Pollution at Schools and Children's Aerobic Capacity

Michelle Marcus*
Vanderbilt University and NBER

August 24, 2021

Abstract

Poor respiratory health is a major cause of mortality and morbidity worldwide, and children are especially vulnerable. Existing research in economics has documented the effect of pollution on severe health outcomes, such as hospitalizations for asthma and infant death. However, evidence on the effect of air pollution on less extreme measures of respiratory health is limited, because these effects are difficult to measure. Using a more sensitive measure, aerobic capacity (VO_2max), I study the impact of air pollution on respiratory performance of children. I combine school-grade level data from the California Physical Fitness Test from 2009-2017 with local air pollution and weather data to estimate the impact on student aerobic capacity of fluctuations in air pollution levels on testing days. Ozone affects child aerobic capacity at levels even below the EPA thresholds.

JEL Codes: I14, I18, Q53

*Vanderbilt University, Department of Economics, 415 Calhoun Hall, Nashville, TN 37240. E-mail: michelle.marcus@vanderbilt.edu

Poor respiratory health is a major cause of mortality and morbidity worldwide, with children being especially vulnerable. Air pollution can also have long lasting impacts on children's lung function and growth, exacerbation and development of asthma, oxidative stress and damage, and enhanced respiratory sensitization to aeroallergens (Chen et al., 2015; Guarnieri and Balmes, 2014; Brunst et al., 2015). While the US EPA sets National Ambient Air Quality Standards (NAAQS) with the goal of protecting public health from the harmful effects of air pollution, these regulations are costly to industries. It is important to accurately quantify the associated benefits from pollution reductions, especially for levels of pollution at or below the current thresholds set by the EPA. Although existing research in economics has focused on the effects of air pollution on extreme health outcomes, such as hospitalizations and death (Neidell, 2004; Currie and Neidell, 2005), understanding the effects on less extreme measures can help provide a more complete understanding of the respiratory effects of air pollution.

In general, children are more sensitive to pollution than adults due to their higher respiratory rate, narrower airways, and developing lungs. While child hospitalizations for asthma or another respiratory illness are likely to be a more sensitive measure than adult hospitalizations, it is possible that pollution may have an even more subtle effect on child health. Even if a child is not sick enough to go to the hospital, he or she may experience some respiratory distress or difficulty breathing. Moreover, pollution may impact the aerobic capacity of non-asthmatic children as well. If this is the case, estimates of the impact of pollution on hospitalizations and emergency room visits do not capture the full range of health effects of pollution.

In this paper, I use a more sensitive measure, aerobic capacity (VO_2max), to study the impact of air pollution on respiratory performance at levels below current NAAQS thresholds for three important criteria pollutants: ozone, nitrogen dioxide, and fine-particulate matter. I combine school-grade level data from the California Physical Fitness Test from 2009-2017 with local air pollution and weather data to estimate the impact of fluctuations in test date pollution levels on the aerobic capacity for students in grades 5, 7, and 9.

Estimates measure the respiratory impact of air pollution near schools, which is where children spend the majority of their time, outside of their residence. Although children spend a considerable amount of time at school, our current understanding of the impact of air pollution exposure at schools is incomplete. In fact, epidemiological research suggests that asthma risk increases with traffic-related pollution exposure near both homes and near schools, and that a disproportionate number of economically disadvantaged and nonwhite children attend high-exposure schools in

California (McConnell et al., 2010; Green et al., 2004).

Quantifying the effect of pollution near schools on child respiratory health is complicated by the fact that families with different preferences for clean air may sort across both residential locations and school districts. For example, a family with a high valuation for clean air may choose to live in an area with strict regulations on polluting industries. On the other hand, a family with low valuation for clean air might choose to purchase a house in a heavily polluted industrial area, to take advantage of low property values. In estimating the impact of pollution on respiratory health, it is important to account for sorting behaviors that lead to endogenous pollution exposure. In this paper, regressions include school fixed effects to control for differences in pollution exposure across school locations and to capture any differential characteristics driven by sorting of residents across schools. Identification comes from variation in pollution levels on testing dates at a given school location over time, rather than across locations.

