Clean Air and Cognitive Productivity

Heyes, Anthony; Cook, Nikolai; Rivers, Nicholas

DOI:
10.1086/724951

Clean Air and Cognitive Productivity: Effect and Adaptation*

Nikolai Cook
Wilfrid Laurier University

Anthony Heyes
University of Birmingham

Nicholas Rivers
University of Ottawa

January 4, 2023

Abstract

We observe 1.8 million university course grades for 88,959 adults who learn and complete examinations in a much less polluted environment than previously studied. We use a within-student identification strategy and find robust evidence of a negative and causal effect of exam-day outdoor air pollution on course performance. The effect of pollution persists beyond the same-day effect. Female students are more sensitive than males, and effects greatest when engaged in unfamiliar tasks. We explore two margins of adaptation, one infrastructural, one behavioral. Working in a new building, and particularly if it is high quality (LEED Gold), provides significant mitigation. Relocating to a floor above ground-level also offers partial protection.

Keywords: Air Pollution - Cognitive Function - Particulate Matter - Productivity

1 Introduction

Exposure to polluted air, over various time frames, is well understood to damage human health (Landrigan et al. (2018)). In addition, important recent research points to a negative

*Heyes is corresponding author and can be contacted at a.g.heyes@bham.ac.uk. We are very grateful to two referees from this journal, Matthew Neidell, Michele Baggio, Peng Zhang, Xiaohui Zhang and seminar participants at University of Connecticut, University of Ottawa, The Chinese University of Hong Kong and Peking University HSBC School of Business for insightful advice and comments at various stages of this project. Heyes acknowledges funding for this project from SSHRC under Insight Grant #435-2017-1069 entitled "Air Pollution and Human Well-being". Heyes and Rivers acknowledge financial support from the Canada Research Chair programme.

effect of pollution on mental function. It is this line of inquiry which we take forward here, with a particular focus on short-term (same-day) exposure in low- and very-low exposure environments. Using a large sample of undergraduate student course grades mapped to outdoor pollution concentration on the day of the course final exam, along with a quasi-experimental design, we show that contemporaneous pollution negatively affects exam performance and course scores even in low-pollution settings. We also provide novel evidence on the protective effect of buildings.

Any detrimental effect of pollution on mental acuity would be an important outcome in its own right, but also has the potential to provide a link in causal chains from air quality to a diverse set of outcomes important for human well-being. These include productivity in non-physical workplace tasks (Chang et al. (2019)), work quality (Archsmith et al. (2018)), quality of communication (Heyes et al. (2018)), financial decision-making (Huang et al. (2020)), consumer behavior (Ding et al. (2021)), affect (Zheng et al. (2019)), and interpersonal violence (Herrnstadt et al. (2021)). Accounting for such effects is essential for evaluating the full economic and social burden of pollution and the cost-benefit analysis of policies that improve air quality.

This paper evaluates how the ambient level of outdoor fine particulate matter ($PM_{2.5}$) at the time of the course final exam affects the grade received on the course. Final exam scores are required to make up 40 to 60% of the final course grade, allowing us to infer impacts on exam performance from observed course grade scores. We observe the 1.8 million course grades received in the period from 2007 through 2019 by 89 thousand distinct undergraduate students at the University of Ottawa, a large, comprehensive, research-intensive public university, which operates from a main campus located in the heart of the Canadian capital city of Ottawa.

At least three features or our setting make it ideal for this research:

(1) Exams provide a good quality measure of mental function in a diverse set of incentivized tasks which plausibly correlates with performance in a range of brain-intensive tasks

outside the academic environment. On average each student is observed under about 20 plausibly exogenous pollution treatments, including frequent low exposure conditions. The panel structure of the data allows us to estimate the effect of within-student differences in pollution, removing any time-invariant individual characteristics that might otherwise confound inference. The dataset is large, allowing for precise estimation of effects even in specifications containing a rich set of fixed effects and other controls.

(2) The scheduling and location of each subject/treatment pair is determined exogenously, being set by University authorities well in advance of the event, such that we can ignore issues with respect to selection into treatment that pose important challenges in some other artifactual settings.

(3) The granular character of data means that we know not just the building in which a particular subject is assigned to work under each treatment but also room location. This allows us to probe secondary research questions, including the protective benefits of working in a building of more modern design. We also provide first evidence of an important protective benefit of a simple non-technological response to pollution, namely relocation to a higher floor. The latter is consistent with a building engineering literature that points to a negative building-envelope and within-building gradient between pollution exposure and height (for example Jung et al. (2011) and citations therein).[1]

Under panel estimation, including a rich set of fixed effects and other controls, we find evidence of a substantial negative effect of same day ambient outdoor $PM_{2.5}$ concentrations on mental productivity. We find that performance is particularly compromised for females and in STEM (Science, Technology, Engineering, and Mathematics) courses, especially for

[1] Our interest throughout is in the role of outdoor air pollution, since that is what the policy-maker influences when she imposes air quality regulations. The mapping between indoor and outdoor air at any time depends on a range of factors, including building design, ventilation systems, and how the building is used by its occupants. Particulate matter also resides in the body for some time, as can the effects induced by exposure, so that the effects of pollution in the ambient outdoor environment are plausibly imported from outdoors within the person. A secondary question, therefore, is the extent to which working in a building of ‘good’ design protects against external conditions.

students for which STEM is outside their ‘usual’ program of study.

We supplement our fixed effects estimates with an instrumental variables estimator based on atmospheric thermal inversions. A thermal inversion is a meteorological phenomenon in which a layer of warm air high in the atmosphere prevents atmospheric convection, trapping air pollutants at ground level. About 30% of exams in our dataset take place under conditions characterized by thermal inversion. The quasi-random occurrence of thermal inversions in the city generates plausibly exogenous variations in local $PM_{2.5}$ levels that we exploit for identification. The instrumental variables estimator helps to correct for measurement error that results from inability to measure individual pollution exposure as well as potential omitted variables. Results point to larger impacts of pollution on performance than the fixed effects estimator, and suggest that the main fixed effects estimates may under-estimate true effect sizes.

In addition to estimating the impact of contemporaneous (on the day of the exam) pollution on course grades, we also investigate the role of pollution on days leading up to the exam. We implement a variety of distributed lag models, in which course grades are regressed on pollution on the day of the exam as well as on preceding days. These models suggest that pollution exposure during the ten days preceding the exam results in a deterioration of performance. The distributed lag models suggest that the cumulative impact of a sustained increase in pollution is larger than an increase in exam-day pollution only, and again suggest that our baseline fixed-effects estimates focused on contemporaneous pollution are a conservative estimate of overall impacts.

Ottawa is a very clean city. The annual average $PM_{2.5}$ concentration in 2018 was 6.0 μg/m^3, compared to 15.6 μg/m^3 in Paris, 50.9 μg/m^3 in Beijing and 113.5 μg/m^3 in Delhi (IQAir, 2019). The fact that air quality continues to detract from cognitive performance even in this low-pollution setting suggests that there may not be a threshold below which air pollution no longer matters. Indeed, using a non-linear specification, we show that cognitive performance deteriorates as a result of contemporaneous air pollution concentrations of 5 to

$10\mu g/m^3$ (relative to even cleaner days of 0 to $5\mu g/m^3$).

We are also interested in learning about adaptation. First, we document an important protective effect of buildings. Working in a new building (defined as one opened since 2000) mitigates around a quarter of the negative effect of ambient pollution on exam performance, and a LEED-certified one around a further quarter.[2] We also find that working on a higher floor (one above ground level), holding other things as fixed (including building) mitigates around half of the effect.

Our results complement and extend recent research on the effect that outdoor air quality has on indoor mental acuity. Zhang et al. (2018), using two waves of the China Family Panel Survey, find no effect of same-day pollution on how respondents perform on unincentivized cognitive exercises. They do however find impacts of longer term exposure, concentrated in low-educated and older males, particularly on verbal as opposed to mathematical exercises. In an early study Ebenstein et al. (2016) find a negative relationship between air pollution and scoring on university-admittance exams, with persistent impacts on wages later in life. Roth (2016) find a negative association between indoor air quality and exam performance.[3]

Two important recent studies are most pertinent for us here, relying as they do on methods explicitly designed for causal inference.[4]

First, in a recent laboratory-based study of students in Brazil, Bedi et al. (2021) find that same day $PM_{2.5}$ levels negatively impact the fluid reasoning of respondents, but find no

[2]LEED (Leadership in Energy and Environmental Design) is a building certification scheme developed by the US Green Building Council. LEED-certified buildings aim to reduce energy use and improve occupant comfort and air quality relative to non-certified buildings. See: https://www.usgbc.org/leed.

[3]In two related studies: (a) Kunn et al. (2019) study a panel of competitive chess players, finding that diminished indoor air quality is associated with an increase in the likelihood of a player making an erroneous move (against the benchmark of a chess computer); (b) Nauze and Severini (2021) find that exposure to $PM_{2.5}$ reduces performance of players in an online "brain-training" game.

[4]Graff Zivin et al. (2020) find a negative association between (upwind) agricultural fires and test scores in China and, separately, fires and air quality. They do not, however, apply the fire count data as an instrument, their main specification directly regressing test performance against fire count.

effect on the four other measures of cognition collected. One limitation, which the authors acknowledge explicitly, is that the relatively small sample from their experimental setting (n = 464) meant that the study may have been under-powered for detecting small effects across a broader set of cognitive measures. Their data is also cross-sectional rather than panel in character - they do not observe the same subject under alternative conditions - making it hard to control for relevant time-invariant individual characteristics, many of which may be unobserved and/or unrecorded.

Second, and the study closest to ours is Carneiro et al. (2021), executed contemporaneously with this. Using a large sample they find a negative impact of contemporaneous air pollution (in that case PM_{10}) on performance in university entrance tests in Brazil. Students sit exams over two days, so each subject is observed under two treatments, with pollution at exam location instrumented using wind direction.

It is this evidence that we seek to complement and extend here. As already noted, a key attraction of our setting is that observations are drawn in a low-pollution setting. Pollution exposure in Canada is amongst the lowest of any country in the world. Mean 24-hour exposure in our data is 5.5 μg/m^3 (with standard deviation of 3.9). For purposes of calibration the United States Environmental Protection Agency (USEPA) defines air quality to be "Good", its best available designation, on a day that the 24-hour measure does not exceed 12 μg/m^3, which is the case for 93% of dates in our sample. In addition, both the mean $PM_{2.5}$ concentration in our sample and the concentration on all of the days we observe fall below the relevant national standards.[5] As a result, our study provides evidence on the effect of contemporaneous pollution on cognitive performance in a setting with low to very low $PM_{2.5}$ exposure.

This contrasts with earlier studies relating to much more polluted places. In the much

[5]See: https://www.ccme.ca/en/air-quality-report. The 2020 $PM_{2.5}$ standards set by the Canadian Council of Ministers of the Environment are 27 μg/m^3 for 24-h exposure and 8.8 μg/m^3 for annual average exposure. Our sample mean concentration of 5.5 μg/m^3 is well below the annual average standard, and every day in our sample falls below the 24-h standard.

smaller sample of days (54 in total) in Bedi et al. (2021), mean $PM_{2.5}$ is 18.5 μg/m^3, and in Ebenstein et al (2016) mean $PM_{2.5}$ across the whole sample is 21.1 μg/m^3. In Carneiro et al. (2021) the whole sample average of PM_{10} is 21.1 μg/m^3.

Evaluating whether pollution has impacts on cognition in a low-exposure setting helps to establish whether existing thresholds are appropriate and to shed light on the nature of the relationship between marginal pollution impacts and ambient concentrations. The richness of our dataset allows us to uncover substantial negative effects even when $PM_{2.5}$ falls below 10 μg/m^3.

2 Data

First, we detail the administrative student data provided to us by the university - our measure of cognitive performance. Second, we detail air quality data we collect from the Ontario Ministry of the Environment and meteorological data from Environment and Climate Change Canada. Third, we detail our data on thermal inversions. Last, we report and offer some discussion of the summary statistics.

2.1 Student performance data

We obtained privileged access to administrative data from student files at the University of Ottawa as the basis for our measure of cognitive performance. We observe the universe of course outcomes for undergraduates at the University for the academic years 2007-08 through 2018-19 inclusive. We observe a comparatively large number of cognitive tasks - over 1.8 million courses and associated final exams completed by almost 89,000 unique adults over this period.

We focus on courses taken during the Fall and Winter semesters. Courses taken during the Fall semester hold final exams throughout the month of December, and exams for courses offered during the Winter semester cover most of April, with exact dates varying slightly from year to year. Exams are written during one of three time 'slots'; in the morning from 9 a.m. until 12 p.m., in the afternoon from 2 p.m. to 5 p.m., and in the evening from 7 p.m. to 10 p.m. Exams are scheduled on all seven days of the week.

Our data captures the grade that students receive in a course, rather than the grade obtained on the final exam. This feature of our data introduces a complication. While we hypothesize that final exam day air quality impacts performance on the final exam, assessment for each course also involves other elements from earlier in the semester, such as midterms and course-work.

