the appendix 44

A1.3 Appendix tables 44

A2 Data **51**

A2.1 Student Test Score Data 53

A2.2 Student Address and School Information 53

A2.3 School Geocoding 54

A2.4 Student Demographics 54

A2.5 Treatment Variables 55

A2.6 Instruments 55

A2.7 Additional Identifiers and Controls 56

A2.8 Heterogeneity Measures 56

A2.9 Analysis Samples 57

A3 Method for Interpreting Magnitudes **57**

A3.1 Student-Level Pollution Exposure and Scaling 57

A3.2 Average-Student Effects of Pollution Exposure 58

A3.3 Interpreting Magnitudes in Intuitive Units 59

A3.4 Counterfactual Changes in Pollution Exposure 64

A3.5 Decomposition of the Black–White Achievement Gap by Pollution Exposure 65

A3.6 Decomposition of Achievement Gaps by School Challenge Level 68

A3.7 Changes in Achievement Associated with Declining Pollution Exposure . . . 72

A1 Appendix Exhibits

A1.1 Appendix figures

1. Change in pollution distribution over time
2. Change in smoke distribution over time
3. Residualized pollution distribution
4. Residualized smoke distribution
5. Histogram of days by level of pollution

Figure A1

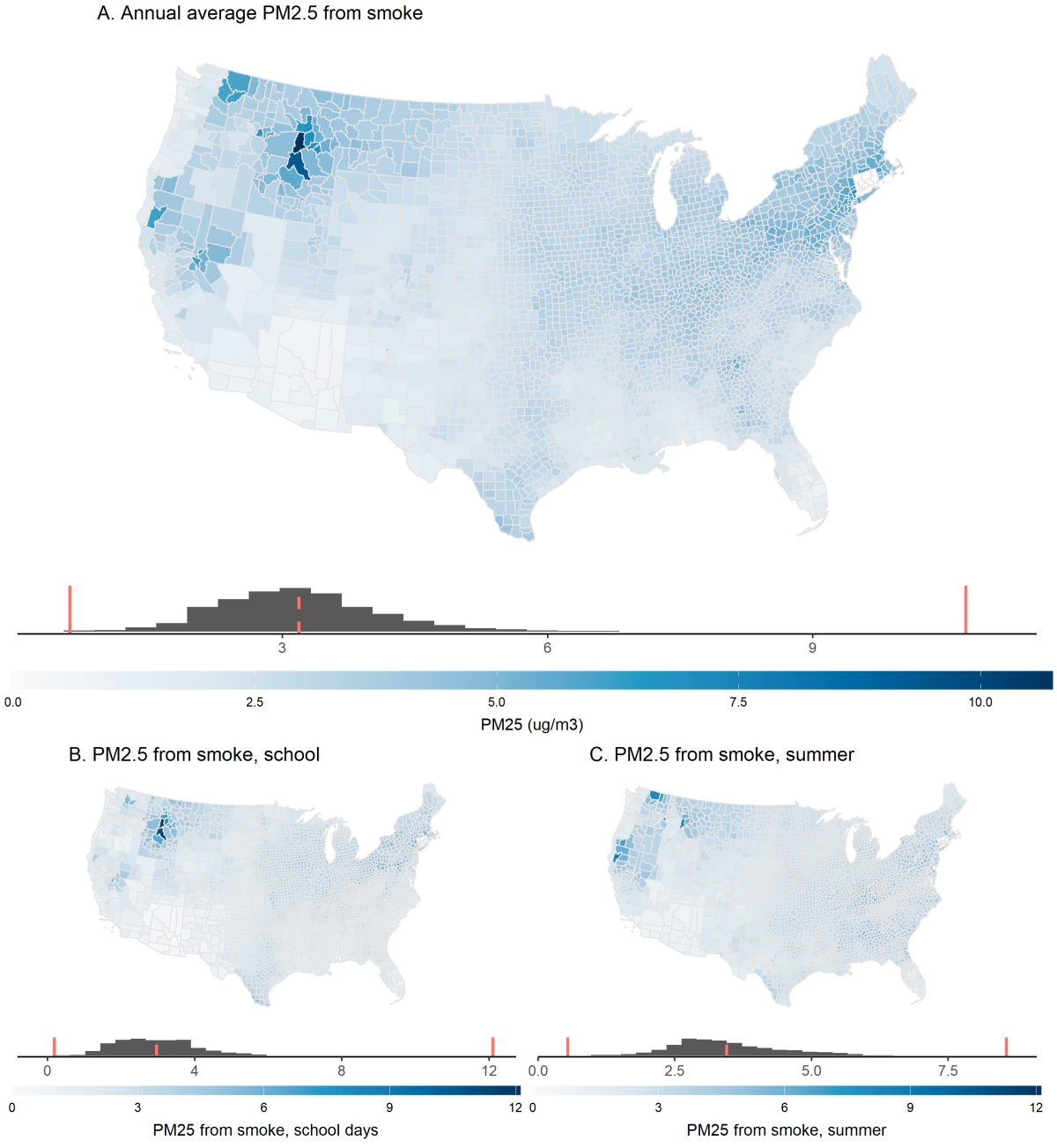

A. Annual average PM2.5 from smoke
3
6
9
0.0
2.5
5.0
7.5
10.0
PM25 (ug/m3)
B. PM2.5 from smoke, school
C. PM2.5 from smoke, summer
0
4
8
12
0
3
6
9
12
PM25 from smoke, school days
0.0
2.5
5.0
7.5
0
3
6
9
12
PM25 from smoke, summer

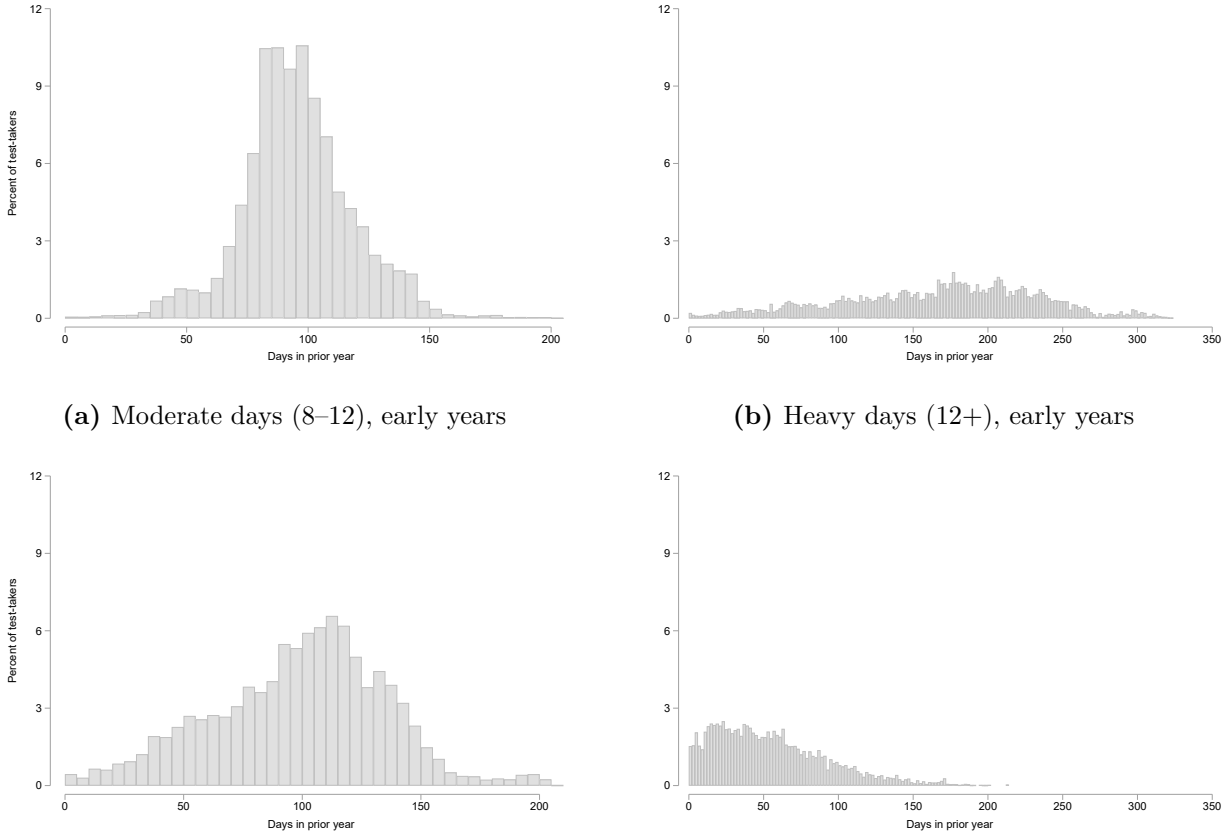
Figure A2


A. Average pre-2005 PM2.5
B. Average post-2010 PM2.5
5
10
15
2.5
5.0
7.5
10.0
12.5
C. Pre-2005 PM2.5, school
D. Pre-2005 PM2.5, summer
5
10
15
5
10
15
20
25
E. Post-2010 PM2.5, school
F. Post-2010 PM2.5, summer
5
10
2.5
5.0
7.5
10.0
12.5
0
5
10
15
20
PM25 (ug/m3)

A1.2 Potential maps for the appendix

A1.3 Appendix tables

1. Main specification with robustness to alternative trends – OLS
2. Main specification with robustness to alternative trends – IV
3. IV results with different distance thresholds
4. IV results using LIML instead of 2SLS