Previous economics literature establishing the link between air pollution and health has exploited natural experiments to avoid the inherent endogeneity problems of cross-sectional comparisons. These studies find important effects of air pollution on extreme health outcomes, such as hospitalizations and emergency room visits (Jans et al., 2018; Schlenker and Walker, 2015; Moretti and Neidell, 2011; Lleras-Muney, 2010; Neidell, 2009, 2004) and infant mortality and birth outcomes (Mccoy and Zhao, 2021; Knittel et al., 2016; Arceo et al., 2016; Currie et al., 2009b; Currie and Neidell, 2005; Chay and Greenstone, 2003b,a). While these outcomes are important and costly, they do not capture more subtle effects of pollution on respiratory performance. In addition to these extreme measures, some work has shown effects of air pollution on adult work performance, productivity, and labor supply (Archsmith et al., 2018; He et al., 2019; Chang et al., 2019, 2016; Graff Zivin and Neidell, 2012; Hanna and Oliva, 2015). Ozone has also been shown to negatively impact collegiate track and field performance (Mullins, 2018). Among children, air pollution has an important impact on school absences (Heissel et al., 2019; Liu and Salvo, 2018; Hales et al., 2016; Currie et al., 2009a; Ransom and Pope III, 1992; Romieu et al., 1992). Other work demonstrates that air pollution, from highways and industrial sources for example, can impact children's test scores and behavioral incidents (Persico and Venator, 2019; Heissel et al., 2019). These findings are important as they document an impact of air pollution on more sensitive outcomes among a broader set of individuals. However, they do not measure child respiratory performance directly.

This work contributes to our understanding of the effects of air pollution by studying a more sensitive measure of respiratory performance, aerobic capacity (VO_2max). Previous research has shown that pollution exposure during exercise can impair lung function and reduce aerobic performance (see Giles and Koehle (2014) for a review of the literature). Most existing work focuses on experimental settings or documents correlations between air pollution and poor aerobic capacity. For example, Ignatius et al. (2004) find that primary school children had significantly lower aerobic capacity in a high pollution district relative to a low pollution district in Hong Kong. Austin et al. (2019) use quasi-experimental variation in diesel emissions from retrofitting school buses to look at the impact on aerobic capacity and test scores, but do not measure air pollution directly. This paper builds upon this existing work to document a causal relationship between daily variation in air pollutants and aerobic capacity in a non-experimental setting with a focus on using fixed effects to account for endogenous residential sorting that can bias cross-sectional comparisons.

Unlike looking at student athletes or using extreme events, such as hospitalizations or death, aerobic capacity provides a measure of respiratory performance for all children in grades 5, 7, and 9 in California, as there is no parent opt-out for the physical fitness test. Although non-asthmatic children would be unlikely to show up in hospitalization data, they may still be negatively affected by air pollution, and aerobic capacity can provide a measure of their respiratory performance. Moreover, aerobic capacity may be useful in detecting effects of pollutants at low levels that might not show up when using measures such as hospitalizations or death.

Beyond contributing to our understanding of the child respiratory impacts of air pollution at low levels, this work expands our knowledge of the importance of exposure to pollution at locations other than the home. Because children spend a large amount of time at school and exposure to pollution at school and home tends to be higher among non-white and economically disadvantaged students, these results are especially important in understanding respiratory health gradients.

I find that exposure to ozone increases the percent of students with unacceptable levels of aerobic capacity at levels of pollution even below the NAAQS threshold. These results are robust to numerous alternate specifications, including adding additional demographic and economic controls, alternate pollution measures, and adding a variety of weather controls. Relative to testing all students when ozone levels are 0-25 percent of the NAAQS threshold, testing all students when ozone is above the NAAQS threshold leads to a 5.4 percentage point (13 percent) increase in the number of students with poor aerobic capacity. I also find harmful effects of testing students when ozone is below the threshold, even as low as 25-50 percent of the threshold. I do not find statistically significant impacts of nitrogen dioxide or fine particulate matter on aerobic capacity, perhaps because these pollutants remain at lower levels more often during the study period, or because ozone tends to reach its peak during the school day, making it easier to detect its effects on respiratory health during the physical fitness test. These findings suggest large potential gains in respiratory performance from a reduction in ozone near schools.