University academic regulations require the final exam's weight to be no lower than 40% and no higher than 60% in course assessment. The weight of the final exam in the course grade is determined prior to the start of the semester, and thus should not be correlated with realized pollution on the day of the final exam. However, while this additional variation in weighting across courses is not correlated with our regressor of interest, it is not observable to us, and thus adds measurement error to the dependent variable. Measurement error of this sort would not be expected to bias our main coefficient estimates, but to reduce precision, making significance claims conservative. This feature further requires that in interpreting effect sizes, we use a multiplier to reflect the fact that any impact of exam-day conditions on exam performance has a muted impact on course-level performance - we impute the variation in exam performance as a factor of two times the variation in course performance, consistent with an assumption that the final exam carries 50% weight in every course. In doing so, a 2% decrement in overall course score, for example, would be taken to imply a 4% decrement in performance on final exam.[6]

[6]Formally, we are interested in the coefficient γ_1 in the stylized regression: $\text{exam\_grade} = \gamma_0 + \gamma_1\text{pollution} + \nu$. Lacking data on exam grades, we instead estimate $\text{course\_grade} = \beta_0 + \beta_1\text{pollution} + \epsilon$, where ν and ϵ are error terms and pollution is measured pollution on the day of the exam. The course grade scheme is determined prior to the start of the semester and is given by $\text{course\_grade} = \theta\text{exam\_grade} + (1-\theta)\text{other\_grade}$, where θ is the weight on the final exam and other_grade is the grade obtained on other course components. Substituting in, we obtain $\text{exam\_grade} = \beta_0/\theta + \beta_1/\theta\text{pollution} + \epsilon/\theta - (1/\theta - 1)\text{other\_grade}$. Because grades on other course components are orthogonal to pol-

The fact that we observe course grades rather than exam grades directly results in two limitations for our study. First, using course grade rather than exam grade as our dependent variable results in less precise coefficient estimates. In our setting, with a large number of observations, this limitation is not of major consequence. Second, if the weighting scheme is systematically different in different course types or for different students or buildings (e.g., if science courses weight the final exam higher than non-science courses) then our heterogeneity analysis may confound differences in causal effects across disciplines with differences in exam weights across disciplines. We touch on this point further below when we analyze heterogeneity in results across disciplines.

We also collect data on gender from the student files, as well as data on program of study (such as whether the student is enrolled in a STEM or non-STEM program). In addition, we obtain exam schedule and exam location data from the registrar, as well as information on the subject matter and level of each course.

2.2 Air quality and meteorological data

We collect historical air quality data from a publicly accessible on-line portal provided by the Ontario Ministry of the Environment.[7] We use data from monitor 51001, which is in the same urban environment as the campus and less than 1.3 kilometers to the northeast.[8] Particulate matter coverage is excellent - with 98.7% of valid hourly $PM_{2.5}$ observations for December and April during the sample period.[9] The station also collects data on a host of other pollutants, including carbon monoxide, nitrogen oxides, ozone, and sulfur dioxide.

lution on the day of the exam (these grades are determined prior to the final exam), we can drop other_grade from the regression without generating bias. We can thus recover γ_1, the coefficient of interest, by dividing the estimated coefficient β_1, from the regression of course_grade on exam day pollution, by θ.

[7] http://www.airqualityontario.com/history/index.php

[8] In a robustness check, we substitute pollution data from another nearby monitor, located to the southwest and slightly further away (the data on the second is unavailable for 2018 onward however).

[9] We use linear interpolation to fill in missing air pollution observations.

We aggregate the hourly measures of $PM_{2.5}$ into 24-hour averages. Specifically, if an exam is scheduled on a particular day, we assign the arithmetic mean of that date's 24 hours to the exam date - consistent with usual practice in research on pollution effects. Fine particles may reside in the body for some time after inhalation, and exam performance may be affected by conditions leading up to the exam via impacts on sleep, mood, or studying. As a result, in an extension of our main results, we also examine the impacts of pollution leading up to the final exam on performance.

We collect daily weather data from Environment and Climate Change Canada's weather station operated to the southeast of the campus (ID=6105978). Weather data includes temperature, precipitation, wind speed and direction, relative humidity, and atmospheric pressure.

2.3 Thermal inversions

In addition to surface-level meteorological data, we obtain data on above-surface temperature, which we use to determine the presence of a thermal inversion. A thermal inversion occurs when air temperature does not fall monotonically with declining air pressure (increasing altitude). Our data on above-surface temperature are obtained from the ERA5 climate reanalysis.[10] The ERA5 reanalysis incorporates large amounts of historical data (from ground monitoring stations, satellites, radiosondes, etc.) into a physically-consistent climate model, and produces hourly reanalyses of a large number of climatic phenomena. We obtain air temperature estimates for the ERA5 grid cell containing the city and University of Ottawa (75.5°W, 45.5°N) for model pressure levels within 100 hPa of surface.[11] We extract temperature data at four periods each day (0:00, 6:00, 12:00, 18:00 GMT). We define an inversion as present if the air temperature in the first model pressure level above surface is below that of the surface air temperature.

[10] See https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

[11] The ERA5 model resolves temperature at 25 hPa increments starting at 1,000 hPa.

2.4 Summary statistics

Table 1 presents summary statistics. We observe exam performance of almost 89,000 distinct undergraduate students. Each student takes an average of about 20 exams in our period, and in total we observe about 1.8 million exams, taken in 1,026 different exam slots (each exam 'slot' is a unique date-time and is thus exposed to a potentially unique level of pollution). The average performance is just over 73 percent.

The main independent variable is ambient fine particulate matter ($PM_{2.5}$).[12] The mean $PM_{2.5}$ concentration in the sample is 5.5 μg/m^3. Figure 1 plots the distribution of $PM_{2.5}$ levels on day of exam, the variation from which we identify our results.

3 Methods

3.1 OLS

We use OLS to estimate a fixed effects model of the following form:

$$\text{course\_grade}_{ise} = \beta_0 + \beta_1 \text{pollution}_{se} + \Gamma_i + \eta_s + \Delta_e + \varepsilon_{ise}. \tag{1}$$

$\text{course\_grade}_{ise}$ is the course grade of individual i in semester s in course with exam e. For ease of interpretation in the tables that follow, standardized performance will be expressed as a Z-score. In our central specification pollution_{se} is the 24-hour concentration of fine particulate matter measured in μg/m^3 on the day of the exam. The regressor of interest, β_1, is the effect of exam-day air pollution on the grade in course with exam e. As stated

[12]While we focus on $PM_{2.5}$ we recognize the potential independent role of other pollutants. Disentangling the role of individual pollutants is fraught with difficulty and particulate matter (or the closely related concept of aerosol optical depth) is often treated as a general proxy for "bad air". Figure A1 reports the variations in four other common pollutants across exam dates in our sample. As a robustness exercise we will confirm that our main results are not meaningfully disturbed when including levels of these pollutants as additional regressors in our main specifications.

above, because we observe course grades rather than exam grades, in our main results we report $2 \times \beta_1$ as the effect of pollution on exam performance (we also multiply standard errors by 2). For completeness, we also report un-scaled coefficient estimates, which can be intepreted as the effect of exam day pollution on course grades. Γ_i is a vector of individual fixed effects. η_s is a vector of semester fixed effects, with a semester defined by a Fall/Winter designation and a year, for example 'Winter 2017'. Δ_e is a vector of controls relating to exam e and includes time-of-day fixed effects, day of week fixed effects, and outdoor weather variables. Outdoor weather variables included in the regression are temperature, precipitation, relative humidity, pressure, and latitudinal and longitudinal wind. In addition, we include all two-way interactions between weather variables (e.g. precipitation by temperature).[13]

By including individual fixed effects we control for unobserved but time-invariant differences in student characteristics, such as academic ability, and identification is based on within-student variation in air pollution across dates. Exam dates are assigned by the university centrally and published weeks in advance of the start of the exam session, allaying concern about selection into treatment on the basis of actual or forecast pollution. Because of this feature of the setting, in which students cannot manipulate exposure to air pollution by selecting exam dates, we also test a model without student fixed effects.

In all of our main specifications standard errors are two-way clustered at the student and exam slot level (we apply non-nested panel clustering following Cameron et al. (2011)) which are robust to arbitrary within-student autocorrelation and contemporaneous cross-student correlation. The exam slot is the most natural level at which treatment is assigned making this approach to calculation of standard errors consistent with the advice contained in Abadie et al. (2017), though in the context of our robustness exercises we will show that

[13]The effect of outdoor temperature on indoor cognitive performance has been examined in a number of papers including Park (2020) and Cook and Heyes (2020). The rooms in which the exams are written are all protected from external temperature fluctuations by high quality climate control. Evidence that indoor temperature at this university are held a constant 20 degrees Celsius with little within-room or across-room variation is contained in Figures 4 and 5 in Cook and Heyes (2020). That variation that does occur is uncorrelated with exterior temperature.

qualitative results are the same under alternative approaches to inference.

3.2 Non-linear model

In addition to the main OLS specification, which holds the marginal effect of pollution constant, we consider a specification which allows flexibility in the marginal effect of pollution on grades:

$$\text{course\_grade}_{ise} = \beta_0 + \sum_{n=1}^{5} \beta_n \times \mathbf{1}[Bin_n(PM_{2.5})] + \Gamma_i + \eta_s + \Delta_e + \varepsilon_{ise}. \tag{2}$$

Our coefficients of interest in this non-linear specification are the β_n, which are the estimated coefficients on indicator variables which take the value of 1 if the $PM_{2.5}$ level is within the range of Bin_n and 0 otherwise. We use five exhaustive bins for $PM_{2.5}$ levels, with 5μg/m^3 increments, and drop the 0-5μg/m^3 level as the reference to which other levels are compared.

3.3 Lagged pollution

Our main OLS specification regresses course grade on exam day particulate matter concentration. However, particulate matter may remain in the body for a period of time, and may impact sleep and studying behaviour and effectiveness leading up to the exam. As a result, we estimate a model that aims to capture exposure to pollution in the days leading up to the exam. We estimate a distributed lag model, following the approach of Schwartz (2000), Zanobetti et al. (2003), and He et al. (2019). We begin by estimating an un-restricted distributed lag model, by augmenting Equation (1) to include not only concurrent pollution on the day of the exam, but also lagged pollution (along with other control variables and fixed effects). We estimate models with pollution lagged by up to $Q = 14$ days before the exam (in the specification with $Q = 14$, we estimate 14 additional coefficients). The cumulative impact of a permanent increase in pollution on exam scores is then $\beta_1 + \sum_{q=1}^{Q} \kappa_q$, where each

κ_q is the coefficient on pollution q days before the exam.

Pollution is serially correlated over time, so un-restricted distributed lag models often exhibit significant multicollinearity and recover κ_q coefficients with large standard errors, and with coefficients that bounce up and down over consecutive days of lags (Wooldridge, 2012). As a result, it is standard practice to impose some structure on the coefficients, to require that lag coefficients follow some smooth (polynomial) function: the restricted polynomial distributed lag (Schwartz, 2000; He et al., 2019; Dell et al., 2012). Using a restricted polynomial distributed lag approach, κ_q coefficients are restricted as follows: $\kappa_q = \sum_k^K \alpha_k q^k$, where the degree of the polynomial is represented by K (and where $K < Q$). The restricted polynomial distributed lag approach has the advantage of requiring the estimation of fewer coefficients (K instead of Q) and imposing some structure on the coefficients. We present results for a quadratic, cubic, and quartic ($K = 2, 3, 4$) polynomial. In each case, we differentiate between the *impact effect*, which is the coefficient (β_1) on pollution on the day of the exam, and the *cumulative effect*, which is the sum of coefficients on contemporaneous pollution as well as all lags of pollution ($\beta_1 + \sum_{q=1}^Q \kappa_q$).

3.4 Instrumental variables (IV)

Two main concerns arise in inference based on OLS estimation in this context. First, measurement error might arise with respect to the pollution exposure of our subjects. Exposure is proxied by conditions at the nearest available air quality monitor, around 1.3 kilometers from the University campus, and individual exposure is also plausibly sensitive to the behavior and movement of individuals during the day, which is unobserved. Such measurement error, if classical, would be expected to attenuate the estimated effect sizes and understate the statistical significance of those effects. Second, there are variables that are unobserved or otherwise omitted from the regression that plausibly correlate with both daily air quality and student performance. For example traffic congestion in the vicinity of campus might cause stress to students arriving for exams, or noise disturbance, along with air pollution.

To address these and other endogeneity issues, and reinforce the causal nature of the relationship, we re-estimate Equation (1) but replacing measured pollution with predicted level of pollution, using presence of thermal inversion as an instrument.

Thermal inversions have been used as an instrument for air pollution in several other recent papers, with other outcomes of interest, to address endogeneity and measurement error in air pollution (He et al. (2019), Chen et al. (2017, 2018), Dechezlepêtre et al. (2019), Jans et al. (2018), Sager (2019), Arceo et al. (2016), Heyes and Zhu (2019)). Under standard atmospheric conditions, air temperature falls with increasing altitude (decreasing pressure) above surface. During thermal inversion episodes, a pocket of cooler air becomes trapped below a mass of warm air, inverting the normal monotonically negative relationship between air temperature and altitude. Thermal inversions inhibit the dissipation of air pollution from the surface, and can give rise to elevated levels of air pollution, making them a relevant instrument for pollution.

A number of distinct meteorological phenomena can give rise to thermal inversions, including the continental-scale movement of air masses, the high-latitude warming of upper air masses before lower air masses in winter as a result of low-angle insolation, and overnight surface cooling (Dechezleprêtre et al. (2019)). Because thermal inversions are an upper-atmosphere phenomenon, after controlling for surface-level meteorology, there is no reason to believe they affect surface-level outcomes except through their effect on pollution.