(a) Moderate days (8–12), early years

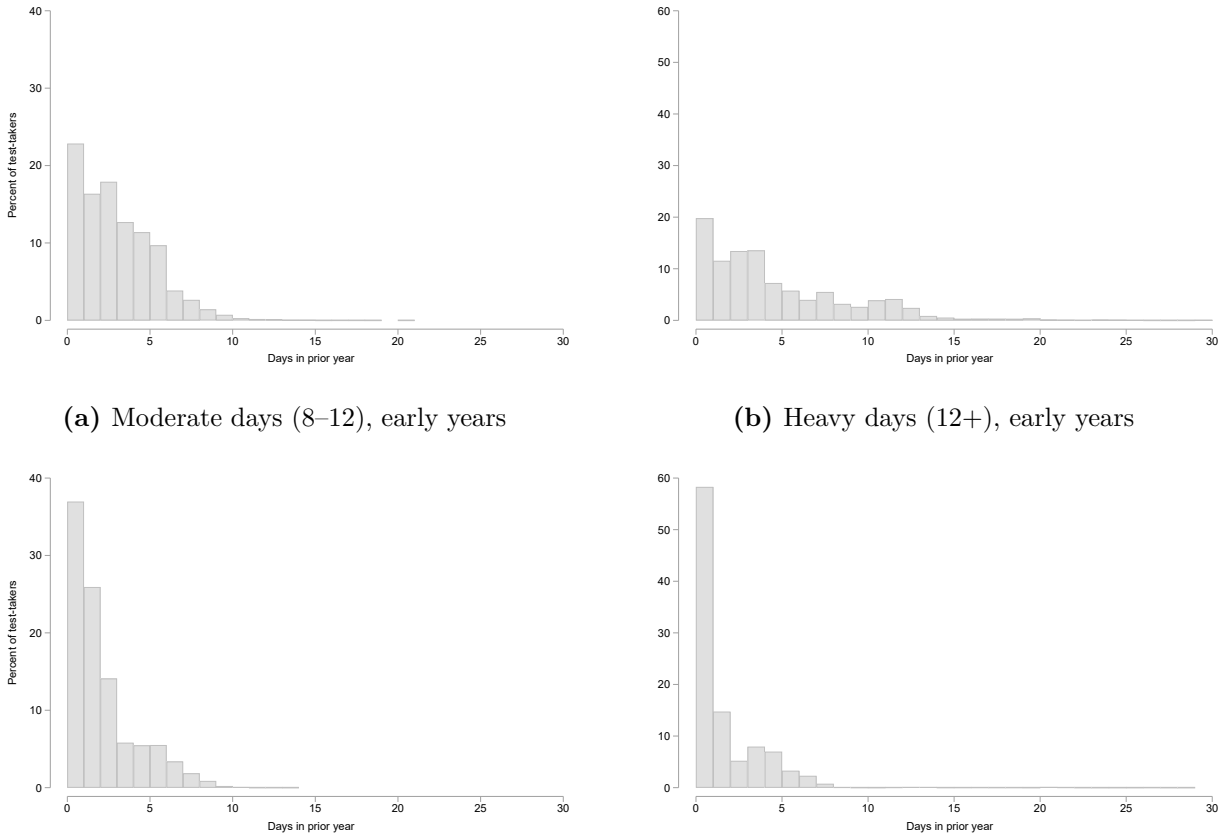
(b) Heavy days (12+), early years

(c) Moderate days (8–12), late years

(d) Heavy days (12+), late years

Figure A3: Distribution of PM2.5 polluted days, early vs. late sample years

Notes: Each panel plots a weighted histogram of the number of days in the year prior to the PSAT exam with PM2.5 in the indicated range. Moderate days are defined as $[8, 12)$ μg/m^3 and heavy days as ≥ 12 μg/m^3. Histograms are constructed at the school-by-exam-year level and weighted by the number of PSAT retakers in each school-year. The sample is restricted to exam years 2001–2016. Early years are defined as 2001-2002; late years as 2015–2016.



(a) Moderate days (8–12), early years

(b) Heavy days (12+), early years

(c) Moderate days (8–12), late years

(d) Heavy days (12+), late years

Figure A4: Distribution of PM2.5 polluted days from wildfire smoke, early vs. late sample years
Notes: Each panel plots a weighted histogram of the number of days in the year prior to the PSAT exam with PM2.5 in the indicated range, attributable to wildfire smoke. Moderate days are defined as $[8, 12)$ μg/m^3 and heavy days as ≥ 12 μg/m^3. Histograms are constructed at the school-by-exam-year level and weighted by the number of PSAT retakers in each school-year. The sample is restricted to exam years with available smoke data, 2007–2016. Early years are defined as 2007-2008; late years as 2015–2016.

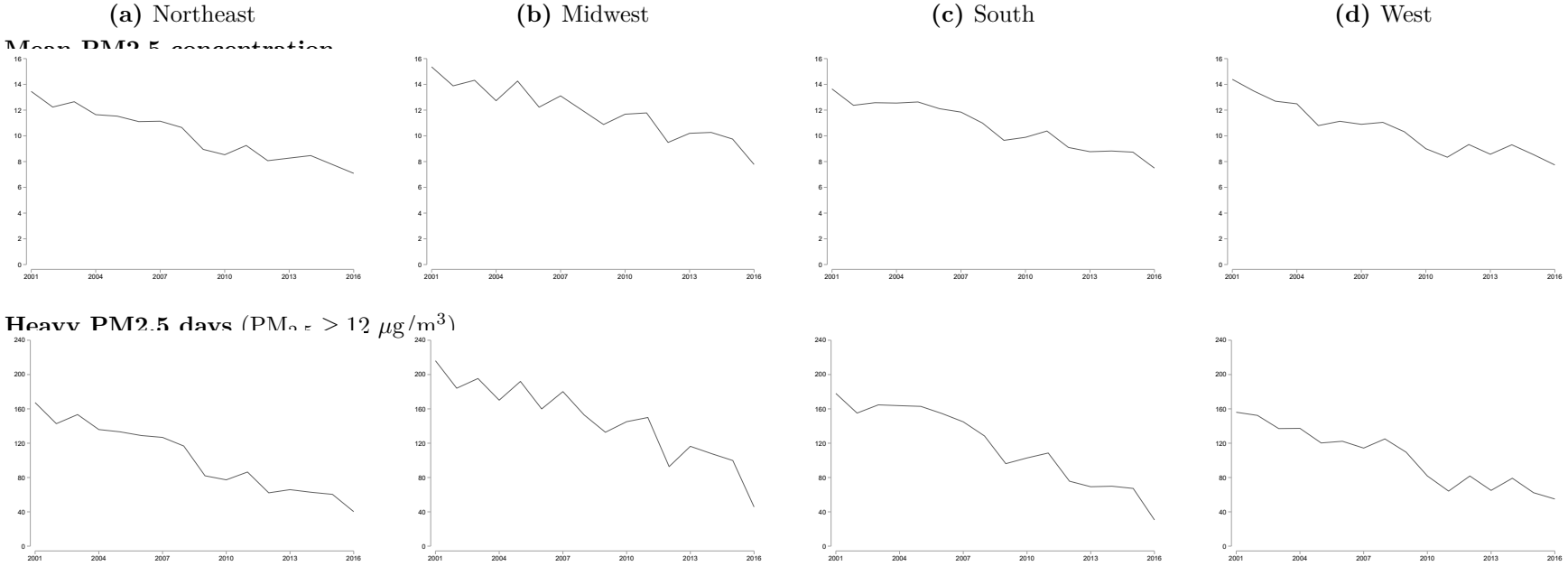


Figure A5: Regional trends in PM2.5 exposure

Notes: The top row plots student-weighted mean PM2.5 concentrations in the year prior to the PSAT exam. The bottom row plots the student-weighted mean number of heavy PM2.5 days (days with $PM_{2.5} \geq 12$ $\mu g/m^3$) in the prior year. Exposure measures are aggregated to Census region by exam year and weighted by the number of PSAT takers in each school–year. The sample includes exam years 2001–2016.

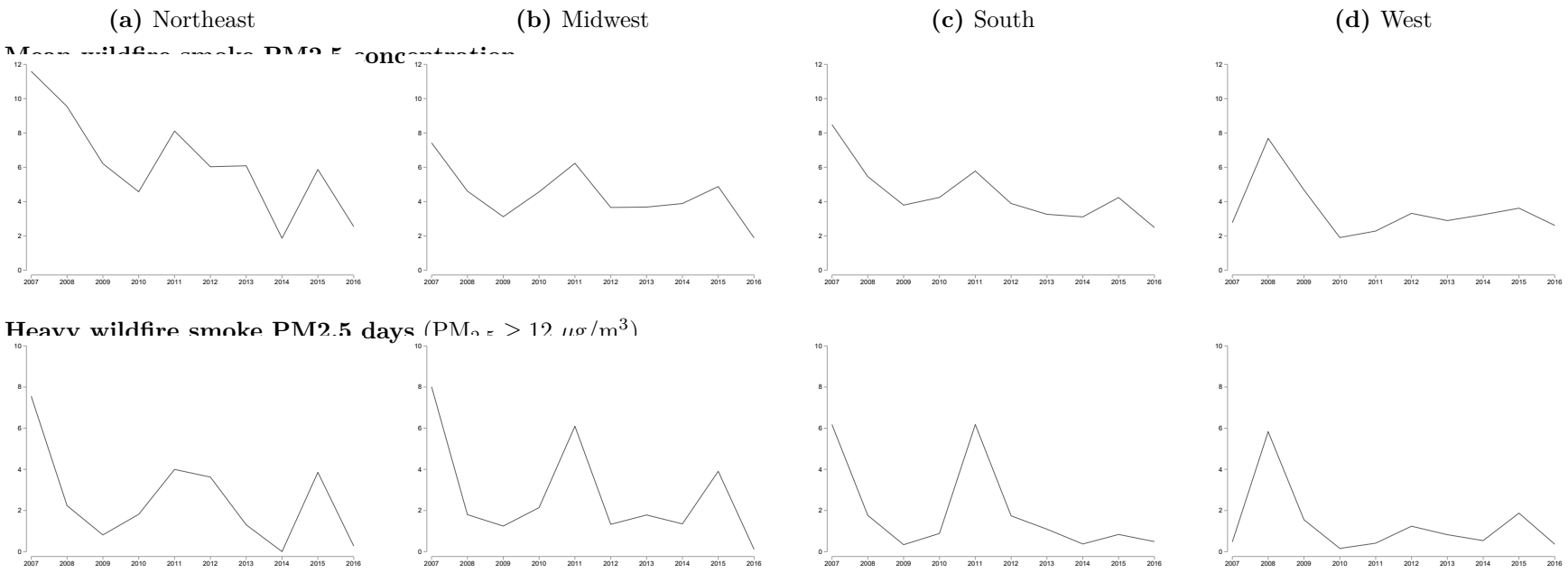


Figure A6: Regional trends in wildfire smoke PM2.5 exposure

Notes: The top row plots the student-weighted mean concentration of PM2.5 attributable to wildfire smoke in the year prior to the PSAT exam. The bottom row plots the student-weighted mean number of heavy wildfire smoke PM2.5 days (days with $PM_{2.5} \geq 12\ \mu g/m^3$) in the prior year. Exposure measures are derived from the ECHO wildfire smoke dataset (Burke et al.), aggregated to Census region by exam year, and weighted by the number of PSAT takers in each school–year. The sample includes exam years 2007–2016, the period for which wildfire smoke PM2.5 data are available.