1 Background

Outdoor air pollution almost always occurs as a mixture, which makes it difficult to disentangle the causal impact of individual pollutants (Barnes, 1995; Esposito et al., 2014). In general, we know relatively little about the exact mechanisms through which each pollutant impacts health in isolation. The EPA regulates six criteria air pollutants: ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide, particulate matter, and lead. However, there are also many additional harmful pollutants that are not currently regulated as criteria air pollutants, such as benzene, perchlorethlyene, and methylene chloride. A lack of widespread measurement of these pollutants limits the study of these additional harmful contaminants. This study focuses on the impact of ozone, nitrogen dioxide, and fine particulate matter. These pollutants have been linked to respiratory and health problems and are widely measured during the sample period.

Ozone is not emitted directly into the air, but is a secondary pollutant formed by chemical reaction between nitrogen oxides (NO_x) and volatile organic compounds (VOCs) in the presence of warmth and sunlight. Although ozone is most likely to reach unhealthy levels during warm weather, it can still reach high levels in colder months. Ozone is thought to cause constriction of muscles in the airway, trapping air in the alveoli, which can cause wheezing and shortness of breath. Ozone is associated with many respiratory problems, such as difficulty breathing, inflammation and damage of the airways, increased susceptibility to infection, and chronic obstructive pulmonary disease (COPD). In addition, ozone is likely to aggravate respiratory diseases, such as asthma, emphysema, and chronic bronchitis.

Particulate matter includes all pollution particles under a certain diameter and can be made up of many different toxic materials. Small particles can be inhaled deep into the lungs and may even enter the bloodstream. While previous studies often focused on particulate matter under 10 micrometers in diameter due to data limitations, this study focuses on fine-particulate matter, which includes particles with diameters less than 2.5 micrometers. Fine-particulate matter is considered by many experts to be a better measure of harmful pollutants because smaller particles may more easily enter the lungs and are more likely to contain toxic materials. Studies have linked particulate matter to many health problems, including heart attacks, premature death for those with poor lung function, aggravated asthma, and other respiratory problems.

Nitrogen dioxide is part of a group of highly reactive gasses known as nitrogen oxides (NO_x), which include nitrous acid and nitric acid. Measures of nitrogen dioxide are used as indicators for the presence of the larger group of nitrogen oxides. Inhalation of nitrogen dioxide can irritate airways and cause respiratory distress, aggravate respiratory diseases, and increase hospital and emergency room admissions. Longer exposure may increase the chance of developing asthma or respiratory infections. When nitrogen oxides react with other chemicals in the air, they can form both particulate matter and ozone. These reactions make it particularly important to consider possible correlations across these pollutants and to estimate their effects simultaneously in a multiple-pollutant model, rather than in isolation.

Children are especially vulnerable to air pollution. Not only can air pollution exacerbate preexisting asthma, but it is also associated with new-onset asthma, increased asthma prevalence, risk of bronchitis and wheezing, reduced lung function growth and function, and airway inflammation (Chen et al., 2015; Guarnieri and Balmes, 2014; Brunst et al., 2015). In addition, air pollution can induce oxidative stress and damage, airway remodeling, inflammation and immunological responses, and enhanced respiratory sensitization to aeroallergens (Guarnieri and Balmes, 2014). Children are especially susceptible to these negative health impacts due to the fact that they spend large amounts of time outdoors engaging in physical activity, which increases the amount of particles that can enter the lower airways. Children also tend to have a high breathing rate and are predominantly oral breathers, which increases the intake of pollutants into the respiratory tract (Esposito et al., 2014).