Our identifying assumption is that the induced increase in fine particulate matter concentrations is the only channel through which inversion influences mental function, conditional on surface-level weather controls included in the regression. One potential violation of this exclusion restriction can occur if thermal inversions affect concentrations of pollutants other than fine particulate matter. While fine particulate matter is considered the most important pollutant for determining health in many contexts, concentrations of other pollutants can also affect health outcomes. As a result, the results using the instrumental variables specification can be considered the gross effect of inversion-generated pollution increases, which are proxied by fine particulate levels.

The first stage in our two-stage least squares regression is:

$$\widehat{pollution}_e = \alpha_0 + \alpha_1 Inversion_e + \Gamma_i + \eta_s + \Delta_e + u_{ise} \tag{3}$$

where $Inversion_e$ is a binary variable that takes the value 1 if there is a thermal inversion (as defined above) in the 6-hour period immediately preceding the date and time of the examination and 0 on other days. The second stage in our 2SLS model is:

$$\text{course\_grade}_{ise} = \beta_0 + \beta_1 \widehat{pollution}_e + \Gamma_i + \eta_s + \Delta_e + \varepsilon_{ise}. \tag{4}$$

We also report the reduced form, which captures the effect of inversions on course grades directly:

$$\text{course\_grade}_{ise} = \beta_0 + \beta_1 Inversion_e + \Gamma_i + \eta_s + \Delta_e + \varepsilon_{ise}. \tag{5}$$

In their study of weak instruments in practice, Andrews et al. (2019) note that many estimators and tests exist for settings with independent and homoskedastic data, assumptions that are commonly violated. We follow their recommendation by reporting the F-statistic of Olea and Pflueger (2013) which is equivalent to the Kleibergen and Paap (2006) statistic in our single regressor case. We also report the critical value of the Stock and Yogo (2005) test statistic.

4 Results

Before proceeding to regression results we present some preliminary graphical analysis of the data.

Figure 3 provides a simple plot of exam performance and exam day $PM_{2.5}$ with the latter in bins 0.1 μg/m^3 wide, no controls applied and the size of bubble proportionate to the number of data points contained. Given this is a mixture of observations across many years, students and disciplines, etc., we should not over-interpret, however visually the plot is weakly suggestive of a negative association. The aim of the rest of the paper is to explore that relation in more detail.

4.1 OLS

Table 2 summarizes the results from estimation of Equation (1). All specifications include student and semester fixed effects.

Column 1 reports the sparsest specification which includes only semester (e.g. 'Winter 2017'), day-of-week, and time-of-day fixed effects. The top two panels also include student fixed effects. Column 2 controls for exam day weather (temperature, precipitation, relative humidity, air pressure, latitudinal and longitudinal wind speed). Column 3 also controls for interactions of weather variables with one another (all possible two-way interactions are included).

For each specification, we report three results (each deriving from a separate regression). The top panel is the main result as described in the text. It is derived from a regression of course grades on exam day pollution and is scaled by a factor of two to capture the approximately 50% weighting of the exam in the course grade. The middle panel reports un-scaled results, from the same regression as in the top panel but without the multiplicative scaling. These results correspond to the causal effect of exam-day pollution on the overall course grade. Our aim is to understand how exam-day pollution affects exam-day performance, and so we focus on results in the top panel (subsequent results in later tables also follow this approach, implementing the $2\times$ scaling of coefficients and standard errors, and using student fixed effects). The bottom panel reports results from a specification without student fixed effects. Eliminating student fixed effects helps to preserve additional variation in the data, and since exam dates are set in well in advance, it should not come at the cost of identification.

In each of columns 1 through 3 the coefficient on $PM_{2.5}$ is negative and significant at conventional levels. Coefficient values increase in absolute value as we control for weather.

We focus on the results in the top panel. In column 3 the estimate of -0.607 achieves significance at a level much higher than 1% level (the associated t-statistic is -4.9). The estimate in column 3 implies that a $1\mu g/m^3$ increase in $PM_{2.5}$ is associated with a decrement in performance of 0.6% of a SD. Alternatively, a 1 standard deviation increase in $PM_{2.5}$ is associated with a decrement in performance of 2.4% of a standard deviation.[14]

4.2 Non-linear effects

Table 3 and Figure 6 provide a characterization of non-linear impacts of air quality on cognitive performance. We construct dummy variables that bin the particulate matter concentration into exhaustive bins with width $5\mu g/m^3$. We run the same regression as above, but with these dummy variables replacing the pollution variable. We treat the first bin ($0\text{-}5\mu g/m^3$) as the reference to which others are compared. Columns in Table 3 capture the effect of controlling more exhaustively for weather. All specifications show that increasing pollution generates larger decrements in exam performance. The effect appears to be approximately linear. Importantly, the results suggest that even when pollution is in the range of 5 to $10\mu g/m^3$, exam performance is compromised relative to when pollution is below $5\mu g/m^3$. Thus the evidence suggests that the current Canadian 24-h standard of $27\mu g/m^3$ for 24-h exposure (see Footnote 5) is not conservative with respect to cognitive impacts from air pollution.

4.3 Lagged pollution

The prior results relate pollution on the day of the exam to course and exam performance. However, pollution levels on days preceding the exam may also affect exam performance, potentially by affecting performance while studying for the exam, sleep quality, if pollution remains in the body for multiple days, or if any health decrement from pollution takes time to resolve. In this section, we conduct robustness checks by including lags of pollution to ensure that our main results are not disturbed. We report results both using an unrestricted

[14] From Table 1, the standard deviation of $PM_{2.5}$ in our sample is $3.9\mu g/m^3$.

finite distributed lag model, as well as results where lag coefficients are constrained to follow a polynomial structure.

Results are reported in Table 4. In each column of the table, we report the *impact effect* as well as the *cumulative effect* for regressions of course grade on exam-day pollution as well as lagged pollution, where the number of lagged days of pollution included in the regression varies by column.[15] The impact effect is the coefficient on exam day pollution, and captures how contemporaneous pollution affects exam-day performance. The cumulative effect is the sum of all coefficients on lagged pollution plus the coefficient on contemporaneous pollution, and captures how a sustained increase in pollution on all days leading up to and on the exam affects exam-day performance. We report results from four separate regressions for each choice of lag length, with model A reflecting unrestricted lags, and subsequent panels reflecting restricted quadratic, cubic and quartic polynomial distributed lag structures.

The results from each of the four models are quite similar and suggest two general outcomes with respect to lagged pollution. First, the impact effect is relatively undisturbed as a result of the addition of lagged pollution. Echoing earlier results, it suggests that exam-day pollution is harmful to exam performance. In Table 4, the impact effect of a 1μg/m^3 increase in contemporaneous pollution varies from 0.5% to 1.3% of a standard deviation and is highly statistically significant in all cases, again pointing to the significant decrement in performance from elevated exam-day performance of similar magnitude to that found in Table 2. In each case, the cumulative effect is larger than the impact effect, suggesting that the impact of pollution persists beyond the immediate. The four models show that the effects of $PM_{2.5}$ on exam-day performance continue to increase up to about 10 days prior to the exam (after this, increasing the number of lags does not increase the cumulative effect). A sustained increase in pollution of 1 μg/m^3 for 10 days leading up to the exam causes a decrement in exam performance of about 8% of a standard deviation; much larger than the impact effect. Other studies using a similar approach point to similar sustained impacts of

[15]As in prior results, coefficients are scaled by a factor of 2 to reflect the average 50% weight of exams in course grades.

pollution, and likewise find cumulative impacts that are significantly larger than contemporaneous impacts alone. He et al. (2019) find that pollution impacts productivity for up to 25 days following an increase in pollution, and Zanobetti et al. (2003) find pollution impacts mortality in the month leading up to death. In each case, these studies find that cumulative effects are substantially in excess of impact effects. For example, He et al. (2019) report that cumulative impacts of pollution on productivity are more than 10 times greater than contemporaneous impacts alone, and Zanobetti et al. (2003) report that cumulative impacts of air pollution on mortality are between two and five times greater than contemporaneous impacts alone.

4.4 Instrumental variables approach

Table 5 reports the results of the instrumental variables regression. The first line of the table is the first stage, the regression of $PM_{2.5}$ concentration on the binary inversion variable. Controls and sample restrictions coincide with those in Table 2. The preferred full-sample specification is column 3, which includes controls for weather (temperature, precipitation, relative humidity, air pressure, latitudinal and longitudinal wind speed) as well as all two-way interactions between these variables. Column 3 confirms the relevance and potency of the instrument, consistent with theoretical understanding of the air quality effects of thermal inversions and the large confirmatory empirical literature. The association between inversion and $PM_{2.5}$ is positive and highly statistically significant in all specifications. The main whole sample estimate with the full set of weather variables and interactions, in column 3, implies that inversion on day of exam increases $PM_{2.5}$ concentration by $1.2\mu g/m^3$ or approximately one third of a standard deviation. In addition to these summary variables, Figure 5 plots the distribution of $PM_{2.5}$ on days with and without a thermal inversion. The figure shows that the presence of a thermal inversion is associated with a shift in the distribution of surface-level air pollution concentrations, confirming visually the relevance of the instrument.

Second stage IV estimates follow in the second row of Table 5. The specification in column 3 is estimated on the whole sample and includes the full set of controls. As in prior tables, coefficients and standard errors are scaled by two to reflect the 50% weighting of the final exam in course grades. The estimated coefficient in column 3 implies that a 1μg/m^3 increase in PM$_{2.5}$ causes a 9% SD decrement in exam performance. Alternatively, a 1 SD increase in PM$_{2.5}$ causes performance to decrease by 35% of a SD. This coefficient estimate is substantially larger than that from column 3 in Table 2, consistent with our expectation that OLS estimates would be attenuated.

Table 5 reports the first stage F statistic, which is around 30 to 57 depending on the specification. In each case, the F statistic is larger than the Stock and Yogo (2005) critical value, suggesting that our selected instrument is strong.

For completeness Table 5 also reports results from the reduced form, that is the regression of the exam outcome on the binary inversion variable. Other things equal our preferred whole sample estimate (Column 3) is that thermal inversion on day of exam reduces cognitive performance by 11% of a SD. Our identifying assumption is that once controls are included the only mechanism through which that reduction occurs is the induced effect on air quality.

4.5 Heterogeneity

Here we explore heterogeneity of effect across sex of student and type of assessment task, by interacting the pollution regressor with student sex or course type (science, technology, engineering and mathematics (STEM) vs. non-STEM course). We focus on heterogeneity in these dimensions because there is some prior evidence that air pollution has differential impacts on females compared to males, and because there is evidence that air pollution impacts performance on different types of tasks differentially. Note that while we make strong causal claims about the effect of pollution on exam scores, our heterogeneity analysis requires more cautious interpretation. Specifically, we compare the causal effect of pollution on exam scores across groups (e.g., male vs. female). Because there may be unobserved differences across groups (e.g., males and females may choose different courses), our heterogeneity analysis is suggestive, rather than causal.

Column 1 in Table 6 repeats the central whole sample estimate from the main regressions in Table 2. We begin our analysis of heterogeneity by focusing on differences in the effect of air pollution on exam scores by sex in Column 2, by adding a regressor that interacts the $PM_{2.5}$ variable with a dummy variable that takes the value 1 if a student is female, 0 otherwise. Prior work on air pollution has found differences in the response by sex over various time horizons. Zhang et al. (2018) use data from two waves of the China Family Panel Study and compare within-individual test scores taken on days with different pollution, roughly four years apart. They find that pollution causes a larger decrement in verbal reasoning scores for men compared to women. In contrast, the reverse is true for math scores (but the gender differences are much smaller for math scores). They explain the large impact of pollution on male verbal reasoning physiologically: males use a smaller amount of "white matter" – brain cells that play a connecting role within the brain – for verbal reasoning than females, and impacts of pollution appear to be concentrated on white matter. Carneiro et al. (2021) and Ebenstein et al. (2016) report a consistent finding: in their studies of university entrance exam-takers in Brazil and Israel, males are more severely impacted by pollution than females. They speculate that this may be a result of a higher prevalence of asthma in males compared to females, as well as because male respiratory function is more inhibited by pollution compared to females. In our regression, the estimated coefficient on the interaction between female and air pollution is negative and significant. The central estimate suggest that performance of females is around three times as sensitive to variations in air quality as that of males. The finding that females are more sensitive than males in our sample contradicts most recent findings in this literature.

In the remaining columns we differentiate between courses that are coded as STEM (science, technology, engineering or mathematics) and those that are not. Just as different jobs rely on different brain functions, so too plausibly do exams in different disciplines, and those functions may vary in the extent to which they are compromised by pollution exposure.[16] Before continuing it is worth noting that a limitation of these results stems from the fact that we do not observe exam scores directly, but only course grades, as discussed previously. Specifically, in comparing the impact of exam-day pollution on course grades in different course types, it is possible that part of the difference in the effect of pollution on course grades is due to different weighting schemes applied in different courses. For example, if STEM courses use higher weights for exams that non-STEM courses, this will result in our approach recovering a higher marginal effect of pollution on STEM grades than non-STEM grades. Because the weight of the final exam in the course grade is constrained to fall between 40 and 60% (see data discussion) it is possible that differing weighting schemes account for as much as a 50% difference in the marginal effect of pollution (60/40-1). The following results can be interpreted with this caveat.