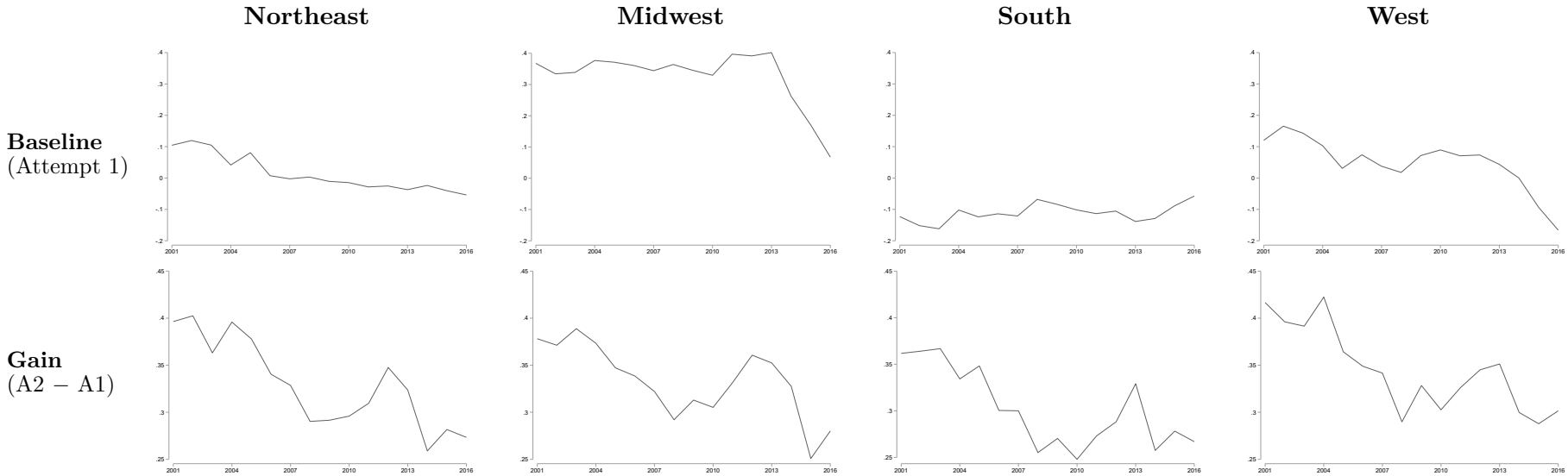
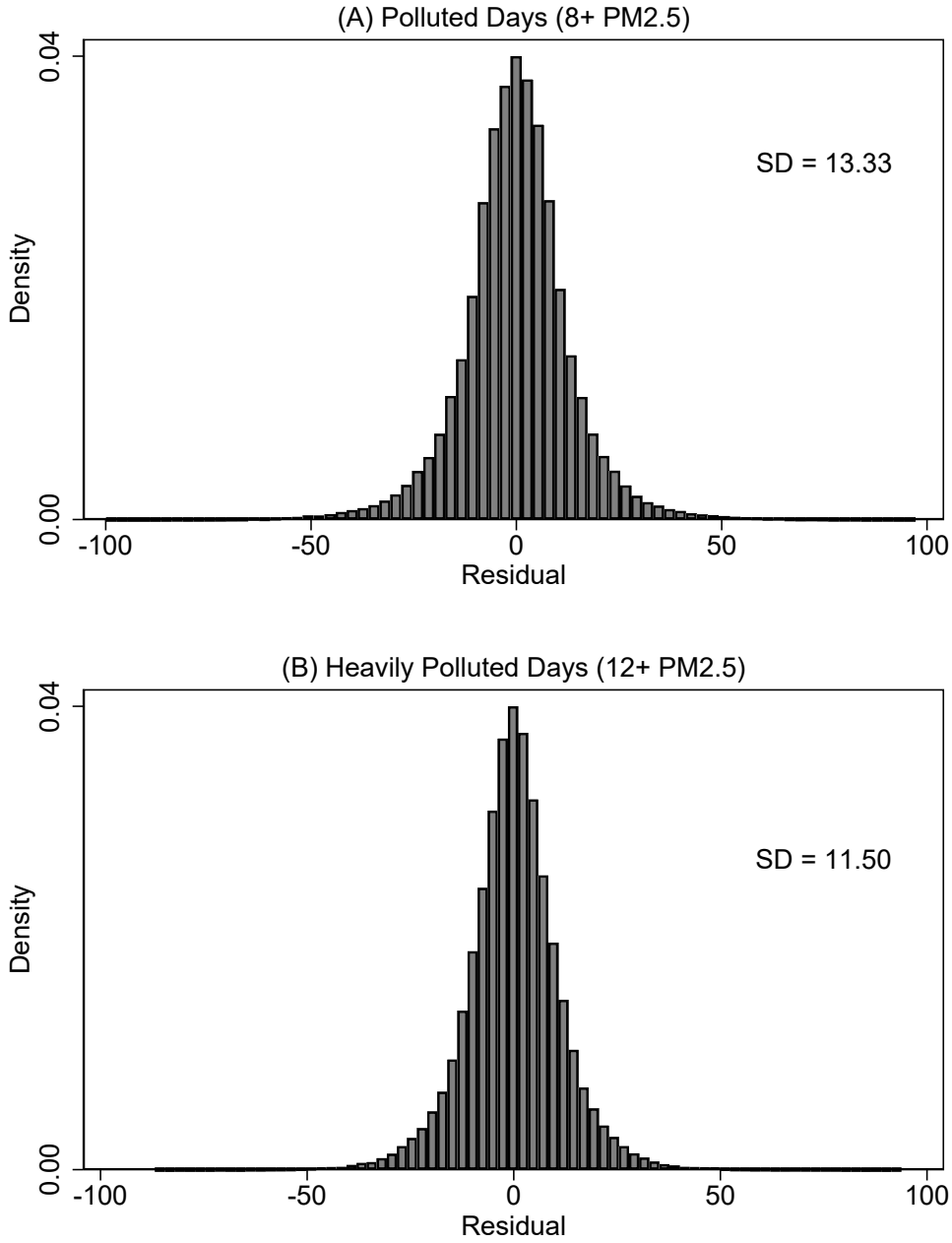


Figure A7: Cohort trends in baseline PSAT performance and learning gains by Census region

Notes: Cohorts are defined by the year of a student's first PSAT attempt. The top row plots the mean standardized PSAT score at the first attempt (baseline); the bottom row plots within-student gains between the first and second attempts (Attempt 2 − Attempt 1). Scores are standardized within fall assessment period. Series are computed for students observed with at least two attempts and aggregated to Census region by cohort year.

Figure A8: Residuals of Prior Year Pollution Days



Notes: The above figure shows the distribution of residuals resulting from regressions on student fixed effects of the number of polluted days (panel A) and number of heavily polluted days (panel B) experienced by students at their school location in the 365 days prior to taking the PSAT. All regressions include fixed effects for each combination of cohort, test date and take number. The figure excludes residuals with magnitude above 100 (panel A). The standard deviation of the full set of residuals is shown in each panel.

Table A1: Impact of Prior Year Pollution on Test Scores, Sensitivity to Geographic Trends

	Census Division (1)	State (2)	County (3)
(A) Overall pollution			
Pollution days	-0.038***	-0.036***	-0.020*
	(0.013)	(0.013)	(0.011)
(B) By pollution intensity			
Moderate pollution days	-0.078***	-0.065***	-0.056***
	(0.018)	(0.017)	(0.014)
Heavy pollution days	-0.007	-0.014	0.008
	(0.017)	(0.016)	(0.015)
N	26,127,773	26,127,773	26,127,773

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measures shown. Pollution is measured as days in designated bin ranges in the 365 days preceding the PSAT take. In panel A, polluted days are those with PM2.5 of at least 8 μ/m^3. Panel B distinguishes between moderately polluted days (PM2.5 of 8-12 μ/m^3) and heavily polluted days (PM2.5 $>$ 12 μ/m^3). All regressions include student fixed effects and fixed effects for each combination of cohort, test date and take number, as well as controls for prior year temperature and precipitation, test day pollution levels, test day temperature and precipitation. Column 2 adds controls for Census division-specific linear time trends. Column 3 adds controls for state-specific linear time trends. Column 4 adds controls for county-specific linear time trends. All geographies used for linear time trends reflect the location of the school the student last attended. The sample comprises all students from the high school classes of 2001-22 who took the PSAT more than once and had available pollution data.

Table A2: OLS - Impact of School Year and Summer Pollution

	(1)	(2)	(3)
(A) Overall pollution			
School year pollution days	-0.026		-0.025
	(0.020)		(0.021)
Summer pollution days		0.005	-0.017
		(0.039)	(0.040)
(B) By pollution intensity			
School year moderate pollution days	-0.107***		-0.115***
	(0.028)		(0.028)
School year heavy pollution days	0.043		0.056**
	(0.027)		(0.027)
Summer moderate pollution days		0.030	0.049
		(0.049)	(0.050)
Summer heavy pollution days		-0.014	-0.051
		(0.049)	(0.051)
N	16,742,701	16,742,701	16,742,701

Notes: Heteroskedasticity robust standard errors clustered by student high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of thousandths of a standard deviation in PSAT (math plus reading) z-scores on the pollution measures shown. Pollution is measured as days in the year preceding the PSAT take, separating the calendar into days during the school year and days during the summer break. In panel A, polluted days are those with $PM_{2.5}$ of at least 8 $\mu g/m^3$. Panel B distinguishes between moderately polluted days ($PM_{2.5}$ of 8-12 $\mu g/m^3$) and heavily polluted days ($PM_{2.5} > 12$ $\mu g/m^3$). All regressions include student fixed effects and fixed effects for each combination of cohort, test date and take number, as well as controls for prior year temperature and precipitation, test day pollution levels, test day temperature and precipitation, and Census division-specific linear time trends. The sample comprises all students from the high school classes of 2007-16 who took the PSAT more than once and had available pollution and smoke data.

A2 Data

We utilize a comprehensive dataset of standardized student test scores obtained from the College Board, focusing specifically on PSAT re-takers. We combine these detailed, student-level administrative records with high-resolution measures of local air pollution, wildfire smoke exposure, and weather. Below we describe the PSAT exam data, student address and school information, demographics, school geocoding, and the resulting analysis samples used in our OLS and IV specifications.

This dataset includes information on all students who have taken the PSAT. We focus specifically on the subset of XX million students who take this exam at least twice, allowing us to observe within-student changes over time. In addition, we have detailed geographic information on both the students' residences and the schools they attend, as well as the date of each exam sitting. This enables us to control for air quality on the specific day of testing, and to incorporate local (state-level?) academic calendars to measure the impact of air pollution during the school year relative to non-school periods such as summer months.