2 Data

Aerobic Capacity Data

The California Physical Fitness Test (PFT), administered in the spring, provides data on students in grades 5, 7, and 9, yearly from 2009 to 2017. Fitness measures are released for each grade in each school, and additional sub-categories report fitness measures by gender, race, ethnicity, and economic status.[1] Although the fitness measures are not at the individual level, additional

[1] The California Department of Education determines a students "economically disadvantaged status based on two fields in the PFT: (1) Parent/Guardian Highest Education Level and (2) Eligibility for the National School Lunch Program (NSLP), which is also known as Free and Reduced Price Meals (FRPM). It is important to note that eligibility for the NSLP does not necessarily mean the student is participating in the NSLP.

information provided by the California Department of Education contains individual dates of test administration for all students in each grade in each school for each year.[2] These test dates are important because they can be linked to pollution and weather information at the daily level to better measure exposure to poor air quality on the actual test dates, rather than using air pollution over the entire testing window. For each grade-school-year, exposure is measured as either the average level of pollution across all test dates or as the fraction of all student test dates that fall within a particular range of air pollution, relative to each pollutant's NAAQS threshold. For example, suppose a given school-grade-year administers the test on two dates t and t+15 with ozone levels x and y, respectively. If the fraction of students taking the test on date t was 80 percent and on date t+15 was 20 percent, then the average ozone exposure calculated for the school-grade-year is 0.8x + 0.2y. These measures are meant to approximate the intensity of exposure to air pollution for the entire grade, since this is the level at which we observe the outcome of interest.

Having exact testing date information for all students in each grade-school-year is especially important since most students within a grade-school-year take the physical fitness test on the same date, as shown in Figure 1. The left panel shows the distribution of percent of students within a grade-school-year who take the PFT on the most common date, presumably the "scheduled test date." On average, about 80 percent take the test on the scheduled date. The right panel shows a histogram of the range of days (i.e. latest date - earliest date) over which students in a grade-school-year take the test. In most cases, all students in the grade take the test within a few days of each other. On average, all students within the grade-school-year take the test within a 17 day range.

PFT data includes the fraction of students in the "healthy fitness zone" (HFZ) with regard to aerobic capacity, body fat, flexibility, abdominal strength, trunk strength, and upper body strength. Table A1 shows the average fraction of students in each grade-school-year cell that are not in the "healthy fitness zone." For this paper, I will focus on the measure of aerobic capacity reported as VO_2max, which is the maximum rate that oxygen can be taken into and used by the body during exercise. The Cooper Institute designs the tests for aerobic capacity and sets the healthy fitness level to capture the range in which health is not at risk. Good aerobic capacity has been shown to reduce the risk of high blood pressure, coronary heart disease, obesity, diabetes, metabolic syndrome, and some forms of cancer.

[2] These data do not include information for charter schools, so these schools are not included. Less than 1 percent of students take the PFT outside of the official testing window, which is between February to May.

During the PFT testing window from February to May, students are administered one of three tests to determine their aerobic capacity: One-Mile Run, 20 meter Progressive Aerobic Cardiovascular Endurance Run (PACER), or Walk Test. For the One-Mile Run, students are instructed to run a mile as fast as possible, which is recorded in minutes and seconds. For the 20 meter PACER, students run a 20 meter distance repeatedly to a specified pace set to music which gets faster each minute. Aerobic capacity is estimated from the number of laps completed. For the Walk Test, students are told to walk one mile as fast as possible and aerobic capacity is measured from both heart rate and walk time. Unfortunately, the data do not contain information on the specific test type used or whether tests are conducted indoors or outdoors.[3] On average, 41.2 percent of students are not in the healthy fitness zone for aerobic capacity.

According to program documentation, the test options of the One-Mile Run and the 20 meter PACER begin with age 10 and the Walk Test begins with age 13. These measures of aerobic capacity are not available for children under the age of ten. Grade five students who are nine years old are scored using the standards for ten year old students. Table A2 shows the most recent standards for aerobic capacity for both females and males by age. Some changes to the standards and the calculation of aerobic capacity occurred in 2011 and 2014. Starting in 2011, only the lower end of the HFZ standards by gender and age was provided, while in previous years the HFZ was presented as a range with the lower number representing the lower end of the HFZ. Prior to 2011, only the percent of students in the HFZ and the percent of students outside the HFZ were reported, so the analysis focuses only on this threshold of poor aerobic capacity. In addition, a student's age, gender, and BMI were needed in addition to PACER or One-Mile Run test results to estimate VO_2Max. In 2014, a new proprietary formula to estimate aerobic capacity was introduced for the PACER test, but the HFZ standards remained the same.[4] Year fixed effects should capture large changes to the standards across years. However, because these fitness standards are gender and age specific, one might be concerned that year fixed effects may not sufficiently capture the shifts aerobic capacity over time driven by changing standards. Therefore, I also show the robustness of the results to