In column 3 of Table 6 we add to the basic specification with full controls a term that interacts the $PM_{2.5}$ regressor with a dummy variable that takes the value 1 if the exam in question relates to a STEM-coded course, 0 otherwise. The coefficient on $PM_{2.5}$ remains negative. The coefficient on the interaction term is negative, large in absolute value and significant. The implied decrement to performance of an increase in air pollution on exam day is around 2 times larger in a STEM exam than in an exam from the whole sample, suggesting that the effect that we find in the whole sample is partly driven by what is going on in STEM.

We then probe the question of whether familiarity with a genre of test might mitigate the debilitating effect of pollution exposure. Column 4 re-estimates the specification from column 3 but only on STEM students, column 5 on their non-STEM counterparts. Column 5 would include, for example, a business major taking a computer science course as an elective. Comparing the estimated coefficients on the interaction term across these columns points to the decrement in performance caused by pollution being much greater - more than twice the

[16]Recall that Bedi et al (2021) find that exposure to high levels of $PM_{2.5}$ reduces performance on fluid reasoning but find no effect on several other cognitive tests including attention, arithmetic processing speed, and working memory.

size - for students not habituated to that sort of mental exercise (the non-STEMs).[17] Note that this is not reporting simply that non-STEM students perform worse on STEM courses (these are all within-student estimates), rather their performance is especially compromised by poor conditions.

4.6 Pollution during semester

The focus of this paper is on the effect on performance of short-term or contemporaneous $PM_{2.5}$ exposure. An important parallel strand of research addresses the question of longer-run exposure on mental capacity including on academic achievement of school children (Heissel et al. (2020)) and risk of dementia in older adults (Bishop et al. (2018)).

That result from this literature that identifies negative effects of exposure over longer periods, combined with plausible serial correlation in air quality at our study location, raises the possibility that our results do not capture truly contemporaneous effects but rather erroneously attribute the effect over the course of the semester to exam day conditions. Our subjects may perform poorly in an exam not because their mental function is compromised at that time, but in part or whole because conditions during the semester - when the course was in progress - were not conducive to study and/or efficient learning.[18]

Our within-subject design with day to day variation in pollution can plausibly be expected to have isolated our results from such confounding. Subjects that experienced symmetric semester-level conditions face varying exam-time treatment by virtue of alternative scheduling of exams.

Nonetheless in Table 7 we report the results of additional exercises designed to probe

[17] In their study of online brain-training Nauze and Severini (2021) found effects of exposure to be largest in new or unfamiliar tasks.

[18] Longer run effects bring into play additional potential mechanisms that are not relevant for us. For example Currie et al. (2007) is one of several papers that link air pollution to school absence. Balakrishnan and Tsaneva (2021) link exposure over the course of an academic year to reduced performance by Indian school-children in unincentivised reading and math tests, concluding that "... school attendance is the main mechanism explaining these impacts" (page 1).

this concern explicitly. We do this by adding, one at a time, to the main specification four different regressors chosen to proxy air quality during the semester, to the model with contemporaneous (exam-day) pollution. In order to include these regressors, we must drop the semester fixed effects that we include in the main regressions, replacing them instead with a year time trend. Column 1 includes the mean daily $PM_{2.5}$ level measured during the teaching part of the semester. Column 2 includes median daily $PM_{2.5}$ during the semester. Columns 3 and 4 include a count of the number of days during the semester where daily mean $PM_{2.5}$ concentration exceeded 10 and $12\mu g/m^3$ respectively (these levels have both been used as indicators of cities with acceptable air quality).

In each case the coefficient on the new regressor is negative and statistically significant. However, interpreting these new coefficients is fraught with difficulty, because unlike the coefficients on contemporaneous pollution, they are not causally identified with the same quasi-experimental approach as the coefficients on contemporaneous pollution. As a result, we make no claims about these longer run effects here. The primary purpose of these additional exercises is to ensure the stability of our main estimates to these inclusions. The stability of coefficient estimates across the top row reinforces our claim that the main specifications capture a short-run (same day) impact.

4.7 Adaptation

To understand the protective benefits of modern construction as well as of holding exams on upper floors of buildings, we run our central specification but with additional regressors that interact the treatment variable with various binary variables reflecting the status of the building in which the exam was held or the location within the building. In order to run these regressions, we remove student fixed effects and replace them with course fixed effects. The thought experiment in this case corresponds to comparing grades taken when a given course exam is held in a different building or room. The first column result in Table 8 reports the coefficient estimate from a non-interacted regression. The point estimate is somewhat higher in absolute value than in Table 2, where student fixed effects are included. This is the baseline specification that we use to run our adaptation regressions.

4.7.1 Upper floors

Epidemiological and other studies of the impacts of air pollution typically draw data from monitoring networks designed to capture ground- or surface-level conditions.[19] However, a number of studies suggest that external pollution levels depend upon height above ground level, others show that air quality *within* buildings varies systematically with height: "... vertical pollution dispersion can reduce exposure to ambient pollutants in tall buildings, as concentrations of some ground-source pollutants are diluted at higher floors" (Stephens et al. (2019: 26)). In their own sampling, for example, Stephens et al. (2019) find that PM_1 and $PM_{2.5}$ concentrations are respectively about 34% and 30% lower at the building envelope at the 44th floor compared to the 2nd floor, with the gradient approximately linear. Though sensitive to building design and local circumstances dilution in several studies is found to be particularly pronounced at lower levels. For example Kumar et al. (2009) found that particulate matter at the street level of building envelopes (0.2 - 2.6 m above ground) about six times higher than those at 20 m.[20] Stephens et al. (2019) provide a useful survey.

Dispersion with height provides a potentially low-tech., behavioral adaptation to decrements in performance due to pollution, namely going upstairs (particularly on highly polluted

[19] In European Union countries: "Monitoring stations should be in the breathing zone of people on the ground, i.e. they should be positioned at a height of 1.5 m to 2 m" (European Union (2019: 28)), although the bounds for admissible monitor heights is 1.5 â€" 4 m (European Union (2019)). In the United States the analogous limits on heights for micro- and middle-level particulate matter monitors is 2 m â€" 7 m (United States Code of Federal Regulations (no date)). The monitor used for this study draws from an intake height of 4 m; see: http://www.airqualityontario.com/history/station.php?stationid=51001.

[20] In addition to physical measurements of air quality the diluting effect of height has been argued to contribute to reductions in all-cause mortality associated with increasing residential floor height in high-rise buildings in Switzerland (Panczak et al. (2013)), and higher self-reports of building-related sickness by occupants working on lower floors of office buildings in the US (Mendell et al. (2008)).

days, and particularly when engaged in mentally-demanding tasks). In a similar manner, many householders with access to a basement retreat to it during hot weather. While the subjects in our study are assigned an exam location, such behavior could be a private mechanism of self-protection in some settings, and could inform approaches to management of space by organizations.

Columns 2-4 of Table 8 summarizes the results of adding a regressor that interacts the $PM_{2.5}$ treatment variable with a dummy variable that takes the value 1 if the student writes the exam in question in a room located on an "Upper Floor", 0 otherwise.[21] In all cases "Upper Floor" is defined as anything higher than ground floor.

The interaction term alone is added in column 2. The estimated coefficient on that regressor is positive and significant at better than 1%. The central estimate suggests that moving an exam to an upper floor mitigates, other things being equal, about a half of the overall decrement due to polluted air. Note that these are *within-course* estimates, such that the difficulty of the course is held constant in these estimations.

In the subsequent two columns, we restrict the sample to older and newer buildings, respectively, with a "New "building defined as one that was built after the year 2000. We find that the protective effect of upper-floor classrooms is largest in older buildings (where the effect of pollution is itself greater).

4.7.2 Buildings

An important element of weighing the social and economic benefits of air quality improvements is that most regulations target exterior air quality, whereas people typically spend most of their time indoors. If building quality can partially or completely insulate occupants from the negative effects of external conditions then building expenditure can provide a private defensive technology against a negative public good, just as buildings can be redesigned to

[21]Not unusually for a university located in a cold city, the University of Ottawa has extensive space below grade in a number of its buildings, and around 16% of the exams in our sample take place in rooms at basement level.

be less at risk from emerging flood risks or to extreme heat events related to climate change.

Outdoor pollutants can, to varying extents, penetrate indoor work and living spaces. For example Krebs et al. (2021) examine crowd-sourced data from around 1,000 indoor and outdoor air quality monitors in California, finding that a 10% increase in outdoor particulate matter concentration in the vicinity of a building is associated with a 4.2 to 6.1% increase in the concentration indoors. Penetration occurs rapidly and almost entirely within a time period of 5 hours, with some evidence that the extent of penetration depends on building age.

This raises the natural question of the extent to which being in a modern building can mitigate against the effects identified here, a question which the fine-grained nature of our dataset allows us to explore. The University of Ottawa was founded in 1848 but has grown substantially in recent decades. The campus is comprised of a mixture of old buildings (for instance the original main building, Tabaret Hall, was originally built in 1856, and still houses numerous teaching and examination spaces) and high-specification modern buildings.

We code buildings on two measures. First, we interact pollution concentration with a dummy that takes the value 1 if the exam takes place in a "New" building, defined as built since 2000, zero otherwise. The results of this are reported in column 5 of Table 8. The coefficient on the new regressor is positive and statistically significant. The coefficient size suggests that, other things equal, relocation of an exam from an old to a new building reduces the effect of variation in exterior air quality on indoor performance by about a quarter. This is consistent with the finding of Krebs et al. (2021) that penetration of particulate matter into buildings increases with building age.

Campus construction in recent times has been of high quality and with environmental credentials in mind. A number of buildings, for example, have obtained "Leadership in Energy and Environmental Design" (LEED) certification at various levels.[22]

[22]LEED is probably the most highly-regarded green building rating system globally. It is operated by the non-profit US Green Building Council, can be applied both to new construction and renovations, and provides a framework for efficient, healthy and cost-saving green building. By 2015 there were over 80,000 LEED-certified buildings worldwide, several in our study setting. See Wei et al. (2015) for detail on the air quality requirements for LEED

The specification in column 6 of Table 8 includes, instead of the interaction term just described, a regressor that interacts the treatment with a dummy variable that takes the value 1 if the exam takes place in a building that is LEED-certified, 0 otherwise (we also include the New dummy variable, since LEED buildings are relatively new). Perhaps surprisingly, the coefficient on this is very small and does not come close to significance, though it is worth noting that the LEED-certified exams contribute only about 6% of the sample, so this is a comparatively low-powered test. Interestingly when we re-code the LEED dummy to take the value 1 only when an exam is taken in a building certified at the more stringent LEED-Gold level we find a larger and statistically significant protective effect.[23]

Taken as a whole the results in Table 8 support several claims. First, modern buildings provide an important protection against the cognitive effect of air pollution. Second, there is variation *among* modern buildings as to the extent of that protection. Third, even in high-specification, modern buildings - for example built and operated to a quality consistent with LEED-Gold certification - the mitigation is incomplete, part of the cognitive burden of polluted outdoor air remains.[24] Fourth, particularly within older buildings, moving to upper floors offers some protection from impacts of air pollution.

4.8 Robustness

We investigate the robustness of our main results to a variety of alternative modeling approaches and data choices.

Exposure definition Our preferred measure of same-day ambient outdoor $PM_{2.5}$ is a 24-

certification at various levels.

[23]The only LEED Gold building on campus is the large Faculty of Social Science (FSS) Building, opened in 2013. While space in that building is not dedicated exclusively to teaching and assessment in social science disciplines, those disciplines are over-weighted in our sample. To allay concern that here we are picking up a 'social science' effect, rather than a true building effect, we re-estimate this specification but excluding social science courses (not shown). Results are not significantly disturbed.

[24]This may be because indoor air quality is imperfectly insulated from outdoor, or because of lagged effects from subjects breathing air while outside, or a combination of the two.

hour average based on calendar day and calculated by taking the twenty-four hourly measures for a given date and taking the arithmetic mean. This is a widely-used metric and easy to understand but it does represent a modeling choice. Hourly $PM_{2.5}$ levels within a day are strongly serially correlated so that we do not endeavour to disentangle the relative importance of pollution levels *within* a day. However in Table 9 we re-estimate the main OLS specification with the $PM_{2.5}$ regressor constructed in four alternative ways. Column 2 uses the hourly $PM_{2.5}$ measure at the exam start time. Column 3 uses the mean of the measure in the three hours before the exam start time. Column 4 uses the mean of the measure for the three hours starting at the exam time. In column 5 the pre-working day average (from 1 am through 7 am on the morning of the exam). The main coefficient changes little across columns.

Table 9 reports two further exercises with respect to the $PM_{2.5}$ regressor that relate to the air quality monitor. In column 6, in the preferred OLS specification we replace the $PM_{2.5}$ series from the monitoring station closest to campus with the corresponding series from the next closest monitoring site. This has little effect on the estimated coefficient, consistent with ambient particulate matter concentrations being rather homogenous across locations within a city.[25]

Finally, as part of a national program of improvement, the pollution monitor at the site from which we draw paper was upgraded in 2012. The new instruments, of the sort used by USEPA, measure an additional portion (semi-volatile) of the fine particulate matter mass not captured by the older instruments, making concentrations measured after the change not directly comparable with earlier ones (see page 11 of Environment and Climate Change Canada (2018)). This seems unlikely to be a problem, as the additional portion of matter measured is small. Furthermore, all of our specifications contain semester (in effect year-month) fixed effects. Nonetheless to allay any residual concern column 7 in Table 9 summarizes the result of re-estimating the baseline specification but only on post-2012 data with

[25]The high correlation between daily $PM_{2.5}$ measures at the two stations can be seen from Appendix Figure A2.

little change in conclusion.