Table A3: Other papers that examine learning impacts of $PM_{2.5}$ exposure

Citation	Empirical approach	Results	N
Gilraine and Zheng (2022)	IV with Coal plant closures & Shift share	"We find that each one-unit increase in particulate pollution reduces test scores by 0.02 standard deviations."	700K
Martin et al. (2025)	Downwind of BRT lines (IV)	"$1\mu g/m^3$ additional $PM_{2.5}$ throughout the academic year decreases math scores by 0.0040 standard deviations, this is equivalent to 6% of the mean"	300K
Dang et al. (2024)	Wind direction IV by month	"However, air pollution concentration during the past 12 months does not have significant impacts on the test scores." But they do find adverse impacts of pollution in the 1 month and 3 months prior. They use a linear function with average $PM_{2.5}$.	40K
Wen and Burke (2022b)	Cumulative $PM_{2.5}$ attributable to smoke, no IV	"An additional $10\mu g/m^3$ of cumulative smoke $PM_{2.5}$ on school days decreases average test scores by 0.044% of a standard deviation."	300K

A2.1 Student Test Score Data

PSAT Exam Histories and Outcome Construction

We construct a unified record of PSAT/NMSQT performance for the high-school classes of 2001–2022 by appending cohort-specific exam-history files extracted from the College Board's Cornerstone administrative database. Each PSAT/NMSQT exam administered in the fall is treated as a distinct attempt. For students who tested under the pre-2015 three-section PSAT, we observe section scores in Reading, Math, and Writing, as well as the official composite score. For students who tested under the redesigned PSAT (2015 onward), we observe total scores and section scores for Evidence-Based Reading and Writing (EBRW) and Math.

Scores of zero are treated as invalid, and we discard or correct attempts with inconsistencies between total and section scores. To place scores from different test regimes on a comparable scale, we construct a pseudo two-section score for the three-section period by averaging Reading and Writing and adding Math. We then standardize these composite scores separately within each fall PSAT/NMSQT administration (e.g., Fall 2010). Our primary outcome is the resulting within-administration z-score; EBRW and Math section scores are standardized analogously.

Exam dates are taken from recorded assessment events when available. For attempts with missing dates, we assign the canonical PSAT/NMSQT test date for that cohort and administration year, following the calendar used in Park et al. (2020). A flag identifies imputed dates. When multiple scores appear for the same student during a given fall administration, we retain only the earliest valid attempt.

Sample Restrictions and Retakers

We restrict attention to fall PSAT/NMSQT administrations taken in grades 9–11. We exclude attempts taken in grade 8 or grade 12 and drop exams administered in Fall 2020 and Fall 2021 due to COVID-related disruptions. After these exclusions, we count the number of valid fall PSAT/NMSQT attempts per student and define *retakers* as students with at least two such attempts. Approximately 18 million students meet this criterion, contributing roughly 39 million valid exam observations.

For our main specifications, we further restrict attention to the two most common PSAT/NMSQT testing dates in each exam year. For each fall administration, we identify the two dates with the largest numbers of test takers and drop exams taken on rarer dates. After recomputing attempt numbers, we omit students who become effectively one-time takers under this restriction. This produces a retaker sample in which each student has at least two PSAT/NMSQT scores drawn from common exam dates.

A2.2 Student Address and School Information

Home Address

We extract students' final mailing addresses from Cornerstone, using operational snapshot tables for recent cohorts (2018–2022) and archived closed-cohort tables for earlier cohorts (2001–2017). Address fields—including street, city, state, ZIP5, and ZIP+4—are standardized across cohorts to harmonize legacy and current formats. We retain only students whose final ZIP codes correspond to the 50 U.S. states or the District of Columbia.

High School of Record

We assign each student a high school using College Board AI codes, which we crosswalk to stable organization identifiers (org_ids) maintained in College Board's school directory. We exclude records associated with non–K–12 institutions, online-only or correspondence schools, and schools with uncleanable or non-U.S. addresses. To ensure consistency between home and school locations, we restrict the sample to students whose home and school fall within the same U.S. state and whose ZIP codes are valid domestic ZIP codes. This affects a small share of observations where address information is incomplete or inconsistent.

High school identifiers, once linked to org_ids, enable us to merge school-level geocodes and school characteristics and to assign environmental exposures at the school location.

A2.3 School Geocoding

We construct a master list of high schools by combining all org_ids linked to PSAT records in our sample. For each org_id, we extract physical addresses and any available geocodes from Cornerstone's organization-profile database. These records include street address, ZIP code components, internal latitude–longitude coordinates, and county or census geographic codes when available. All fields are standardized and validated.

To fill missing or low-quality geocodes, we submit standardized school addresses to Melissa Geodata, a commercial rooftop-level address verification and geocoding service. Melissa returns validated postal addresses, rooftop latitude–longitude coordinates, and 11-digit census tract identifiers based on both 2010 and 2020 Census vintages. We merge these results back to the school master file, prioritizing Melissa's coordinates when available and using internal College Board coordinates or ZIP-centroid locations otherwise.

Finally, each school is assigned a county FIPS code using the combination of organization-profile county identifiers, census tract codes, and ZIP-to-county crosswalks. These geographic assignments support (i) construction of pollution exposures using a 2-km radius around the school location (or school ZIP code when lat/long are missing) and (ii) county-level descriptive mapping and county-level controls in our empirical specifications.

A2.4 Student Demographics

We extract demographic variables—including race/ethnicity, gender, and parental education—from Cornerstone. For earlier cohorts (2001–2016), race/ethnicity is drawn from archived cohort tables; for later cohorts (2017–2022), we supplement snapshot files with demographic fields from operational student tables to recover race information. Because Cornerstone transitioned from separate mother/father education variables to "parent 1" and "parent 2" indicators around 2015, we harmonize parental education across all cohorts into a single measure capturing each student's highest parental education. Gender is standardized into Male, Female, Another Gender, and Unknown. These demographic variables are merged with the exam-history and school-location files using the cohort-consistent student identifier.

A2.5 Treatment Variables

A2.5.1 Air Pollution Data (PM2.5)

Our air pollution measures use PM2.5 levels predicted using a machine learning approach described Di et al. (2019). Their ML approach combines observed PM2.5 from ground-station data, satellite imagery, chemical transport models, and weather reanalysis to produce daily predicted PM2.5 levels over the contiguous United States from 2000 to 2017. The predictions provide estimated maximum PM2.5 concnetrations at a 1km x 1km resolution, daily over this time period.[2]

We aggregate these gridded measures of PM2.5 to schools in two ways. First, we define a buffer around each school in our data using their latitude and longitude coordinates of 10km. We then average the PM levels at all grid-cells that fall within the buffer and assign that value to the school for each day in the sample. Second, for any school that we do not have latitude and longitude coordinates for, we determine the school zip code and average the PM values for all grid-cells that fall within the spatial extent of that zip code. Schools without latitude and longitude values are then assigned the PM associated with their zip code.

This produces a dataset of PM levels by day for every school in our sample for every day from 2000 to 2017. Students are assigned exposure to pollution based on the total number of days that pollution falls into each of several sets of bins (in our primary specification the bins are 0-8 $\mu g/m^3$, 8-12, and 12+) during the 365 days prior to the most likely day that they took the PSAT.[3] In some analyses we further divide days into school and non-school days based on the typical days in which schools were in session in the state in which the student took the PSAT.[4]

A2.6 Instruments

A2.6.1 Wildfire Smoke Data

In addition to standard pollution measures, we incorporate data on the contribution of wildfire smoke to daily PM2.5 levels from Childs et al. (2022). These data are constructed by training a machine learning process that incorporates data from ground stations, satellite imagery, aerosol packet transport models (e.g. HYSPLIT), and weather reanalysis products. The trained ML model is then used to predict the level of PM2.5 (in $\mu g/m^3$) that is attributable wildfire smoke within 10km x 10km grid-cells covering the entire contiguous United States.[5]

We use data provided by the authors of the original study that aggregates the gridded smoke estimates to U.S. ZCTAs and provides the predicted level of PM2.5 on each day of the year (for years spanning 2006-2020) in that ZCTA. The predictions distinguish between days with “no smoke" and therefore no PM from smoke and days with smoke but no predicted PM from smoke. We group

[2]These data have been used extensively to examine disparities in exposure to pollution across the U.S. (Jbaily et al., 2022), the relationship between pollution and dementia (Shi et al., 2020), and the adverse effects of pollution on respiratory disease in the U.S. (Feng et al., 2024), among many other outcomes.

[3]We do not know exactly when a given student took the PSAT but we know the dates on which the PSAT was administered in a given year as well as the share of students who took the examine on each day. We calculate exposures based on the two days in which the largest share of students took the PSAT in a given year.

[4]For example, if schools in Massachusetts typically start on September 1st and end on June 14th, we consider days in between these dates to be during the school year for all schools in Massachusetts and days outside these dates to be summer days. This is the same approach to school vs. summer days that was used in Park et al. (2020).

[5]These data have already been used extensively in studies in both the economics and general science literature documenting the adverse impacts of wildfire smoke on mortality (Gould et al., 2024; Connolly et al., 2024), morbidity (Zhang et al., 2025), and the specific contribution of wildfires to PM2.5 trends in the U.S. (Burke et al., 2023).

these days together. We then crosswalk ZCTAs to zipcodes and assign schools smoke PM levels based on the zipcode in which they are located.

A2.7 Additional Identifiers and Controls

In all specifications, we include a fixed effect for the combination of expected cohort year, exam date, and attempt number—a "CDT group"—constructed by grouping students by graduation cohort, the PSAT administration date, and whether the score corresponds to their first, second, or (rarely) third attempt. We also attach county-level unemployment rates using each school's county and the exam year, using unemployment in the exam year and, where missing, in the adjacent years. These controls, along with weather and exam-day atmospheric measures described above, are included in relevant OLS and IV specifications.

A2.8 Heterogeneity Measures

We use two primary measures to examine heterogeneity of effects, a measure quantifying the level of challenge a student from a given high school experiences in accessing higher education, and students' scores on their first exam attempt relative to mean scores in specific geographic units (i.e., national, state).

High School Challenge Index

We measure school-level challenge using the College Board Landscape index, a nationally normed percentile (1–100) that summarizes the educational and socioeconomic conditions in the communities served by each high school. Landscape is constructed at the census-tract level using all College Board assessment takers—SAT, PSAT, and AP—across five consecutive graduating cohorts. For each edition (e.g., Landscape 2024), census tracts are ranked nationally based on a composite of neighborhood characteristics and weighted by the number of College Board test takers residing in each tract.

We merge each high school in our data to its corresponding Landscape percentile based on the school's geocoded location and the last available Landscape edition for that school. For the heterogeneity analysis in Section 4, we construct terciles of this index *within our retaker sample*, so that the low-, middle-, and high-challenge groups reflect the distribution of schools actually attended by students in the analysis sample rather than the national distribution of all high schools. Throughout, we refer to these terciles as low-, medium-, and high-challenge schools.

Baseline Achievement Groups

For our heterogeneity-by-achievement analysis, we classify students according to their PSAT performance at their first observed attempt. Let z_{i1} denote student i's standardized PSAT (math plus reading) score on their first PSAT/NMSQT attempt, where scores are standardized within fall exam period as described above. We construct an indicator for "low baseline achievement" equal to one if z_{i1} falls below the mean standardized score among all test takers in the nation in that exam year, and zero otherwise. We then carry this indicator forward to all subsequent attempts for the same student, so that baseline achievement is defined once, prior to the pollution histories that differ across attempts.