[3] Local education agencies (LEAs) have the local authority in deciding how to meet the requirements of the PFT. Multiple test options are provided for most fitness areas so that all students, including those with disabilities, have the maximum opportunity to participate in the tests. It is not necessary for all students within a school or all schools within a LEA to select or use the same test option for a fitness area. Testing instructions require only the PACER test to be conducted indoors. It is likely that the Mile Run and Walk tests are conducted on a track, such that only schools administering the Run or Walk test that have both indoor and outdoor track facilities would be able to switch the testing location as a margin of response to pollution. Most schools are unlikely to have both indoor and outdoor track facilities. However, to the extent that PFT administrators can respond in this way, the effects of air pollution on aerobic capacity would be understated for higher levels of pollution.

[4] Further documentation of changes in the Fitnessgram Healthy Fitness Zones can be found here: www.cde.ca.gov/ta/tg/pf/healthfitzones.asp

including interactions between year dummies and the percent of students by gender, grade, and race.

Air Pollution Data

Daily data on air pollution comes from the EPA's Air Quality System (AQS) Data Mart through AirData. I obtain daily air quality summary statistics for the criteria pollutants O_3, NO_2, and $PM_{2.5}$ by monitor for the state of California. O_3 is measured as the daily maximum 8-hour concentration (ppm), NO_2 is measured as the daily maximum 1-hour concentration (ppb), and $PM_{2.5}$ is measured as the daily mean concentration (μ g/m^3). I create a weighted average of daily pollution from all monitors within 8 km (about 5 mi) of the school location, using the inverse of distance squared to the school as the weight.[5] I show the results based on alternate distances to pollution monitors and based on the nearest monitor within various distances in the robustness section. A number of schools are excluded from the sample if there are no pollution monitors close enough to the school. Table A3 shows that characteristics of all schools and schools in the analysis sample are similar in terms of aerobic capacity, grade levels, county-level unemployment rates and gender balances. However, the analysis sample has slightly more students per grade, students receiving free/reduced price lunch, and Hispanic students. These differences should be taken into account when interpreting the implications and generalizability of the results.

The EPA thresholds for each of the criteria pollutants are shown in Table A4. The table shows the mean and standard deviation for each pollutant, averaged across test days for each grade-school-year cell. On average, ozone, nitrogen dioxide, and fine particulate matter are 0.04ppm, 31ppb, and $11\mu g/m^3$, respectively. Figure A1 shows the distribution of the mean pollution levels relative to the EPA threshold, which helps illustrate how large a per unit change is relative to the EPA threshold for each pollutant. Because each pollutant is measured in different units and the distribution of each pollutant is on a separate scale, a one unit increase is not comparable across pollutants. First, I address this by scaling each pollutant by its mean and standard deviation, such that a one unit change is equivalent to a standard deviation. In addition, I classify daily pollution levels on test days as within 0-25%, 25-50%, 50-75%, 75-100%, or above 100% of the EPA pollution threshold for each pollutant. As shown in Table A4, about 1.2 percent of testing days in the analysis sample are over the threshold for ozone and about 13 percent are just below the threshold, within 75 to 100

[5] Neidell (2004); Currie and Neidell (2005) also test the validity of inverse distance weighted averages by comparing the actual level of pollution at each monitor location in California with the level of pollution that would be assigned using their method if the monitor in question was not located there. These correlations between actual and predicted levels of pollution were very high (0.77-0.92).

percent of the threshold. For nitrogen dioxide, less than 1 percent of testing days are between 75 and 100 percent of the threshold level, and there are zero testing days in the analysis sample above the threshold for nitrogen dioxide. For fine particulate matter, about 1.4 and 3 percent of testing days are over the threshold and between 75 to 100 percent of the threshold, respectively.[6]

Weather Data

It is important to control for weather given its direct impact on pollution levels. Additionally, California Physical Fitness Test documentation suggests that the aerobic capacity measure could be affected by "extreme weather," and to avoid potential health and safety issues with students and with generating invalid estimates, tests are not to be administered in unusually high temperatures or humidity or when the wind is strong.