Other pollutants We have focused on $PM_{2.5}$, consistent with most of the literature, such as those papers already cited, on the mental impacts of pollution but also acute health effects. Table 10 reports the outcome of adding daily measures of other common (AQI) air pollutants, sequentially, to our preferred OLS specification.[26] The distribution of other pollutants on exam day is given in Appendix Figure A1. Our interest is not on the coefficient on any of those co-pollutants, but rather, to be assured that accounting for variations in those other pollutants does not substantively disturb inference with respect to $PM_{2.5}$. The estimated coefficients on the $PM_{2.5}$ regressor varies somewhat across columns. In particular, controlling for carbon monoxide causes the effect of $PM_{2.5}$ to become insignificant and flip sign. However, many of these pollutants have common sources with $PM_{2.5}$(e.g., CO is released from the transport sector) and controlling for both may be inappropriate.

Outliers In Table 11 we report three additional exercises to confirm that the headline results, taken from estimation over the whole support, are not driven wholly or predominantly by extreme values of the treatment variable. In column 2 we winsorize the $PM_{2.5}$ series from below at the 10% level, in column 3 we winsorize the series from above, and in column 4 both winsorizations at the same time. As can be seen the estimated coefficients on the $PM_{2.5}$ regressor remain negative and highly significant in all cases, and relatively undisturbed in value. In other words even if we ignore all of the variation in the pollution regressors lowest and highest 10% of values the analysis delivers similar conclusions to the baseline. These results confirm earlier results suggesting that low-dose exposure to $PM_{2.5}$ is detrimental to cognitive productivity.

Alternative standard errors Statistical significance of our main results was assessed based on heteroskedasticty-robust standard errors two-way clustered at the student and exam slot level (recall that an exam slot comprises a date and time). We believe this most closely consistent with Abadie et al. (2017). However Table 12 reports the result of three alterna-

[26]The number of exams in the estimating samples varies slightly between columns because of missing data in each pollution series.

tive approaches to calculation of standard errors that might plausibly have been used. Our selected approach to inference is conservative relative to alternatives.

Placebos Placebo tests are falsification tests of study design that involve estimating the proposed specification but with the regressor series of interest replaced by an alternative or placebo series that we know to be irrelevant. If such estimation delivers significant results that suggests that something in the design of the study itself is flawed, and the main results may be spurious. We conduct a placebo test for our main OLS specification and report results in Figure 6. The figure reports the results from running 1,000 placebo regressions where the true value of $PM_{2.5}$ on the day of the exam is replaced with a randomly selected value of $PM_{2.5}$ drawn from within 500 days before and 500 days following the exam date (we exclude pollution within 30 days of the exam date in running these placebo regressions). The left panel shows a histogram of the coefficients on placebo $PM_{2.5}$. These are clustered around zero, and indicate that pollution on days far removed from the exam day has no effect on exam performance (as expected). The diagram also indicates the main effect of contemporaneous $PM_{2.5}$ on exam performance, and shows that this is substantially different from the distribution of placebo coefficients. In the right panel, we report a similar distribution for t-statistics associated with placebo coefficients. Again, these are clustered at zero, and none are as large as the t-statistic reported in our main regression. This placebo test reinforces the causal interpretation of our prior findings.

5 Conclusions

While continuing to flesh out our understanding of the health impacts of air pollution remains a high priority, potentially one of the most important insights from recent empirical research in this area is that pollution exposure can impair mental function.

We take this line of work forward in a number of ways. We find highly statistically significant effects of same day, outdoor $PM_{2.5}$ on how adults perform in a series of high stakes tests. This is done in a big data setting using methods that nullify a number of important endogeneity challenges and allow for explicitly causal inference. The results extend the important recent findings of Bedi et al (2021) and Carneiro et al (2021).

In addition to causally identifying the effect of pollution on exam performance, our setting allows us to estimate effects in a low exposure environment. The USEPA 24-hour standard is that the 98th percentile of 24-hour $PM_{2.5}$ not exceed 35 μg/m^3 and categorizes air quality conditions on a particular as "Good" if the 24-hour average is below 12 μg/m^3. The mean 24-hour measure in our dataset is 5.5 μg/m^3 and only 7% of exams are taken on days where levels exceed 12 μg/m^3.

Even in this relatively low-pollution environment, we find significant impacts of air pollution on cognition (exam performance). Our causal estimates suggest that each additional μg/m^3 of fine particulate matter results in a decrement of exam performance equal to 0.6% of a standard deviation. Our non-linear specifications show that detrimental impacts of pollution persist even when particulate concentrations are below 10 μg/m^3. These results suggest that even air quality characterized as "Good" quality by regulators can have detrimental impacts on performance. We supplement the main results with an instrumental variables specification based on thermal inversions as well as a distributed lag specification, both of which point to detrimental impacts of air pollution being substantially larger than our headline estimate.

Secondary results point to the performance of females being more sensitive than that of males, counter to existing evidence, and the effect being *much* larger in STEM-coded activities. Further research aimed at pinning down precisely which dimensions of mental function are most affected (the sort of work pursued by, for example, Bedi et al (2021)) would be extremely valuable.

With respect to mechanism, in an important sense the pathways that link air quality to outcomes matter only to the extent that they might inform adaptation. Regulation generates reductions in ambient $PM_{2.5}$ levels and so improved outcomes, of one sort or another. While conventionally the outcomes considered have largely related to health, this recent line of work points to potential gains in cognitive performance. Weighing the costs of air quality improvements against the benefits is central to efficient policy prescription, and that process requires knowledge only of the ‘reduced form’ link from air quality to outcome. Nonetheless, on mechanisms and adaptation our results lead to several conclusions;

First, contemporaneous, not just lagged exposure “carried in” from a subject’s time outside, matters. If that were not the case the character of the internal environment in which performance was observed would not matter in our design.

Second, executing the task in a building built since 2000 protects against about a third of the effect.

Third, building quality matters. If in addition to being new a building is LEED-certified this mitigates a further third or so of the overall effect. Importantly, even in a building built to exacting standards, for example to a specification sufficient to attain LEED Gold status, protection is incomplete; exterior conditions still matter.

Fourth, *within* a particular building relocating from the ground floor to a higher floor of a building protects against a portion (up to a half) of the overall impact, consistent with some existing research on vertical gradients in particulate matter both within buildings and at their external envelope. While here subjects do not select into location, this finding suggests a potentially low-tech. adaptive strategy to elevated pollution levels, namely going upstairs.

Finally, parts of our heterogeneity analysis point to the possible role of individual-level task habituation. Different work tasks are routinized and familiar to varying degrees. The decrement to performance due to elevated pollution levels is *much* greater (up to three times the size) when a subject is engaged in a task outside of what they are used to.

6 Tables

Table 1: Summary Statistics

	Mean	Std. Dev.
Final Course Grade (Percent Scale)	73.02	14.78
$PM_{2.5}$ (Daily Avg. $\mu g/m^3$)	5.52	3.94
90 Day Average $PM_{2.5}$	5.60	1.36
90 Day Median $PM_{2.5}$	4.50	1.33
Days in Past 90 with $PM_{2.5}$>10	11.41	6.17
Days in Past 90 with $PM_{2.5}$>12	6.58	3.90
Temperature (Daily Avg. °C)	0.86	8.27
Precipitation (Daily Avg. mm)	0.06	0.17
Relative Humidity (%)	71.86	16.50
Pressure (kPa)	100.61	0.91
Longitudinal Wind Vector (Daily Avg. km/h)	-0.58	3.13
Latitudinal Wind Vector (Daily Avg. km/h)	0.18	2.37
Inversion Present	0.33	0.47
CO (Daily Avg. ppm)	0.23	0.07
NO_2 (Daily Avg. ppb)	9.10	5.50
O_3 (Daily Avg. ppb)	26.30	9.89
SO_2 (Daily Avg. ppb)	0.47	0.81
Female	0.59	0.49
STEM Course	0.30	0.46
New Building (Built After 2000)	0.33	0.47
LEED Certification	0.06	0.23
Upper Floor (Above Ground Level)	0.41	0.49
Observations	1,806,513	
Students	88,959	
Exam slots	1,026	

Each observation is a course grade. Final course grade presented here in levels for exposition only. Daily average $PM_{2.5}$ (and other weather variables) are the arithmetic mean of 24 hour observations. Precipitation includes snow following Environment and Climate Change Canada's 10:1 water content conversion. Longitudinal wind speed is the sum of 24 hourly wind speeds multiplied by the cosine of hourly direction. Summary statistics restricted to regression sample.

Table 2: Exam Day Pollution and Final Grade

	(1)	(2)	(3)
Preferred Specification			
PM2.5	-0.325***	-0.669***	-0.607***
	(0.099)	(0.115)	(0.124)
No Scaling			
PM2.5	-0.163***	-0.334***	-0.304***
	(0.050)	(0.058)	(0.062)
No Student Fixed Effects			
PM2.5	-0.596***	-0.944***	-0.783***
	(0.130)	(0.149)	(0.160)
Weather Controls		Y	Y
Weather Interactions			Y
Observations	1,806,513	1,806,513	1,806,513
Students	88,959	88,959	88,959
Exams	1,026	1,026	1,026

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day average $PM_{2.5}$. Equation 1 estimated using ordinary least squares. Every column (except in third panel) includes student fixed effects. Every column includes semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 3: Exam Day Pollution on Final Grade (Non-Linear)

	(1)	(2)	(3)
PM2.5 $\in [0, 5)$	Baseline	Baseline	Baseline
PM2.5 $\in [5, 10)$	-3.857***	-5.089***	-7.932***
	(0.635)	(0.643)	(0.676)
PM2.5 $\in [10, 15)$	-3.787***	-6.257***	-9.142***
	(0.734)	(0.766)	(0.817)
PM2.5 $\in [15, 20)$	-5.388***	-8.682***	-12.006***
	(0.850)	(0.912)	(0.974)
PM2.5 $\in [20, 22.69]$	-16.331***	-18.805***	-20.338***
	(1.497)	(1.527)	(1.599)
Weather Controls		Y	Y
Weather Interactions			Y
Observations	1,806,513	1,806,513	1,806,513
Students	88,959	88,959	88,959
Exams	1,026	1,026	1,026

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is binned exam day average $PM_{2.5}$. Equation 2 estimated using ordinary least squares. Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 4: Polynomial Distributed Lag

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Lags Included	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Model A: Unrestricted Distributed Lags														
Impact Effect	-1.29	-1.25	-1.23	-1.17	-1.20	-1.03	-1.04	-1.14	-1.10	-0.90	-1.00	-0.84	-0.86	-0.52
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
Cumulative Effect	-8.55	-7.64	-8.67	-7.94	-8.19	-6.05	-5.84	-5.16	-4.83	-3.71	-2.96	-2.40	-2.00	-0.84
	(0.61)	(0.54)	(0.49)	(0.43)	(0.38)	(0.34)	(0.32)	(0.30)	(0.27)	(0.25)	(0.23)	(0.20)	(0.18)	(0.15)
Model B: Quadratic Distributed Lags														
Impact Effect	-0.840	-0.79	-0.91	-0.92	-0.92	-0.76	-0.83	-0.83	-0.89	-0.75	-0.52	-0.52	-0.86	-0.52
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)	(0.12)	(0.13)	(0.13)
Cumulative Effect	-8.05	-7.69	-8.65	-8.28	-7.62	-6.13	-5.93	-5.29	-4.81	-3.73	-2.80	-2.43	-2.00	-0.84
	(0.60)	(0.53)	(0.48)	(0.42)	(0.37)	(0.34)	(0.32)	(0.30)	(0.27)	(0.25)	(0.23)	(0.20)	(0.18)	(0.15)
Model C: Cubic Distributed Lags														
Impact Effect	-0.92	-0.99	-0.96	-0.90	-0.72	-0.86	-0.87	-0.92	-0.70	-0.58	-0.67	-0.84	-0.86	-0.52
	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.13)	(0.13)
Cumulative Effect	-8.10	-7.75	-8.65	-8.28	-7.56	-6.16	-5.94	-5.30	-4.79	-3.70	-2.80	-2.40	-2.00	-0.84
	(0.60)	(0.53)	(0.48)	(0.42)	(0.37)	(0.34)	(0.32)	(0.30)	(0.27)	(0.25)	(0.23)	(0.20)	(0.18)	(0.15)
Model D: Quartic Distributed Lags														
Impact Effect	-0.96	-0.85	-0.89	-0.92	-1.05	-0.85	-0.88	-0.78	-0.81	-0.83	-1.00	-0.84	-0.86	-0.52
	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
Cumulative Effect	-8.12	-7.66	-8.56	-8.30	-7.79	-6.16	-5.94	-5.20	-4.82	-3.72	-2.96	-2.40	-2.00	-0.84
	(0.60)	(0.53)	(0.49)	(0.43)	(0.37)	(0.34)	(0.32)	(0.30)	(0.27)	(0.25)	(0.23)	(0.20)	(0.18)	(0.15)
Student FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Semester FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Day-of-Week FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time-of-Day FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m	1.8m

Significance markers removed to save space. The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day average $PM_{2.5}$. Estimated using ordinary least squares. Model A uses no restrictions on the estimated coefficients. Model B,C, and D require the coefficients on lagged $PM_{2.5}$ to follow a quadratic, cubic, and quartic polynomial respectively. Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level.