Although we also construct analogous indicators based on exam-year means at the state and high school levels (using separate lookup files for state-by-year and school-by-year mean scores), our

main heterogeneity results by initial achievement use the national mean as the reference. In Table 9, we therefore compare students whose first PSAT attempt placed them at or above versus below the national mean in their exam year.

A2.9 Analysis Samples

Baseline OLS Sample

Our baseline OLS sample ("full retaker sample") includes all retakers with complete PSAT score, address, school, and demographic information, and who tested on one of the two most common PSAT/NMSQT dates in their testing year. This sample incorporates pollution exposure measures only from years and locations where school-level and ZIP-level pollution data are available.

IV-Eligible Sample

For our instrumental-variables analyses, we further restrict the retaker sample to observations with complete exposure to both total particulate pollution and smoke-attributable PM2.5 from the ECHO dataset. Because ECHO smoke data are available beginning in 2007, we limit the IV sample to exam years 2007–2016. Following best practice for plausibly exogenous smoke-driven variation, we exclude school-year observations whose schools lie within 10 km of a wildfire boundary. The resulting IV-eligible sample is therefore defined by the intersection of (i) retakers with two or more valid attempts, (ii) availability of PM2.5 and smoke exposure measures, (iii) the two-most-common exam-date restriction, and (iv) school locations more than 10 km from any fire.

A3 Method for Interpreting Magnitudes

A3.1 Student-Level Pollution Exposure and Scaling

This section constructs the student-level pollution exposure measures used throughout the magnitude calculations. We characterize both the level and dispersion of exposure across students in the analytic sample.

A3.1.1 Average Exposure Across Students

We begin by restricting the data to observations included in the main estimation sample. The unit of observation in the underlying dataset is a student-by-exam record, so students may appear multiple times if they take the PSAT more than once. To obtain a student-level exposure measure, we collapse the data to one observation per student.

For each student, we compute the mean number of days with pollution in each exposure category across all of their observed exam years. Specifically, we calculate the average number of days per year in which particulate matter concentrations fall within the moderate range (PM2.5 between 8 and 12 $\mu g/m^3$), the heavy range (PM2.5 above 12 $\mu g/m^3$), and the combined set of polluted days (PM2.5 at or above 8 $\mu g/m^3$). This produces a student-level average exposure profile that reflects the pollution conditions experienced over the period in which the student is observed.

After constructing these student-level averages, we summarize the distribution of exposure across students. We compute the mean of these student-level averages, which gives the expected number of polluted days experienced by the average student, with each student weighted equally. This differs from aggregating at the observation level, where students with more test attempts would

receive greater weight. In addition, we compute the median of the student-level averages to provide a measure of the central tendency that is less sensitive to outliers.

These summary statistics serve as inputs for subsequent magnitude calculations, where estimated pollution effects are scaled by the typical exposure experienced by students. By working at the student level and weighting each student equally, this approach ensures that magnitude interpretations correspond to the experience of the average student in the analytic sample.

A3.1.2 Dispersion in Exposure and One-Standard-Deviation Scaling

To complement the characterization of average exposure, we quantify the dispersion in pollution exposure across students. Specifically, we compute the standard deviation of students' mean annual number of moderate-pollution days (PM2.5 between 8 and 12 $\mu g/m^3$), where each student's exposure is first averaged across all observed exam years, as described in the previous subsection.

This statistic captures the extent of variation in moderate-pollution exposure across students in the analytic sample. Because it is constructed from student-level averages, it reflects differences in typical exposure levels across students rather than variation across individual exam observations.

We use this dispersion measure to construct one-standard-deviation changes in exposure, which provide a convenient benchmark for interpreting the magnitude of pollution effects. In subsequent calculations, estimated per-day pollution effects are scaled by this standard deviation to express the impact of a typical increase in moderate-pollution exposure. These scaled effects are then translated into alternative units, including standardized test-score units and measures of learning over time.

A3.2 Average-Student Effects of Pollution Exposure

This section computes the effect of pollution exposure for a representative student by scaling the estimated per-day pollution effects using the exposure profile of the average student. These quantities serve as the baseline magnitudes used in subsequent interpretation and counterfactual analyses.

A3.2.1 Total Achievement Cost for the Average Student

We first translate the estimated pollution effects into changes in test scores for a representative student. The goal is to quantify the total achievement cost associated with the level of pollution exposure experienced by the average student in the analytic sample.

To do so, we combine the estimated per-day effect of pollution on PSAT scores with the average number of polluted days experienced by students, as described in Section 1. This yields a model-implied reduction in test scores relative to a benchmark in which no polluted days occur. Throughout, we interpret this quantity as the total effect of exposure accumulated over the course of a year, rather than the effect of a marginal change in exposure.

We report estimates of this quantity under alternative specifications that differ in how pollution exposure is defined.

Baseline Scaling Using Student-Level Exposure We begin by estimating the total achievement cost of pollution for the average student using the binned exposure specification. This calculation combines the estimated per-day effects of moderate- and heavy-pollution days with the average number of days in each category experienced by students, as constructed in Section 1.

Specifically, let $\beta_{8\text{–}12}$ and $\beta_{\geq 12}$ denote the estimated effects of one additional day of moderate and heavy pollution, respectively. Let $\mathbb{E}[D_{8\text{–}12}]$ and $\mathbb{E}[D_{\geq 12}]$ denote the average number of days in each category across students, where each student is weighted equally. The total achievement cost for the average student is then given by:

$$\Delta_{\text{avg}} = \beta_{8\text{–}12}\,\mathbb{E}[D_{8\text{–}12}] + \beta_{\geq 12}\,\mathbb{E}[D_{\geq 12}].$$

We estimate this quantity using a nonlinear combination of the regression coefficients and the average exposure levels, which allows us to recover both the point estimate and its associated standard error and confidence interval. This approach accounts for uncertainty in the estimated pollution effects when constructing the scaled magnitude.

The resulting estimate represents the model-implied annual change in PSAT scores for the average student, measured relative to a benchmark in which no moderate- or heavy-pollution days occur. This interpretation relies on the linearity of the estimated relationship between pollution exposure and test scores, so that the total effect of exposure is the sum of the per-day effects across pollution categories.

Alternative Specification Using an 8+ Exposure Measure

As an alternative to the binned specification, we construct an average-student magnitude using a specification that aggregates moderate- and heavy-pollution days into a single measure of polluted days. In this specification, exposure is defined as the total number of days with PM2.5 at or above 8 μg/m^3.

Let β_{8+} denote the estimated effect of one additional polluted day under this specification, and let $\mathbb{E}[D_{8+}]$ denote the average number of polluted days experienced by students, as constructed in Section 1. The total achievement cost for the average student is then given by:

$$\Delta^{8+}_{\text{avg}} = \beta_{8+}\,\mathbb{E}[D_{8+}].$$

We estimate this quantity using a nonlinear combination of the regression coefficient and the average exposure level, which yields both the point estimate and its associated standard error and confidence interval. As in the baseline case, this approach accounts for uncertainty in the estimated per-day effect when constructing the scaled magnitude.

This specification provides a more parsimonious representation of pollution exposure by collapsing moderate- and heavy-pollution days into a single count. The resulting estimate can be interpreted as the model-implied annual change in PSAT scores for the average student relative to a benchmark in which no polluted days occur.

A3.3 Interpreting Magnitudes in Intuitive Units

This section translates the estimated pollution effects into more interpretable units to facilitate economic interpretation. Building on the average-student effects constructed in Section 2, we express these magnitudes in standardized test-score units, approximate percentile shifts, and time-based measures of learning.

These transformations provide complementary ways of understanding the size of the estimated effects relative to the distribution of scores and typical patterns of learning, without altering the underlying empirical estimates.

A3.3.1 Standardized Units and Distributional Interpretation

We first interpret the estimated effects in terms of standardized test-score units and their implied position in the distribution of scores. These measures provide a distributional benchmark for assessing the magnitude of the estimated effects.

Effects in Standard Deviation Units

We first express the estimated effects in standardized test-score units. Because the PSAT outcome is scaled such that 1000 points correspond to one standard deviation, we convert point-scale effects into standard deviation units by dividing the estimated change in scores by 1000.

Let Δ denote the average-student scaled effect in points, as defined in Section 2. The corresponding effect in standard deviation units is given by:

$$\Delta_{\text{SD}} = \frac{\Delta}{1000}.$$

Confidence intervals are transformed analogously by applying the same scaling to the lower and upper bounds.

We apply this transformation both to the baseline estimate based on the binned specification and to the alternative estimate based on the 8+ specification. This provides a standardized measure of effect size that is directly comparable across specifications and facilitates interpretation relative to the distribution of test scores.

Approximate Percentile Shifts Near the Median

We next translate the estimated effects into approximate changes in percentile rank. To do so, we use a normal approximation to the distribution of standardized test scores and interpret the estimated effect as a shift in the mean of a standard normal distribution.

Let Δ_{SD} denote the effect in standard deviation units, as defined in the previous subsection. Under the assumption that test scores are approximately normally distributed, the implied change in percentile rank for a student near the median is given by:

$$\Delta_{\text{pct}} = [\Phi(\Delta_{\text{SD}}) - 0.5] \times 100,$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

Confidence intervals for the percentile shift are constructed by applying the same transformation to the lower and upper bounds of the effect in standard deviation units. This yields an approximate mapping from the estimated effect size to a change in position within the score distribution.

We apply this transformation both to the baseline estimate and to the alternative estimate based on the 8+ specification. The resulting quantities provide an interpretable measure of effect size in terms of movement within the distribution of test scores, particularly in the neighborhood of the median where the normal approximation is most accurate.

A3.3.2 Time-Based Interpretations of Learning

We next translate the estimated effects into time-based measures of learning. While standardized units and percentile shifts describe the magnitude of effects relative to the distribution of scores,

these measures can be difficult to interpret in terms of academic progress. To provide a more intuitive benchmark, we convert score changes into equivalent units of learning over time.

These conversions are based on observed patterns of score growth within students, which provide an empirical measure of typical learning over a given period. By expressing pollution-induced score changes relative to these benchmarks, we obtain interpretable measures such as months of learning and the share of annual learning gains associated with pollution exposure.

Months of Learning

To express the estimated effects in time-based units, we convert score changes into equivalent months of learning using an empirically estimated learning rate. This learning rate is derived from within-student score growth between consecutive PSAT attempts in the analytic sample.

Specifically, for each student with multiple test attempts, we compute the change in test scores between adjacent exams and divide by the number of days between those exams. This yields a student-level measure of score growth per day. We then annualize this quantity by scaling it to a 365-day period and average across students to obtain the mean annual gain in test scores.