Daily temperature and precipitation data for each 2.5 by 2.5 mile square in California is based on the Parameter-elevation Regressions on Independent Slopes Model (PRISM) weather data set, as used in Schlenker and Roberts (2009). For each school, I calculate the inverse distance-squared weighted average of maximum temperature and precipitation values for each day using grid squares with centroids within 8km of the school. Next, I calculate the percent of student tests in the school-grade-year with a daily maximum temperature that falls within each of 7 temperature bins in degrees Celsius: below 0, 0-5, 5-10, 10-15, 15-20, 20-25, and over 25.[7] In addition to these temperature bins, I include controls for the percent of testing days with any precipitation and the average precipitation on testing days for all regressions.

In addition, data on humidity and wind speed are calculated from daily weather data from the National Climatic Data Center's (NCDC) Global Surface Summary of the Day. These data comprise daily averages computed from global hourly station data. Daily absolute humidity, in pounds mass per cubic foot (lbm/ft^3), is calculated from daily dew point and temperature, following standard meteorologic formulas (Parish and Putnam, 1977). For each school, I calculate the inverse distance-squared weighted average of humidity and wind speed for each day using weather

[6] Table A4 restricts the sample to observations with measures of all three pollutants, as in the analysis sample. Table A5 in the appendix shows that the distribution of testing days within each bin is similar for the unrestricted sample.

[7] The results are robust to finer temperature bins, as shown in Table A6 in the appendix. Column 2 includes finer bins at high temperatures by replacing the highest pollution bin (over 25 C) into four bins: 25-27, 27-29, 29-31, and over 31 C. Column 3 includes 34 bins of temperature where there is one bin for temperature below zero, one bin for each degree Celsius between 0 and 33 degrees, and one bin for temperatures above 33 degrees Celsius. The results are also robust to controlling for sunlight. Column 4 includes average sunlight on testing days based daily sunlight measured at the population weighted county centroid from the National Solar Radiation Database (NSRDB). The results are very similar across all four columns.

monitors within 20km of the school.[8] All regressions include controls for the average humidity and average wind speed on testing days.

Other Data

Additional control variables include employment rates, school-grade enrollment demographics, and school-grade free/reduced price lunch.

Employment rates at the county-month level come from the Bureau of Labor Statistics' Local Area Unemployment Statistics (LAUS) data. I use the mean employment rate during the testing month and link to schools by county. This will help control for local economic activity that could be related to pollution and child outcomes.

School enrollment and free/reduced price lunch information comes from the California Department of Education's school data files. From the enrollment data, I calculate the percent of students by race and gender in each school-grade for each year. From the free and reduced price lunch data, I calculate the percent of students receiving free or reduced price lunch in each school for each year. These controls help account for changes in respiratory capacity due to changing student body characteristics.

3 Methods

Previous epidemiological research suggests that a disproportionate number of economically disadvantaged and nonwhite children attend schools with high exposure to air pollution in California (Green et al., 2004). However, residential location is endogenously determined, as families may sort into high exposure schools based on preferences. To overcome this challenge, the empirical strategy used here exploits variation in pollution over time at different school locations in California. School fixed effects will capture time-invariant characteristics of schools, such as average income of families, school facilities, and location in relation to highways and other stationary sources of pollution.

First, I estimate the following single-pollutant model,

$$LowVO2_{gst} = P_{gst} + \omega_{gst} + \gamma_g + \delta_t + \sigma_s + \epsilon_{gst} \tag{1}$$

[8] Although the PRISM data includes gridded information across all of California, weather monitors are more dispersed, necessitating a larger radius in order to maintain a sufficient sample size.

where g indexes grade, s indexes school, and t indexes school-year. P represents either the level or z-score of three criteria pollutants: O_3, NO_2, and $PM_{2.5}$. The single-pollutant model is estimated separately for each of the three pollutants. The unit of observation is at the grade-school-year. $LowVO2_{gst}$ is the percentage of the grade that is outside of the "Healthy Fitness Zone" in grade g in school s and in year t. The regression is weighted by the number of students in each grade. State-wide trends are captured non-parametrically with year dummies, δ_t. School fixed effects are included as σ_s, and γ_g is a set of grade dummies. ω_{gst} is a set of weather controls for the percent of test days within 7 temperature bins, the percent of days with any precipitation, and the average precipitation, humidity, and wind speed on test days. Standard errors are clustered at the school level.