Table 5: Instrumented Exam Day Pollution on Final Course Grades

	(1)	(2)	(3)
First Stage		DV: Exam Day PM2.5	
Inversion Before Exam	2.105***	1.682***	1.235***
	(0.854)	(0.593)	(0.454)
Second Stage		DV: Final Grade	
PM2.5	-4.950***	-6.011***	-9.032***
	(0.433)	(2.073)	(1.458)
Reduced Form		DV: Final Grade	
Inversion Before Exam	-10.421**	-10.111**	-11.152**
	(4.195)	(5.064)	(5.159)
Weather Controls		Y	Y
Weather Interactions			Y
Observations	1,806,513	1,806,513	1,806,513
Students	88,959	88,959	88,959
Exams	1,026	1,026	1,026
Kleibergen-Paap F	57	48	30
Stock & Yogo (2005)	16.38	16.38	16.38

First Stage: The dependent variable is exam day $PM_{2.5}$ concentration. The primary independent variable is an indicator for the presence of a temperature inversion during the exam. **Second Stage:** The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is instrumented exam day $PM_{2.5}$. **Reduced Form:** The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is an indicator for the presence of a temperature inversion during the exam. **All Panels:** Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 6: Heterogeneity

	(1)	(2)	(3)	(4)	(5)
PM2.5	-0.607***	-0.282	-0.560***	-0.423	-0.067
	(0.124)	(0.172)	(0.135)	(0.265)	(0.161)
PM2.5 × Female		-0.553***			
		(0.187)			
PM2.5 × STEM Course			-0.502**	-1.646***	-5.109***
			(0.196)	(0.281)	(0.778)
Sample Restriction	All	All	All	STEM	Non-STEM
Weather Controls	Y	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y	Y
Observations	1,806,513	1,806,513	1,806,513	769,007	1,037,506

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day average $PM_{2.5}$. The secondary independent variables are indicators for whether a student is (a) female or (b) taking a STEM course. Sample restrictions refer to whether a student is enrolled in a STEM program. Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 7: Robustness to Inclusion of Semester-level Pollution Metrics (OLS)

	(1)	(2)	(3)	(4)
PM2.5	-0.623***	-0.574***	-0.659***	-0.585***
	(0.124)	(0.124)	(0.124)	(0.124)
90 Day Average PM2.5	-9.311***			
	(2.144)			
90 Day Median PM2.5		-11.416***		
		(2.608)		
Number of Days in Past 90 with PM2.5>10			-2.329***	
			(0.295)	
Number of Days in Past 90 with PM2.5>12				-1.609***
				(0.438)
Weather Controls	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y
Observations	1,806,513	1,806,513	1,806,513	1,806,513
Students	88,959	88,959	88,959	88,959
Exams	1,026	1,026	1,026	1,026

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day average $PM_{2.5}$. The secondary independent variables include the average PM2.5 measured over the semester, the number of days during the semester with an average PM2.5 reading above 10, the proportion of time during the semester spent inverted, and the number of days in the semster with at least one inversion. Every column includes student, day-of-week, and time-of-day fixed effects and a year time trend. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 8: Adaptation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PM2.5	-1.191***	-2.161***	-3.248***	-0.995**	-1.584***	-1.711***	-1.894***
	(0.170)	(0.316)	(0.509)	(0.475)	(0.264)	(0.299)	(0.299)
PM2.5 × Upper Floor		1.164***	2.283***	0.0651			
		(0.317)	(0.546)	(0.474)			
PM2.5 × New Building					0.446*	0.575*	0.755**
					(0.269)	(0.303)	(0.302)
PM2.5 × LEED Cert.						0.452	
						(0.531)	
PM2.5 × LEED Gold Cert.							1.179**
							(0.529)
Sample Restriction	All Bldg.	All Bldg.	Old Bldg.	New Bldg.	All Bldg.	All Bldg.	All Bldg.
Weather Controls	Y	Y	Y	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y	Y	Y	Y
Observations	1,807,688	1,807,688	1,190,295	595,147	1,785,442	1,785,442	1,785,442

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day $PM_{2.5}$. Upper floor is defined as an exam written on the second floor or higher (i.e. above the basement and ground floors). A new building is defined as one built after the year 2000. LEED certification by a third-party (Leadership in Energy and Environmental Design) certifies a building was built to achieve top marks in sustainable site development, water savings, energy efficiency, and (importantly for us) indoor environmental quality. At the time of writing there are four LEED certified buildings on campus, three of which hold exams. They are the Faculty of Social Sciences (c. 2012, Gold), Learning Crossroads (c. 2018, Silver) and STEM Buildings (c. 2018, Silver). Estimated using ordinary least squares. Every column includes **course**, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the course level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 9: Robustness to PM2.5 Definition (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Pref	T	(T-3,T)	(T,T+3)	1h-7h	Alt. Mtr.	2013 On
PM2.5	-0.607***	-0.634***	-0.632***	-0.269***	-1.169***	-0.649***	-0.970***
	(0.124)	(0.090)	(0.093)	(0.093)	(0.090)	(0.120)	(0.150)
Weather Controls	Y	Y	Y	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y	Y	Y	Y
Observations	1,807,688	1,807,688	1,807,688	1,807,688	1,807,688	854,296	1,114,617

The dependent variable is final course grade, measured in hundredths of a standard deviation. Each column uses a different definition of exam day $PM_{2.5}$ concentration. (1) is the preferred specification, which uses 24-hour average $PM_{2.5}$. (2) uses hour of exam $PM_{2.5}$. (3) uses the average of $PM_{2.5}$ for three hours prior to the exam. (4) uses the average of $PM_{2.5}$ during the three-hour exam time slot. (5) uses the average of $PM_{2.5}$ during 1 a.m. to 7 a.m. morning of the exam. (6) uses 24-hour average $PM_{2.5}$ measured at the next closest air quality monitor. (7) uses 24-hour average $PM_{2.5}$ after TEOM technology was replaced by SHARP for the air quality monitoring system (SHARP has better cold-weather performance). Each observation is a student exam. Data from 2007 through 2019, inclusive. Standard errors are clustered at the student and exam-slot level. (*** p<0.01, ** p<0.05, * p<0.1.)

Table 10: Robustness to Criteria Air Pollutants (OLS)

	(1) Pref. Spec.	(2) Incl. CO	(3) Incl. NO2	(4) Incl. O3	(5) Incl. SO2	(6) Incl. Multi
PM2.5	-0.607***	0.135	-0.826***	-0.272**	-0.285**	-0.171
	(0.124)	(0.163)	(0.157)	(0.131)	(0.129)	(0.171)
CO		Y				Y
NO2			Y			Y
O3				Y		Y
SO2					Y	Y
Weather Controls	Y	Y	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y	Y	Y
Observations	1,806,513	1,801,973	1,806,513	1,806,513	1,806,513	1,801,973

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day average $PM_{2.5}$. The secondary independent variables include daily average concentrations for criteria air pollutants. Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 11: Robustness to Outliers (OLS)

	(1) Pref. Spec.	(2) Wins. High	(3) Wins. Low	(4) Wins. Both
PM2.5	-0.607***	-0.608***	-0.580***	-0.560***
	(0.124)	(0.163)	(0.126)	(0.168)
Weather Controls	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y
Observations	1,806,513	1,806,513	1,806,513	1,806,513

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day average $PM_{2.5}$. Winsorization high refers to winsorizing the highest 10% of PM 2.5 concetrations. Winsorization low refers to winsorizing the lowest 10% of PM 2.5 concetrations. Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

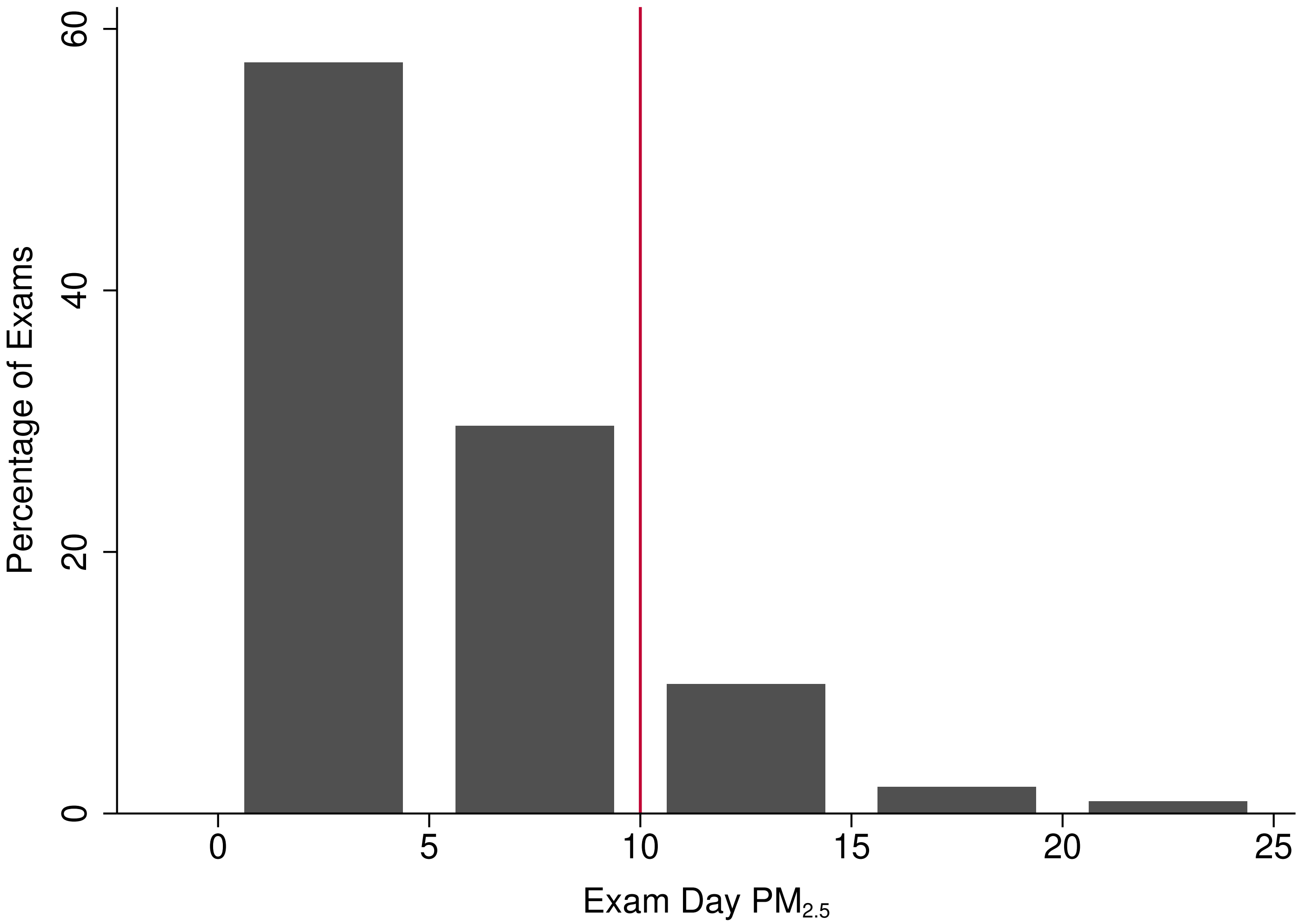
Table 12: Robustness to Alternative Clustering (OLS)

	(1) Clust. Stud. x Slot	(2) Clust. Student	(3) Clust. Stud. x Date
PM2.5	-0.60704***	-0.60704***	-0.60704***
	(0.12397)	(0.12290)	(0.12368)
Weather Controls	Y	Y	Y
Weather Interactions	Y	Y	Y
Observations	1,806,513	1,806,513	1,806,513

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day average $PM_{2.5}$. The first column uses our preferred clustered errors at the student $\times$ exam slot level. The second column uses clustered errors at the student level. The third column uses clustered errors at the student $\times$ date level. Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

7 Figures

Figure 1: Variation in Exam Day PM2.5



Histogram of exam day 24-hour $PM_{2.5}$ concentrations. A vertical line at $10\mu g/m^3$ is provided.