Let G_{ann} denote the resulting average annual gain in test scores, measured in standardized score units (where 1000 points correspond to one standard deviation). The implied rate of learning per month is given by:

$$G_{\text{month}} = \frac{G_{\text{ann}}}{12}.$$

We convert the estimated effect from Section 2 into months of learning by dividing the change in test scores, measured in points, by this monthly learning rate. Confidence intervals are obtained by applying the same transformation to the lower and upper bounds of the estimated effect.

This approach yields an interpretable measure of the magnitude of pollution effects in terms of typical learning progression, based on observed within-student score growth in the data.

Share of Annual Learning Gains

We next express the estimated effects as a share of average annual learning gains. This normalization uses the same empirical learning-rate benchmark constructed in the previous subsection, namely the average within-student score gain annualized to a one-year period.

Let G_{ann} denote the mean annual gain in test scores, measured in standardized-score points. For any estimated score effect Δ, the corresponding share of annual learning is defined as:

$$\text{Share of annual learning} = 100 \times \frac{\Delta}{G_{\text{ann}}}.$$

We apply this transformation to the average-student scaled effect from Section 2, which yields a measure of the extent to which the estimated pollution effect offsets typical yearly score growth. This provides an interpretable benchmark for assessing the magnitude of the estimated effect relative to the amount of learning that students typically accumulate over the course of a year.

We also apply the same normalization to selected counterfactual exposure scenarios considered later in the appendix. This allows those hypothetical score changes to be interpreted on the same time-based scale as the average-student effect. In each case, the calculation uses the same annual

learning benchmark, so differences across exercises reflect variation in the underlying score effect rather than changes in the normalization itself.

Effects of a One-Standard-Deviation Increase in Exposure

As an additional benchmark, we scale the estimated effect of moderate-pollution exposure by a one-standard-deviation increase in students' typical exposure. This calculation uses the dispersion measure constructed in Section 1, namely the standard deviation of students' mean annual number of moderate-pollution days.

Let $\sigma_{D_{8\text{–}12}}$ denote the standard deviation of student-level mean exposure to moderate-pollution days, and let $\beta_{8\text{–}12}$ denote the estimated per-day effect of one additional moderate-pollution day. The effect of a one-standard-deviation increase in exposure is then given by:

$$\Delta_{\text{1SD}} = \sigma_{D_{8\text{–}12}} \cdot \beta_{8\text{–}12}.$$

We estimate this quantity using a nonlinear combination of the standard-deviation exposure measure and the estimated regression coefficient, which yields both the point estimate and its confidence interval. We then translate this effect into several alternative units. First, we express it in standardized test-score units by dividing by 1000, since the outcome is scaled so that 1000 points correspond to one standard deviation. Second, we express the same quantity as a percent of a standard deviation by multiplying the standardized effect by 100. Finally, when the annual learning benchmark from the previous subsection is available, we normalize the one-standard-deviation exposure effect by average annual score growth to express it as a share of annual learning gains.

This calculation provides an interpretable benchmark for the magnitude of the estimated pollution effect under a typical increase in exposure, measured relative to the observed dispersion in moderate-pollution days across students.

A3.3.3 Comparison to Observed Score Changes

As an additional benchmark, we compare the magnitude of the estimated pollution effects to observed within-student changes in PSAT scores across test attempts. This provides a reference point based on actual score variation in the data, rather than a constructed transformation.

To do so, we compute within-student changes in test scores between consecutive PSAT attempts. For each student, we calculate the difference in scores between the first and second attempts, as well as additional differences for students with more than two attempts. We then summarize the distribution of these changes across students, including the mean, median, standard deviation, and the mean absolute change in scores.

Using these quantities, we construct two comparison metrics. First, we compute the ratio of the estimated effect from Section 2 to the mean within-student score change between the first and second attempts. This yields a measure of the extent to which the estimated pollution effect corresponds to typical observed changes in scores. Second, to abstract from the direction of score changes, we compare the absolute value of the estimated effect to the mean absolute within-student change. This provides a magnitude-based benchmark that is not affected by the sign of average score changes.

These comparisons situate the estimated pollution effect relative to the distribution of actual score movements observed in the data, offering an empirical benchmark for interpreting the size of the effect.

A3.3.4 Student-Level Counterfactual Outcomes

Counterfactual Score Levels at the Second Attempt

We next construct student-level counterfactual outcomes to assess how pollution exposure affects the distribution of test scores. Specifically, we estimate how much higher each student's PSAT score at their second attempt would have been under a no-pollution benchmark.

To do so, we use the estimated pollution coefficients from the baseline specification and compute, for each observation, the model-implied contribution of pollution exposure to the observed score. For attempt-2 observations, this pollution component is given by the product of the estimated coefficients and the number of moderate- and heavy-pollution days in the year prior to the exam.

We then construct a counterfactual score by subtracting this pollution component from the observed score. The resulting difference between the counterfactual and observed scores represents the implied increase in performance in the absence of polluted days. This yields a distribution of student-level counterfactual differences, rather than a single average effect.

We summarize this distribution using standard descriptive statistics, including the mean, median, and selected percentiles. These summaries characterize the heterogeneity in the model-implied impact of pollution across students.

This exercise should be interpreted as a partial-equilibrium counterfactual that holds all other covariates and fixed effects at their observed values. The resulting quantities reflect model-implied level differences at the second attempt and should not be interpreted as changes in learning or growth between attempts.

Counterfactual Score Growth Between Attempts

We next consider a growth-based counterfactual that asks how much larger score improvement between the first and second PSAT attempts would have been under a no-pollution benchmark. Unlike the previous subsection, which focuses on counterfactual score levels at the second attempt, this exercise focuses on the change in scores across attempts.

To construct this object, we first compute counterfactual scores under no pollution for all observed attempts. For each observation, we subtract the model-implied pollution component from the observed score, where the pollution component is defined as the sum of the estimated effects of moderate- and heavy-pollution days in the year prior to that exam. This yields a no-pollution counterfactual score for each attempt.

We then compute, for each student with both a first and second attempt, the difference between the counterfactual second-attempt score and the counterfactual first-attempt score. This gives the counterfactual amount of score growth between attempts under a no-pollution benchmark. We compare this quantity to the observed score change between the second and first attempts, and define the difference between the two as the model-implied increase in score growth attributable to the removal of pollution exposure.

This procedure yields a distribution of student-level counterfactual changes in growth, which we summarize using standard descriptive statistics such as the mean, median, and interquartile range. We also translate the average effect into standardized-score units and, when the learning-rate benchmark from the previous subsection is available, into months of learning.

As in the preceding counterfactual exercise, this should be interpreted as a partial-equilibrium,

model-implied comparison that holds all other covariates and fixed effects at their observed values. Because the no-pollution benchmark is applied to both attempts, the resulting object captures the change in score growth that would arise from removing pollution exposure prior to each exam, rather than the effect of changing exposure at only one attempt.

A3.4 Counterfactual Changes in Pollution Exposure

We next consider how student achievement would respond to changes in pollution exposure under a set of counterfactual scenarios. In contrast to the average-student magnitudes in Section 2, which scale estimated effects using observed exposure levels, the exercises in this section impose hypothetical changes in the number of polluted days and translate those changes into predicted score differences using the estimated per-day effects.

We consider two types of counterfactual changes. First, we examine large changes in exposure that are best interpreted as arising from higher-level institutional or environmental shifts, such as regional or policy-driven changes in air quality. Second, we consider smaller reductions in exposure that are more consistent with school- or district-level mitigation strategies. In each case, we use the estimated pollution effects to map changes in exposure into predicted changes in test scores, holding all other factors fixed.

A3.4.1 Higher-Level Institutional Changes in Moderate-Pollution Exposure

A 100-Day Moderate-Pollution Scenario

We first consider a counterfactual scenario in which a school experiences 100 moderate-pollution days in a given year. This scenario is intended to capture the implications of a large change in exposure that is best interpreted as arising from higher-level institutional or environmental conditions, rather than from local mitigation.

To construct this quantity, we use the estimated per-day effect of moderate-pollution exposure from the baseline specification. Let $\beta_{8\text{–}12}$ denote the estimated effect of one additional day with PM2.5 between 8 and 12 μg/m^3. Under the maintained assumption of linearity in days, the effect of a year with 100 moderate-pollution days is given by:

$$\Delta_{100} = 100 \cdot \beta_{8\text{–}12}.$$

In this scenario, exposure to heavy-pollution days is set to zero, so the total effect reflects only moderate-pollution exposure. We estimate this quantity using a nonlinear combination of the regression coefficient, which yields both the point estimate and its associated standard error and confidence interval.

We then express the resulting effect in multiple units. First, we convert the point-scale effect into standardized test-score units by dividing by 1000, since the outcome is scaled such that 1000 points correspond to one standard deviation. Second, we express the same quantity as a percentage of a standard deviation by multiplying the standardized effect by 100.

This scenario provides an interpretable benchmark for the magnitude of the estimated pollution effect under a large change in exposure, illustrating how sustained exposure to moderate levels of pollution translates into changes in student achievement.

A 30-Day Moderate-Pollution Scenario

We next consider a counterfactual scenario in which a school experiences 30 moderate-pollution days in a given year. This scenario represents a smaller, though still substantial, change in exposure relative to the 100-day case and provides an additional benchmark for interpreting the magnitude of the estimated effects.

As in the previous subsection, we use the estimated per-day effect of moderate-pollution exposure from the baseline specification. Under the assumption of linearity in days, the effect of a year with 30 moderate-pollution days is given by:

$$\Delta_{30} = 30 \cdot \beta_{8\text{–}12}.$$

In this scenario, exposure to heavy-pollution days is set to zero, so the resulting effect reflects only moderate-pollution exposure. We estimate this quantity using a nonlinear combination of the regression coefficient, which yields the corresponding confidence interval.

We express the resulting effect in standardized units by dividing the point-scale estimate by 1000, and we additionally report the effect as a percentage of a standard deviation. This provides a direct comparison to the 100-day scenario while illustrating how the estimated effect scales with the magnitude of exposure.

A3.4.2 School- or District-Level Reductions in Moderate-Pollution Exposure

We next consider counterfactual reductions in moderate-pollution exposure that are more consistent with changes that could be implemented at the school or district level. In contrast to the scenarios considered above, which fix exposure at a given level, these exercises evaluate the effect of reducing the number of moderately polluted days relative to a school's observed exposure.

To construct these quantities, we again use the estimated per-day effect of moderate-pollution exposure from the baseline specification. Let $\beta_{8\text{–}12}$ denote the estimated effect of one additional moderate-pollution day. Under the assumption of linearity in days, the effect of a reduction of k moderate-pollution days is given by:

$$\Delta_{-k} = -k \cdot \beta_{8\text{–}12},$$

where $k \in \{5, 10, 15\}$.