In order to allow for non-linear effects of pollution and to estimate the impact of pollution levels below the EPA threshold, I also estimate the following specification, which includes the percentage of testing days for a grade-school-year that are 25-50%, 50-75%, 75-100%, or above 100% of the EPA pollution threshold for each pollutant,

$$LowVO2_{gst} = P^{25-50}_{gst} + P^{50-75}_{gst} + P^{75-100}_{gst} + P^{100+}_{gst} + \omega_{gst} + \gamma_g + \delta_t + \sigma_s + \epsilon_{gst} \tag{2}$$

where P^{25-50}_{gst} is equal to the fraction of testing days that fall within 25 to 50 percent of the EPA threshold level of pollution for pollutant P. P^{50-75}_{gst}, P^{75-100}_{gst}, and P^{100+}_{gst} are defined analogously. The omitted category for each pollutant is 0 to 25 percent of the EPA threshold.

In addition to these single-pollutant models, which study the effect of each pollutant individually, I also estimate models that includes all three pollutants simultaneously. As criteria pollutants often co-vary with each other, this specification is especially important. For example, since ozone is formed primarily through a photochemical reaction between volatile organic compounds and nitrogen oxides, it may be that ozone and nitrogen dioxides are negatively correlated. Therefore, I estimate the following multi-pollutant models,

$$LowVO2_{gst} = \sum_{P=1}^{3} [P_{gst}] + \omega_{gst} + \gamma_g + \delta_t + \sigma_s + \epsilon_{gst} \tag{3}$$

$$LowVO2_{gst} = \sum_{P=1}^{3} [P^{25-50}_{gst} + P^{50-75}_{gst} + P^{75-100}_{gst} + P^{100+}_{gst}] + \omega_{gst} + \gamma_g + \delta_t + \sigma_s + \epsilon_{gst} \tag{4}$$

where P represents one of three criteria pollutants. Pollution is measured as either the level or z-score in equation 3 and is measured in bins relative to the EPA threshold in equation 4. Controls are defined analogously to equation 1. The regression is weighted by the number of students in each grade-school-year and standard errors are clustered at the school level.

To explore possible heterogeneous effects of pollution exposure by gender, income, race and ethnicity, I exploit additional data on aerobic capacity by demographic groups. For this analysis, I estimate equations 2 and 4, where the outcome variables are measured as the percentage of the demographic group in the grade that is outside of the "Healthy Fitness Zone." These regressions are weighted by the number of students in the relevant demographic category for each grade-school-year cell.

4 Results

Columns 1-3 of Table 1 show the results from estimating equations 1 and 2 for ozone, nitrogen dioxide, and particulate matter, respectively. Panels A and B show the results from estimating equation 1 using the level and z-score of each pollutant, respectively. Note that nitrogen dioxide and fine particulate matter are scaled by 100 in Panel A to keep coefficients readable. Panel C shows the results from estimating equation 2 using bins of each pollutant relative to the EPA threshold level. For all columns, the outcome variable is the fraction of the grade that is not in the healthy fitness zone (HFZ) for aerobic capacity. This is the fraction of the grade that has an unacceptable level of aerobic capacity, as defined in Table A2. All regressions include school fixed effects, grade dummies, and weather controls. Across all three panels, the single pollutant models show that only ozone appears to have a statistically significant impact on aerobic capacity. However, since these criteria pollutants are possibly correlated, especially nitrogen dioxide and ozone, it is important to estimate the multiple-pollutant model shown in equations 3 and 4.