Figure 2: Variation in Semester-level PM2.5 Measures

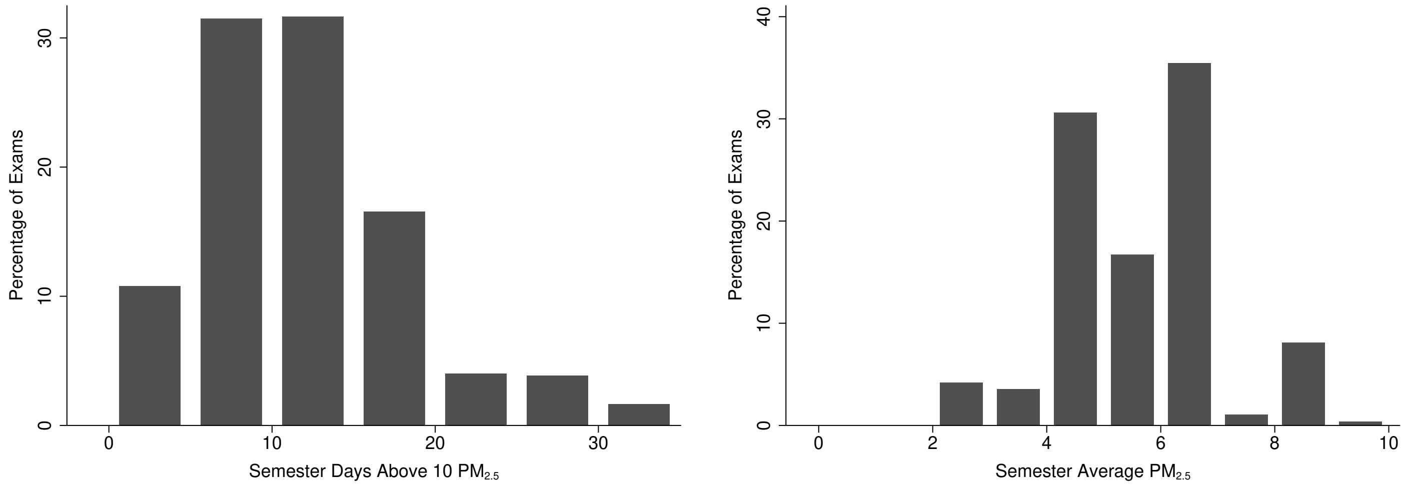
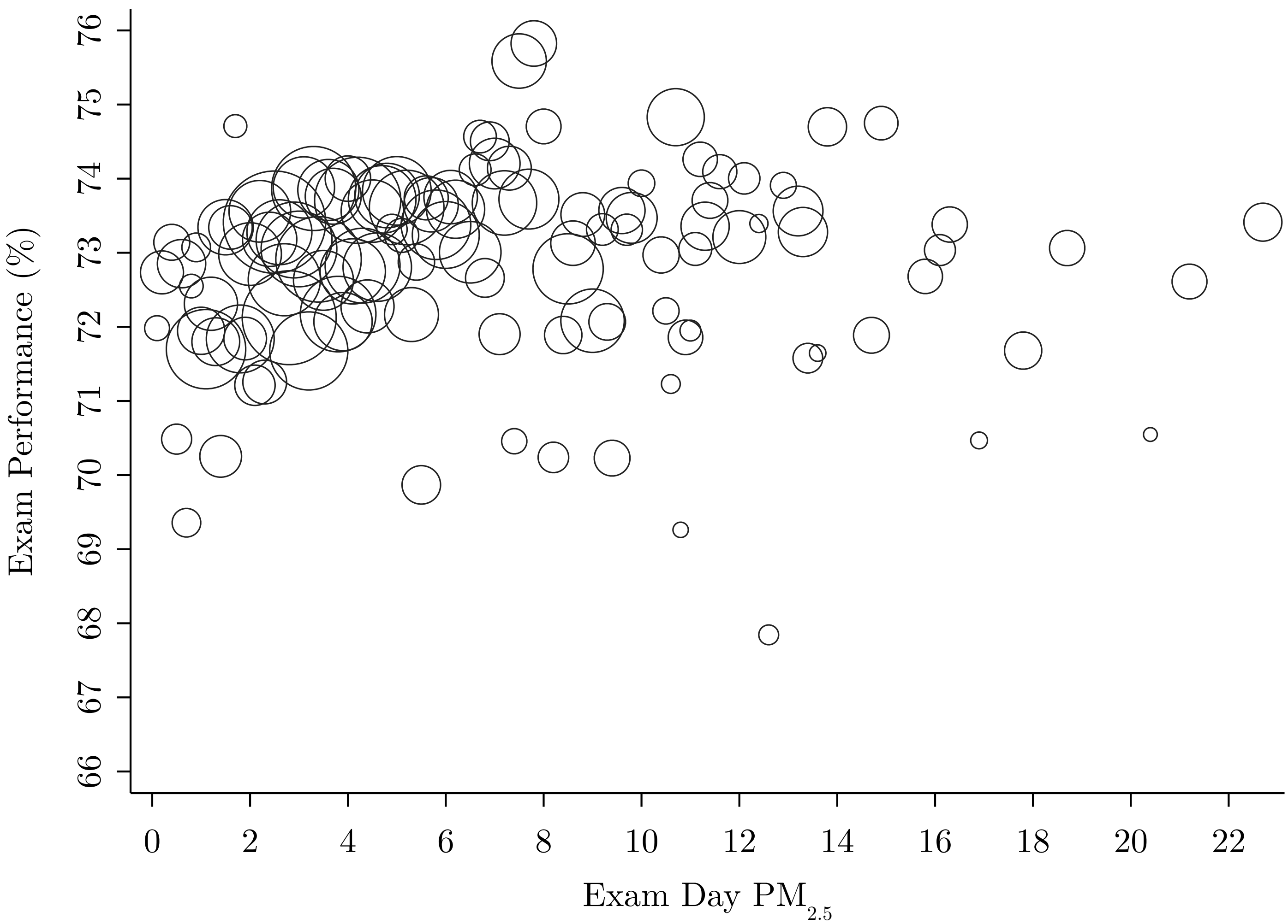
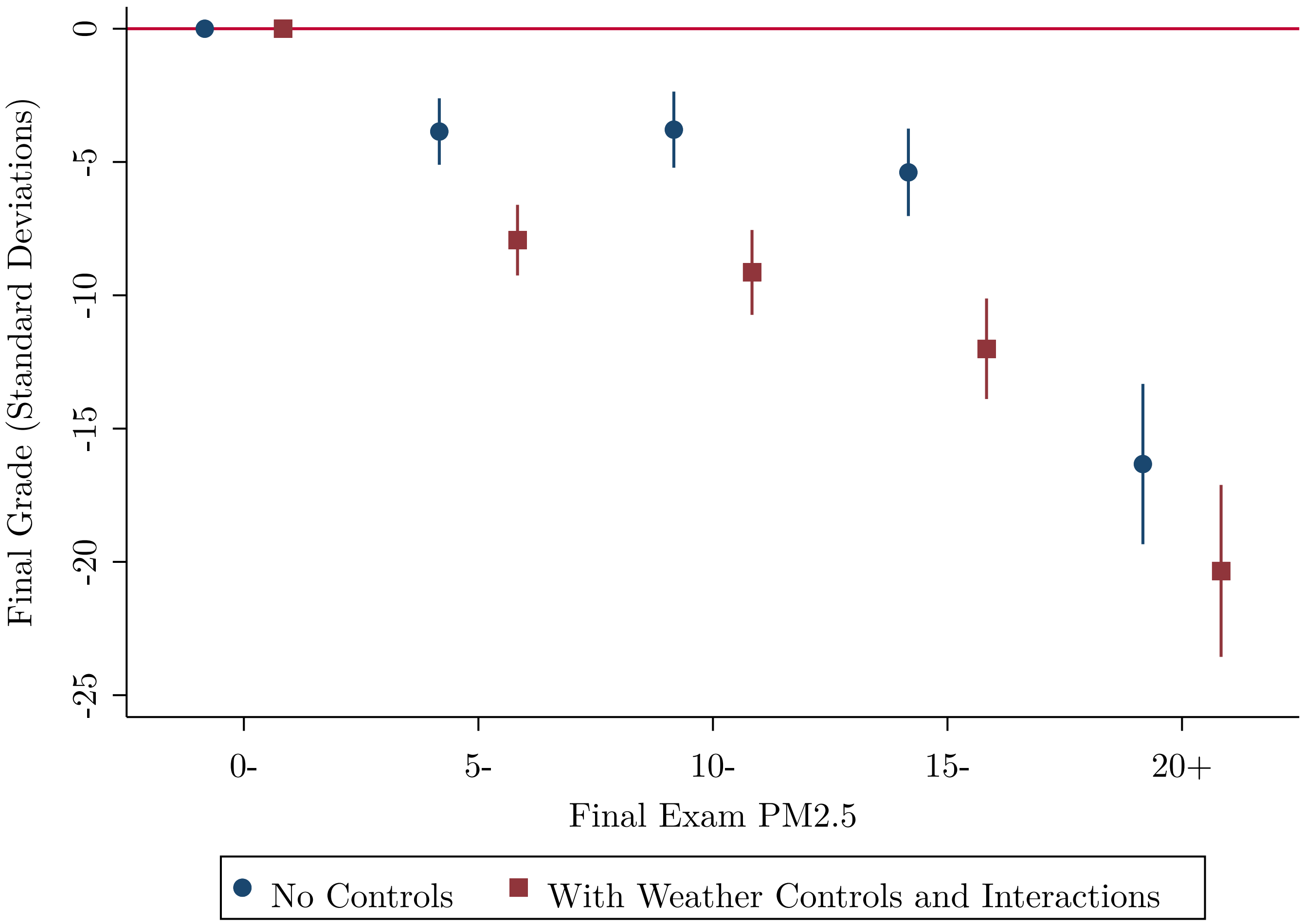


Figure 3: Performance and Pollution (No Controls)



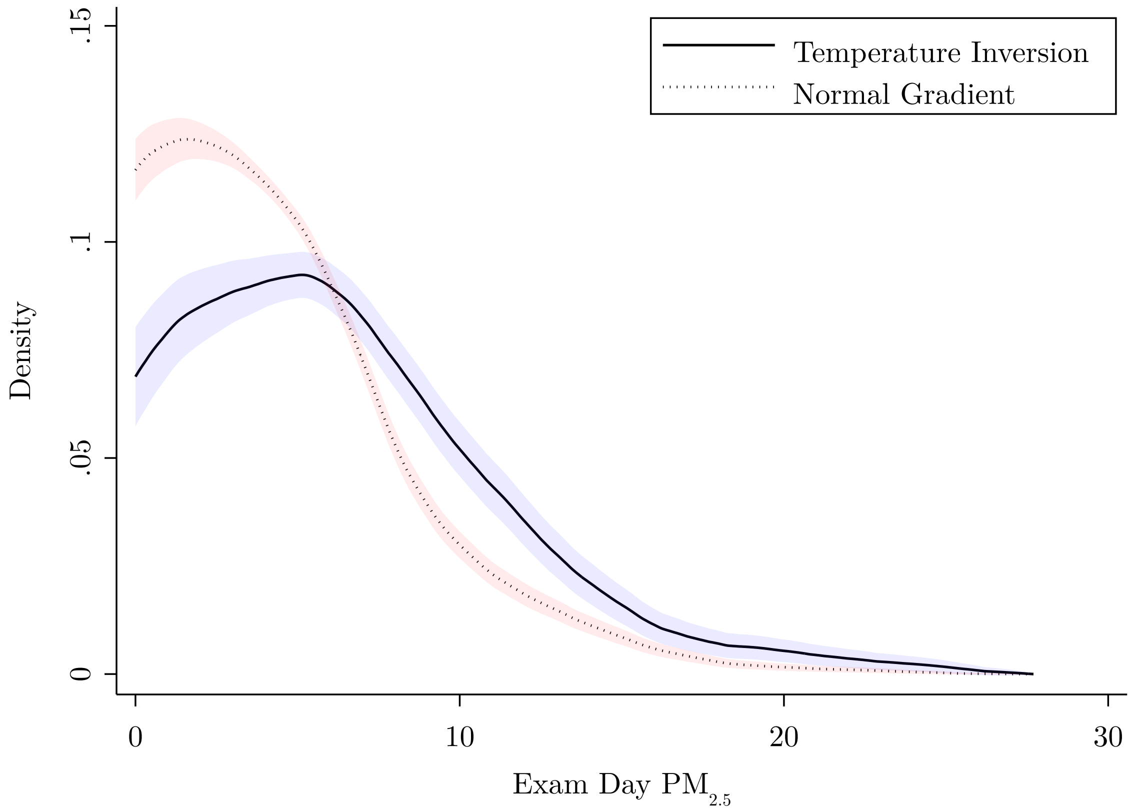
In this figure, we plot realized course grade by exam day particulate matter. $PM_{2.5}$ is rounded to the nearest 0.1 μg/m^3. Markers are proportional to the number of observations they represent.

Figure 4: Exam Day Pollution and Final Grade (Non-Linear)



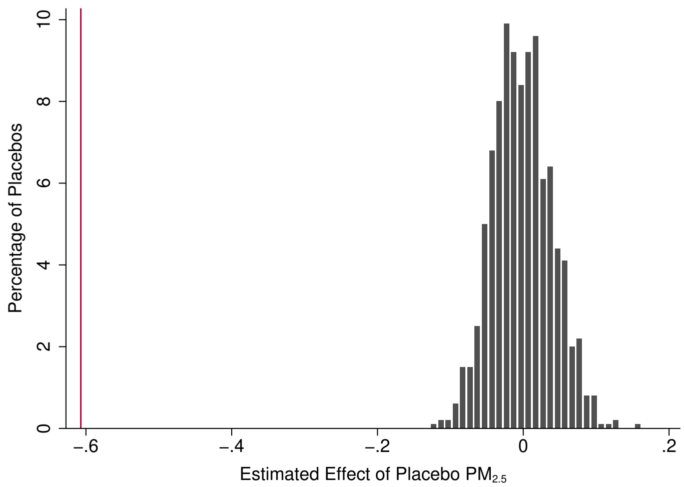
Estimated coefficients from two regressions (also presented in Table 3). Blue markers correspond to a specification with student, semester, and exam-slot fixed effects only. Red markers correspond to a specification with additional weather controls and their interactions. The horizontal axis indicates which 'bin' the exam day PM2.5 corresponds to (for example, 5.5 $\mu g/m^3$ would be assigned the second bin from the left.) The vertical axis represents the magnitude of the estimated coefficients in hundredths of a standard deviation (for example 10 represents a magnitude of 0.10 standard deviations.)