We estimate these effects using nonlinear combinations of the regression coefficient, which yields point estimates and confidence intervals for each scenario. Because the specification is linear in days, the implied change in scores scales proportionally with the size of the reduction in exposure.

These counterfactuals represent partial-equilibrium changes that hold all other determinants of achievement fixed. They provide interpretable benchmarks for the potential gains associated with modest reductions in pollution exposure at a given school or district.

A3.5 Decomposition of the Black–White Achievement Gap by Pollution Exposure

We next assess the extent to which differences in pollution exposure contribute to the observed Black–White achievement gap. Specifically, we combine the estimated per-day effect of pollution with differences in exposure between Black and White students to construct a mechanical decomposition of the gap.

This exercise isolates the portion of the observed gap associated with differential exposure to polluted days, holding the estimated pollution–score relationship fixed. It should be interpreted as an accounting exercise rather than a causal decomposition of broader disparities, as it does not account for other determinants of achievement that may differ across groups.

A3.5.1 Using First-Attempt Scores to Define the Black–White Gap

Moderate-Pollution Exposure

We first define the observed Black–White achievement gap using students' PSAT scores at the first attempt. To do so, we construct a student-level dataset containing each student's mean pollution exposure across observed exam years together with the student's race, and then merge in the student's score from the first PSAT attempt. This allows us to compare group differences in both achievement and typical exposure.

We then compute group means separately for Black and White students. In particular, we calculate the mean first-attempt PSAT score for each group and the mean number of moderate-pollution days experienced by each group, where moderate-pollution days are defined as days with PM2.5 between 8 and 12 μg/m^3.

Let $\overline{D}^{B}_{8\text{–}12}$ and $\overline{D}^{W}_{8\text{–}12}$ denote the mean number of moderate-pollution days for Black and White students, respectively, and let $\beta_{8\text{–}12}$ denote the estimated per-day effect of one additional moderate-pollution day. The contribution of differential moderate-pollution exposure to the first-attempt Black–White score gap is then given by:

$$\text{Contribution}^{8\text{–}12}_{BW} = \beta_{8\text{–}12}\left(\overline{D}^{B}_{8\text{–}12} - \overline{D}^{W}_{8\text{–}12}\right).$$

To express this quantity relative to the observed gap, we divide the implied score contribution by the mean Black–White difference in first-attempt PSAT scores and multiply by 100. This yields the share of the observed first-attempt gap that is mechanically associated with differences in moderate-pollution exposure.

We also construct a related counterfactual quantity by subtracting the implied pollution contribution from the observed Black–White gap. This yields the Black–White score gap that would remain under a benchmark in which Black and White students experienced the same number of moderate-pollution days. Because this calculation uses only the moderate-pollution coefficient, it should be interpreted as an equalized-exposure counterfactual for moderate-pollution days rather than as a full no-pollution benchmark.

As in the rest of this section, this exercise is an accounting decomposition that holds the estimated pollution–score relationship fixed. It should not be interpreted as a causal decomposition of broader racial disparities in achievement.

All Polluted Days (PM2.5 $\geq$ 8)

We next repeat the Black–White gap decomposition using the 8+ pollution specification, which aggregates moderate- and heavy-pollution days into a single measure of polluted days. This provides a decomposition based on the preferred all-polluted-days exposure measure and corresponds to a full no-pollution benchmark.

As in the previous subsection, we define the observed Black–White achievement gap using students'

PSAT scores at the first attempt. We then compute group means for Black and White students, including the mean first-attempt PSAT score and the mean number of polluted days with PM2.5 at or above 8 μg/m^3.

Let $\overline{D}^B_{8+}$ and $\overline{D}^W_{8+}$ denote the mean number of polluted days for Black and White students, respectively, and let β_{8+} denote the estimated per-day effect from the 8+ specification. The contribution of differential exposure to polluted days to the first-attempt Black–White score gap is then given by:

$$\text{Contribution}^{8+}_{BW} = \beta_{8+}\left(\overline{D}^B_{8+} - \overline{D}^W_{8+}\right).$$

To express this quantity relative to the observed achievement gap, we divide the implied score contribution by the mean Black–White difference in first-attempt PSAT scores and multiply by 100. This yields the share of the observed first-attempt gap that is mechanically associated with differential exposure to polluted days under the 8+ specification.

We also use this contribution to construct a no-pollution counterfactual gap. Specifically, we subtract the implied pollution contribution from the observed Black–White gap, yielding the gap that would remain under a benchmark in which neither group experienced any polluted days. Because this calculation is based on the 8+ specification, it corresponds to removing all days with PM2.5 at or above 8 μg/m^3, rather than equalizing exposure only within a specific pollution bin.

As in the rest of this section, this is an accounting decomposition that holds the estimated pollution–score relationship fixed. It should not be interpreted as a causal decomposition of broader racial disparities in achievement.

A3.5.2 Using Student-Level Mean Scores to Define the Black–White Gap

Moderate-Pollution Exposure We next construct an alternative decomposition of the Black–White achievement gap using students' average test scores across observed attempts, rather than scores at the first attempt. This provides a measure of persistent differences in achievement that aggregates performance over time.

To do so, we first compute, for each student, the mean PSAT score across all observed attempts, as well as the mean number of moderate-pollution days experienced over those same years. We then compute group means separately for Black and White students, including both the average score and the average level of moderate-pollution exposure.

Let $\overline{Y}^B$ and $\overline{Y}^W$ denote the mean PSAT scores for Black and White students, respectively, and let $\overline{D}^B_{8\text{–}12}$ and $\overline{D}^W_{8\text{–}12}$ denote the corresponding mean number of moderate-pollution days. Let $\beta_{8\text{–}12}$ denote the estimated per-day effect of moderate-pollution exposure. The contribution of differential exposure to the gap in average scores is then given by:

$$\text{Contribution}^{8\text{–}12}_{BW,\text{mean}} = \beta_{8\text{–}12}\left(\overline{D}^B_{8\text{–}12} - \overline{D}^W_{8\text{–}12}\right).$$

To express this quantity relative to the observed gap, we divide the implied score contribution by the difference in mean PSAT scores between Black and White students and multiply by 100. This yields the share of the Black–White gap in average scores that is mechanically associated with differences in moderate-pollution exposure.

As in the previous subsection, this calculation is based solely on the moderate-pollution coefficient

and therefore corresponds to an equalized-exposure counterfactual for moderate-pollution days. It should be interpreted as an accounting decomposition that holds the estimated pollution–score relationship fixed, rather than as a causal decomposition of group differences in achievement.

All Polluted Days (PM2.5 $\geq$ 8)

We next repeat the decomposition using the 8+ pollution specification, which aggregates moderate- and heavy-pollution days into a single measure of polluted days. As in the previous subsection, the observed Black–White achievement gap is defined using students' mean PSAT scores across all observed attempts.

We compute, for Black and White students separately, the mean PSAT score and the mean number of polluted days with PM2.5 at or above 8 μg/m^3. Let $\overline{Y}^B$ and $\overline{Y}^W$ denote the mean PSAT scores for Black and White students, respectively, and let $\overline{D}^B_{8+}$ and $\overline{D}^W_{8+}$ denote the corresponding mean number of polluted days. Let β_{8+} denote the estimated per-day effect from the 8+ specification. The contribution of differential pollution exposure to the gap in average scores is then given by:

$$\text{Contribution}^{8+}_{BW,\text{mean}} = \beta_{8+}\left(\overline{D}^B_{8+} - \overline{D}^W_{8+}\right).$$

To express this quantity relative to the observed gap, we divide the implied score contribution by the Black–White difference in mean PSAT scores and multiply by 100. This yields the share of the Black–White gap in average scores that is mechanically associated with differential exposure to polluted days under the 8+ specification.

Because this calculation is based on the all-polluted-days specification, it corresponds to a full no-pollution benchmark in which all days with PM2.5 at or above 8 μg/m^3 are removed. As in the rest of this section, the result should be interpreted as an accounting decomposition that holds the estimated pollution–score relationship fixed, rather than as a causal decomposition of broader racial disparities in achievement.

A3.6 Decomposition of Achievement Gaps by School Challenge Level

We next examine whether differences in pollution exposure contribute to achievement gaps across schools of differing academic challenge. In this section, we compare students attending high- and low-challenge schools, where challenge is defined using a percentile-based index assigned at the school level.

We construct a mechanical decomposition of the high–low challenge achievement gap by combining the estimated per-day effect of pollution with differences in exposure between students attending these schools. As in the previous section, this exercise isolates the portion of the observed gap associated with differential exposure to polluted days, holding the estimated pollution–score relationship fixed.

Throughout, we interpret these results as accounting decompositions rather than causal estimates of the effect of school characteristics or socioeconomic differences. The calculations describe how much the observed gap would change if differences in pollution exposure between high- and low-challenge schools were eliminated, without modeling the underlying determinants of school assignment.

A3.6.1 Using First-Attempt Scores to Define the High–Low Challenge Gap

Moderate-Pollution Exposure We first define the observed achievement gap using students' PSAT scores at the first attempt, comparing students attending high- and low-challenge schools.

As in the previous section, we construct student-level measures of average pollution exposure and merge these with first-attempt test scores to characterize group differences in both exposure and achievement.

We compute group means separately for students in high- and low-challenge schools, including the mean first-attempt PSAT score and the mean number of moderate-pollution days, where moderate-pollution days are defined as days with PM2.5 between 8 and 12 μg/m^3. Let $\overline{D}^H_{8\text{–}12}$ and $\overline{D}^L_{8\text{–}12}$ denote the mean number of moderate-pollution days for students in high- and low-challenge schools, respectively, and let $\beta_{8\text{–}12}$ denote the estimated per-day effect of moderate-pollution exposure.

The contribution of differential exposure to the high–low challenge gap is then given by:

$$\text{Contribution}^{8\text{–}12}_{HL} = \beta_{8\text{–}12}\left(\overline{D}^H_{8\text{–}12} - \overline{D}^L_{8\text{–}12}\right).$$

To express this quantity relative to the observed gap, we divide the implied score contribution by the difference in mean first-attempt PSAT scores between high- and low-challenge schools and multiply by 100. This yields the share of the observed gap that is mechanically associated with differences in moderate-pollution exposure.

We also construct a corresponding counterfactual gap by subtracting the implied pollution contribution from the observed gap. This yields the achievement gap that would remain under a benchmark in which students in high- and low-challenge schools experienced the same number of moderate-pollution days. Because this calculation uses only the moderate-pollution coefficient, it should be interpreted as an equalized-exposure counterfactual for moderate-pollution days rather than as a full no-pollution benchmark.

As in Section 5, this exercise is an accounting decomposition that holds the estimated pollution–score relationship fixed and should not be interpreted as a causal decomposition of differences across school types.

All Polluted Days (PM2.5 $\geq$ 8)

We next repeat the decomposition using the 8+ pollution specification, which aggregates moderate- and heavy-pollution days into a single measure of polluted days. As in the previous subsection, the observed achievement gap is defined using students' PSAT scores at the first attempt, comparing students in high- and low-challenge schools.

We compute group means for students in high- and low-challenge schools, including both the mean first-attempt PSAT score and the mean number of polluted days with PM2.5 at or above 8 μg/m^3. Let $\overline{D}^H_{8+}$ and $\overline{D}^L_{8+}$ denote the mean number of polluted days for students in high- and low-challenge schools, respectively, and let β_{8+} denote the estimated per-day effect from the 8+ specification. The contribution of differential exposure to the high–low challenge gap is then given by:

$$\text{Contribution}^{8+}_{HL} = \beta_{8+}\left(\overline{D}^H_{8+} - \overline{D}^L_{8+}\right).$$

To express this quantity relative to the observed gap, we divide the implied score contribution by the difference in mean first-attempt PSAT scores between high- and low-challenge schools and multiply by 100. This yields the share of the observed gap that is mechanically associated with differential exposure to polluted days under the 8+ specification.

We also use this contribution to construct a no-pollution counterfactual gap. Specifically, we subtract

the implied pollution contribution from the observed gap, yielding the achievement gap that would remain under a benchmark in which neither group experienced any polluted days. Because this calculation is based on the 8+ specification, it corresponds to removing all days with PM2.5 at or above 8 μg/m^3.

As in the rest of this section, this is an accounting decomposition that holds the estimated pollution–score relationship fixed and should not be interpreted as a causal decomposition of differences across school types.

A3.6.2 Using Student-Level Mean Scores to Define the High–Low Challenge Gap

Moderate-Pollution Exposure We next construct an alternative decomposition of the achievement gap using students' average test scores across observed attempts, rather than scores at the first attempt. This provides a measure of persistent differences in achievement between students attending high- and low-challenge schools.

To do so, we compute, for each student, the mean PSAT score across all observed attempts, as well as the mean number of moderate-pollution days experienced over those same years. We then compute group means separately for students in high- and low-challenge schools, including both the average score and the average level of moderate-pollution exposure.

Let $\overline{Y}^H$ and $\overline{Y}^L$ denote the mean PSAT scores for students in high- and low-challenge schools, respectively, and let $\overline{D}^H_{8\text{–}12}$ and $\overline{D}^L_{8\text{–}12}$ denote the corresponding mean number of moderate-pollution days. Let $\beta_{8\text{–}12}$ denote the estimated per-day effect of moderate-pollution exposure. The contribution of differential exposure to the gap in average scores is then given by:

$$\text{Contribution}^{8\text{–}12}_{HL,\text{mean}} = \beta_{8\text{–}12}\left(\overline{D}^H_{8\text{–}12} - \overline{D}^L_{8\text{–}12}\right).$$

To express this quantity relative to the observed gap, we divide the implied score contribution by the difference in mean PSAT scores between high- and low-challenge schools and multiply by 100. This yields the share of the high–low challenge gap in average scores that is mechanically associated with differences in moderate-pollution exposure.

As in the previous subsection, this calculation is based solely on the moderate-pollution coefficient and therefore corresponds to an equalized-exposure counterfactual for moderate-pollution days. It should be interpreted as an accounting decomposition that holds the estimated pollution–score relationship fixed, rather than as a causal decomposition of differences across school types.

All Polluted Days (PM2.5 $\geq$ 8) We next repeat the decomposition using the 8+ pollution specification, which aggregates moderate- and heavy-pollution days into a single measure of polluted days. As in the previous subsection, the observed achievement gap is defined using students' mean PSAT scores across all observed attempts, comparing students in high- and low-challenge schools.

We compute group means for students in high- and low-challenge schools, including both the mean PSAT score and the mean number of polluted days with PM2.5 at or above 8 μg/m^3. Let $\overline{Y}^H$ and $\overline{Y}^L$ denote the mean PSAT scores for students in high- and low-challenge schools, respectively, and let $\overline{D}^H_{8+}$ and $\overline{D}^L_{8+}$ denote the corresponding mean number of polluted days. Let β_{8+} denote the estimated per-day effect from the 8+ specification. The contribution of differential exposure to the

gap in average scores is then given by:

$$\text{Contribution}^{8+}_{HL,\text{mean}} = \beta_{8+}\left(\overline{D}^{H}_{8+} - \overline{D}^{L}_{8+}\right).$$

To express this quantity relative to the observed gap, we divide the implied score contribution by the difference in mean PSAT scores between high- and low-challenge schools and multiply by 100. This yields the share of the high–low challenge gap in average scores that is mechanically associated with differential exposure to polluted days under the 8+ specification.

Because this calculation is based on the all-polluted-days specification, it corresponds to a full no-pollution benchmark in which all days with PM2.5 at or above 8 μg/m^3 are removed. As in the rest of this section, the result should be interpreted as an accounting decomposition that holds the estimated pollution–score relationship fixed, rather than as a causal decomposition of differences across school types.

A3.6.3 Allowing Pollution Effects to Differ by School Challenge

The preceding decompositions impose a common pollution–score relationship across students in high- and low-challenge schools. We next relax this restriction by allowing the estimated pollution effect to differ across the two groups. This yields a heterogeneous-slope decomposition in which both exposure and the pollution coefficient may vary by school challenge level.

To do so, we estimate the baseline binned specification separately for students in low-challenge schools and for students in high-challenge schools. Let $\beta^{L}_{8\text{–}12}$ and $\beta^{L}_{\geq 12}$ denote the estimated effects of moderate- and heavy-pollution days in low-challenge schools, and let $\beta^{H}_{8\text{–}12}$ and $\beta^{H}_{\geq 12}$ denote the corresponding effects in high-challenge schools. Let $\overline{D}^{L}_{8\text{–}12}$ and $\overline{D}^{L}_{\geq 12}$ denote the mean number of moderate- and heavy-pollution days experienced by students in low-challenge schools, and define the analogous quantities for high-challenge schools.

Using these group-specific coefficients and exposure means, we compute a group-specific "achievement cost" for each challenge category:

$$\text{Cost}_L = \beta^{L}_{8\text{–}12}\,\overline{D}^{L}_{8\text{–}12} + \beta^{L}_{\geq 12}\,\overline{D}^{L}_{\geq 12},$$

$$\text{Cost}_H = \beta^{H}_{8\text{–}12}\,\overline{D}^{H}_{8\text{–}12} + \beta^{H}_{\geq 12}\,\overline{D}^{H}_{\geq 12}.$$

The heterogeneous-slope contribution of pollution to the high–low challenge gap is then defined as the difference between these two group-specific achievement costs:

$$\text{Contribution}^{\text{het}}_{HL} = \text{Cost}_H - \text{Cost}_L.$$

We express this quantity relative to the observed high–low challenge gap by dividing by the difference in mean first-attempt PSAT scores and multiplying by 100. Unlike the common-β decompositions above, this exercise allows pollution exposure to have different effects across groups, so the resulting object reflects both differences in exposure and differences in the estimated pollution–score slope.

This exercise should be interpreted as an accounting decomposition under heterogeneous slopes, not as a causal estimate of the effects of changing school assignment or socioeconomic conditions.

Because the contribution is a nonlinear function of coefficients estimated in separate samples, formal inference for this object generally requires a bootstrap procedure.

A3.7 Changes in Achievement Associated with Declining Pollution Exposure

A3.7.1 Baseline Accounting Using the 8+ Specification

We next assess how changes in pollution exposure over time translate into changes in student achievement. Specifically, we construct a descriptive accounting measure of how much higher test scores would be expected to be in recent years relative to earlier years, given observed declines in pollution exposure.

To do so, we use the 8+ pollution specification, which defines exposure as the number of days with PM2.5 at or above 8 μg/m^3. Let β_{8+} denote the estimated per-day effect of polluted days, and let $\mathbb{E}[D_{8+}^{\text{now}}]$ and $\mathbb{E}[D_{8+}^{\text{past}}]$ denote the mean number of polluted days in a recent year and in a comparison year several years earlier, respectively.

We define the change in exposure as:

$$\Delta D_{8+} = \mathbb{E}[D_{8+}^{\text{now}}] - \mathbb{E}[D_{8+}^{\text{past}}].$$

The implied change in test scores associated with this change in exposure is then given by:

$$\Delta Y = -\beta_{8+} \cdot \Delta D_{8+}.$$

In practice, we implement this calculation by comparing the most recent year in the sample to the closest available year approximately nine years earlier, reflecting the time span over which IV-based estimates are available. We compute mean exposure in each year and scale the difference by the estimated per-day effect.

This quantity represents the model-implied change in PSAT scores attributable to the observed decline in polluted days over this period, holding the estimated pollution–score relationship fixed. It should be interpreted as a descriptive accounting exercise rather than a causal estimate of how changes in pollution over time affect achievement, as it does not account for other factors that may have evolved concurrently.

A3.7.2 Alternative Specification Using Moderate and Heavy Pollution Bins

As a robustness check, we repeat the time-trend accounting exercise using the binned pollution specification, which distinguishes between moderate- and heavy-pollution days. This allows the contribution of different pollution intensity levels to vary according to their estimated effects.

Let $\beta_{8\text{–}12}$ and $\beta_{\geq 12}$ denote the estimated per-day effects of moderate- and heavy-pollution days, respectively. Let $\mathbb{E}[D_{8\text{–}12}^{\text{now}}]$ and $\mathbb{E}[D_{\geq 12}^{\text{now}}]$ denote the mean number of moderate- and heavy-pollution days in a recent year, and define the corresponding quantities for an earlier comparison year. We compute changes in exposure for each bin as:

$$\Delta D_{8\text{–}12} = \mathbb{E}[D_{8\text{–}12}^{\text{now}}] - \mathbb{E}[D_{8\text{–}12}^{\text{past}}],$$

$$\Delta D_{\geq 12} = \mathbb{E}[D_{\geq 12}^{\text{now}}] - \mathbb{E}[D_{\geq 12}^{\text{past}}].$$

The implied change in test scores associated with these exposure changes is then given by:

$$\Delta Y = -\left[\beta_{8\text{–}12} \cdot \Delta D_{8\text{–}12} + \beta_{\geq 12} \cdot \Delta D_{\geq 12}\right].$$

In practice, we implement this calculation by comparing the most recent year in the sample to the closest available year approximately nine years earlier, and scaling the changes in exposure in each bin by their respective estimated effects.

This specification provides an alternative decomposition of the change in achievement associated with declining pollution exposure, allowing moderate- and heavy-pollution days to contribute differently. As in the baseline case, the resulting quantity should be interpreted as a descriptive accounting exercise that holds the estimated pollution–score relationship fixed, rather than as a causal estimate of the effect of time-varying pollution on achievement.