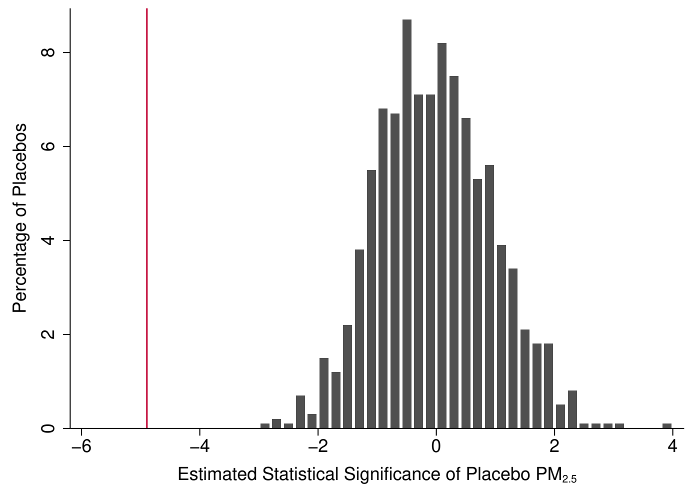
Figure 5: Exam Day $PM_{2.5}$ By Inversion Status



Exam Day PM2.5 by inversion status. Two kernel densities are presented. The dashed line and red confidence intervals correspond to normal temperature gradient conditions. The solid line and blue confidence intervals correspond to temperature inversion conditions. We note a rightward shift of mass under a temperature inversion, corresponding to higher PM2.5 levels.

Figure 6: Placebo Regressions





All Panels: Histograms corresponding to 1,000 placebo regressions. Placebo PM2.5 corresponds to the recorded air pollution level from a randomly chosen day between 500 days before and 500 days after the true date. If a date within 30 days of the exam day was chosen, the placebo date was re-randomized. Vertical lines (in red for online copy) indicate our preferred specification's estimated regression coefficient and statistical significance. **Left Panel:** Histogram of regression coefficients of exam-day PM2.5 on final grade. **Right Panel:** Histogram of the associated t-statistics, where 37 of 1,000 placebos are statistically significant at the 5% level.

8 Bibliography

Abadie, A., S. Athey, C. Imbens and J. Wooldridge (2017). "When should you adjust standard errors for clustering?" MIT Working Paper.

Andrews, L., J. Stock and L. Sun (2019). "Weak instruments in instrumental variables regression: Theory and practice." *Annual Review of Economics* 11: 727-53.

Apte, J. S., Marshall, J. D., Cohen, A. J., and Brauer, M. (2015). Addressing global mortality from ambient PM2. 5. *Environmental Science & Technology*, 49(13), 8057-8066.

Arceo, E., Hanna, R., and Oliva, P. (2016). "Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City." *Economic Journal*, 126(591): 257-280.

Archsmith, J., A. Heyes and S. Saberian (2018). "Air quality and error quantity: Pollution and performance in a high-skilled, quality-focussed profession." *Journal of the Association of Environmental and Resource Economists* 5(4): 827 - 63.

Balakrishnan, U., and Tsaneva, M. (2021). Air pollution and academic performance: Evidence from India. *World Development*, 146, 105553.

Bedi, A., M. Nakaguma, B. 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): 443-74

Bishop, K., J. Ketcham, N. Kiminoff (2018). "Hazed and confused: The effect of air pollution on dementia." *National Bureau of Economic Research* Working Paper #24970.

Burnett, R. T., Pope III, C. A., Ezzati, M., Olives, C., Lim, S. S., Mehta, S., Shin, H., Singh, G., Hubbell, B., Brauer, M., Anderson, R., Smith, K., Balmes, J., Bruce, N., Kan, H., Laden, F., Turner, M., Gapstur, S., Diver, R., Cohen, A. (2014). "An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure." *Environmental Health Perspectives*, 122(4), 397-403.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2), 238-249.

Carneiro, J., M. Cole and E. Strobl (2021). "The effects of air pollution on students' cognitive performance: Evidence from Brazilian university entrance tests." forthcoming *Journal of the Association of Environmental and Resource Economists*.

Chang, T., J. Graff Zivin, T. Gross and M. Neidell (2019). "The effect of pollution on worker productivity: Evidence from call center workers in China." *American Economic Journal: Applied Economics* 11(1): 151-72.

Chen, S., Oliva, P., and Zhang, P. (2017). "The effect of air pollution on migration: Evidence from China." National Bureau of Economic Research Working Paper No. w24036.

Chen, S., Oliva, P., and Zhang, P. (2018). "Air pollution and mental health: Evidence from China." National Bureau of Economic Research Working Paper No. w24686.

Cook, N. and A. Heyes (2020). "Brain freeze: Outdoor cold and indoor cognitive perfromance." *Journal of Environmental Economics and Management* 101 (May).

Currie, J., E. Hanushek, E. Kahn, M. Neidell and S. Rivkin (2009). “Does pollution increase school absences?” *Review of Economics and Statistics* 91(4): 682-94.

Dechezlepretre, A., Rivers, N., and Stadler, B. (2019). “The economic cost of air pollution: Evidence from Europe.” Organisation for Economic Co-operation and Development Working Paper No. 1584.

Dell, M., Jones, B., and Olken, B.. “Temperature shocks and economic growth: Evidence from the last half century.”*American Economic Journal: Macroeconomics* 4(3): 66-95.

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

Ding, Y., J. Zhong G. Guo and F. Chen (2021). “The impact of reduced visibility caused by air pollution on construal level.” *Psychology Marketing* 38: 129-41.

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.

Environment and Climate Change Canada (2018). *Canadian Environmental Sustainability Indicators: Air Quality*. Ottawa: Canada, 113 pp.

European Union (2019). Sampling points for air quality. *Representativeness and comparability of measurement in accordance with Directive 2008/50/EC on ambient air quality and cleaner air for Europe*. Policy Department for Economic, Scientific, and Quality of Life Policies. Luxembourg, 104 pp.

Graff-Zivin, J., Liu, T., Song, Y., Tang, Q., and Zhang, P. (2020). The unintended impacts of agricultural fires: Human capital in China. *Journal of Development Economics*, 147, 102560.

Heissel, J., C. Persico and D. Simon (2020). “Does pollution drive achievement? The effect of traffic pollution on academic performance.” *Journal of Human Resources* 56(2): 406-45.

He, J., Liu, H. and A. Salvo (2019). “Severe air pollution and labor productivity: Evidence from industrial towns in China.”*American Economic Journal: Applied Economics* 11(1): 173-201.

Herrnstadt, E., A. Heyes, E. Muehlegger and S. Saberian (2021). “Air pollution as a cause of violent crime: Microgeographic evidence from Chicago.” Forthcoming *American Economic Journal: Applied Economics*.

Heyes, A. N. Rivers and B. Schaufele (2018). “Pollution and politician productivity: The effect of PM on MPs.” *Land Economics* 95(2): 157 - 73.

Heyes, A. and M. Zhu (2019). “Air pollution as a cause of sleeplessness: Evidence from a panel of Chinese cities using social media metrics.” *Journal of Environmental Economics and Management* 98 (November): 1 - 20.

Huang, J., N. Xu and H. Yu (2020). “Pollution and performance: Do investors make worse trades on hazy days?” *Management Science* 66(10): 4455-4476.

IQAir (2019). *2018 World air quality report: Region and city PM2.5 ranking.* Accessed at: https://www.iqair.com/ca/world-most-polluted-cities.

Jans, J., P. Johansson and J. Nilsson, (2018). “Economic status, air quality, and child health: Evidence from inversion episodes.” *Journal of Health Economics*, 61: 220-232.

Jung, K., K. Bernabe, K. Moors, B. Yan, S. Chillrud, R. Whyatt, D. Carmann, R. Miller (2011). “Effects of floor level and building type on residential levels of uutdoor and indoor PAHs, black carbon and particulate matter.” *Atmosphere* May: 96-109.

Kleibergen, F., and Paap, R. (2006). “Generalized reduced rank tests using the singular value decomposition.” *Journal of Econometrics*, 133(1), 97-126.

Krebs, B., J. Burney, J. Graff Zivin and M. Neidell (2021). “Using crowd-sourced data to assess the temporal and spatial relationship between indoor and outdoor particulate matter.” *Environmental Science and Technology* 55(9): 6107-6115.

Kumar, P., P. Fennell, A. Hayhurst and R. Britter (2009). “Street versus rooftop level concentrations of fine particles in a Cambridge street canyon.” *Boundary Layer Meteorology* 131: 3-18.

Kunn, S. J. Palacios and N. Pestel (2019). “Indoor air quality and cognitive performance.” IZA Discussion Paper No. 12632.

Landrigan, P. J., Fuller, R., Acosta, N. J., Adeyi, O., Arnold, R., Baldoc, A. B., and others (2018). The Lancet Commission on pollution and health. *The Lancet*, 391(10119), 462-512.

Mendell, M., Q. Lei-Gomez, A. Mirer, O. Seppanen and G. Brunner (2008). “Risk factors in heating, ventilating and air-conditioning systems for occupant symptoms in US office buildings.” *Indoor Air* 18(4): 301-16.

Nauze, A. La and E. Severini (2021). “Air pollution and adult cognition: Evidence from brain training.” NBER Working Paper 28785.

Olea, J. L. M., and Pflueger, C. (2013). “A robust test for weak instruments.” *Journal of Business & Economic Statistics*, 31(3), 358-369.

Panczak, R., B. Galobardes, A. Spoerri, M. Zwahlen and M. Egger (2013). “High life in the sky? Mortality by floor of residence in Switzerland.” *European Journal of Epidemiology* 28(6): 453-62.

Park, R. J. (2020). “Hot temperature and high stakes performance.” forthcoming *Journal of Human Resources*.

Pope III, C. A., Cropper, M., Coggins, J., and Cohen, A. (2015). “Health benefits of air pollution abatement policy: Role of the shape of the concentration response function.” *Journal of the Air and Waste Management Association*, 65(5), 516-522.

Roth, S. (2016) “The effect of indoor pollution on cognitive performance: Evidence frm the UK”, mimeograph London School of Economics.

Sager, L. (2019). “Estimating the effect of air pollution on road safety using atmospheric temperature inversions.”*Journal of Environmental Economics and Management*, 98(10): 22-50.

Schaffer, M. (2015). “xtivreg2: State module to perform extended iv/2sls, gmm and ac/hac, limd and k-class regression for panel data models.”

Schwartz, J. (2000). “The distributed lag between air pollution and daily deaths.”*Epidemiology*, 11(3): 320-326.

Stephens, B., P. Azimi and L. Leung, 2019. “How do outdoor pollutant concentrations vary along the height of a tall building?” *Environmental Engineering* 1: 26 - 32.

Stock, J. H., and Yogo, M. (2005). Testing for weak instruments in linear IV regression. Identification and inference for econometric models: Essays in honor of Thomas Rothenberg, 80(4.2), 1.

Thompson, S. (2011). “Simple formulas for standard errors that cluster by both firm and time.” *Journal of Financial Economics* 99(1): 1-10.

United States Code of Federal Regulations. *Probe and Monitoring Path Siting Criteria for Ambient Air Quality Monitoring*. Title 40, Chapter I, Part 58, Appendix E.

Wei, W., O. Ramalho and C. Mandin (2015). “Indoor air quality requirements in green building certifications.” *Building and Environment*.

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Cengage learning.

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Atkinson, R., Le Tertre, A., Bobros, J., Celko, M., Goren, A. and Forsberg, B.(2002). “The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement.”*Epidemiology*, 13(1): 87-93.

Zhang, X., Chen, X., and Zhang, X. (2018). “The impact of exposure to air pollution on cognitive performance.”*Proceedings of the National Academy of Sciences*, 115(37): 9193-9197.

Zheng, S., Wang, J., Sun, C., Zhang, X., and Kahn, M. E. (2019). “Air pollution lowers Chinese urbanites' expressed happiness on social media.”*Nature Human Behaviour*, 3(3): 237-243.

A Additional Tables and Figures

Table A1: Exam Day and Later Pollution on Final Course Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Unrestricted Distributed Leads, Contemporaneous Only Shown							
PM2.5	-1.120***	-0.915***	-0.889***	-0.717***	-0.742***	-0.700***	-0.491***
	(0.132)	(0.134)	(0.134)	(0.136)	(0.136)	(0.136)	(0.137)
Unrestricted Distributed Leads and Lags, Contemporaneous Only Shown							
PM2.5	-1.065***	-1.141***	-1.171***	-1.107***	-1.084***	-1.181***	-0.756***
	(0.137)	(0.143)	(0.144)	(0.153)	(0.153)	(0.156)	(0.162)
Leads or Lags Included	1	2	3	4	5	6	7
Weather Controls	Y	Y	Y	Y	Y	Y	Y
Weather Interactions	Y	Y	Y	Y	Y	Y	Y
Observations	1,806,513	1,806,513	1,806,513	1,806,513	1,806,513	1,806,513	1,806,513
Students	88,959	88,959	88,959	88,959	88,959	88,959	88,959
Exams	1,026	1,026	1,026	1,026	1,026	1,026	1,026

The dependent variable is final course grade, measured in hundredths of a standard deviation. The primary independent variable is exam day $PM_{2.5}$ combined with future $PM_{2.5}$, which serve as a pseudo placebo check. The second panel includes both future and past levels of $PM_{2.5}$. Estimated using ordinary least squares. Every column includes student, semester, day-of-week, and time-of-day fixed effects. Weather controls include daily average temperature, precipitation, relative humidity, pressure, latitudinal wind, and longitudinal wind. Weather interactions include all two-way interactions between the weather variables. Standard errors are clustered at the student and exam-slot level. (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Figure A1: Other Air Pollutants