Column 4 shows the multiple-pollutant models, which also include school fixed effects, grade dummies, and weather controls.[9] Across all specifications, the results are very consistent. Increases in ozone are associated with increases in the fraction of children testing outside the healthy fitness zone for aerobic capacity. Focusing on column 4, panel A estimates that a 0.1ppm increase in ozone is associated with a 7 percentage point increase in students with poor aerobic capacity. Similarly, panel B estimates a 1 standard deviation increase in ozone, approximately 0.01ppm, is associated

[9]There are 5,788 observations at the grade-school-year level for columns 1-4. This is comprised of 804 schools, but not all schools are observed each year and not all schools have 5th, 7th, and 9th grade students. For example, some schools may serve only high school students and report information for only 9th grade. In fact, most schools in the sample (86%) contain only one grade. About 10% contain two grades and about 4% contain all three grades.

with a 0.7 percentage point increase in students with poor aerobic capacity.

In panel C, the fraction of testing days with ozone between 25-50, 50-75, 75-100, and over 100 percent of the threshold level are associated with a higher fraction of children who are outside of the HFZ for aerobic capacity. Point estimates are slightly smaller in magnitude for lower levels of pollution, although not statistically significantly smaller. The estimates suggest that if ozone is above the EPA threshold for all students on the test day, the percent of students with poor aerobic capacity is 5.4 percentage points higher than if ozone had been 0 to 25 percent of the threshold. Relative to a mean of 41.2 percent, this represents a 13 percent increase in the percent of students with poor aerobic capacity. Importantly, ozone also has an effect at levels below the EPA threshold. For example, if all students experience ozone levels that are 25 to 50 percent of the EPA threshold on the test date, the percent of students with poor aerobic capacity is 4.2 percentage points higher than on a low ozone day. Interestingly, the magnitude of this effect not statistically significantly different across each bin above the lowest level. It is important to note that the outcome is a binary measure of the fraction of students outside the healthy fitness zone, and therefore can only measure the fraction of students who cross this threshold, rather than the intensity of the impact on aerobic capacity for those students. For example, elevating ozone above 25 percent of the EPA threshold may induce 4-5 percent of students to cross the threshold into "poor" aerobic capacity, and these same students may experience worsening respiratory conditions as ozone is elevated to higher levels. However, the binary measure of poor aerobic capacity used here will only capture the fraction of students with poor aerobic capacity and will not capture these worsening impacts.

For the other pollutants, I am unable to detect an impact on respiratory performance. For NO_2, this could be due to the fact that there is a relatively small fraction of testing days at higher levels of pollution. As seen in Table A4, less than 1 percent of days are over 75 percent of the threshold for NO_2. This limited variation may make it difficult to detect effects for NO_2. In addition, differences in the peak levels of each pollutant throughout the day may make it easier to detect effects on ozone relative to particulate matter and nitrogen dioxide. While O_3 is elevated throughout the school day and typically reaches its peak around 2-3:00 p.m., at that time NO_2 and $PM_{2.5}$ are typically at a low point.[10] The typical school day lasts from around 8:00 a.m. to about 3:00 p.m., depending on the district. Elevated levels of ozone throughout the school day are likely to make it easier to detect its effect on respiratory health during the physical fitness test, as

[10]Figure A2 in the appendix shows the hourly concentration of each pollutant for all monitors in California in 2017. Ozone peaks around 2-3:00 p.m., nitrogen dioxide peaks around 6:00 a.m. and 8:00 p.m., and fine particulate matter peaks around 8:00 p.m., with a smaller peak around 8:00 a.m.

compared to particulate matter and nitrogen dioxide.

Next, Table 2 explores heterogeneous effects by gender, ethnicity, and income. The sample is restricted to a consistent set of schools with non-missing information for all subgroups.[11] All regressions include school fixed effects, grade dummies, and weather controls. Because fine particulate matter and nitrogen dioxide were not found to have a significant impact on respiratory capacity, the following tables report only coefficients for ozone. The first column shows the results for all students for the set of schools with non-missing subgroup data. Panel A reports the results from the single pollutant model, while panel B reports results from the multi-pollutant model. The overall effects in panels A and B are very similar to the results from columns 1 and 4 in panel C of Table 1. The coefficient estimates across both panels are very similar in magnitude, suggesting the results are robust to controlling for nitrogen dioxide and particulate matter. Test results are reported separately for each of the demographic categories shown in Table 2 for each grade in each school in each year. In these sub-population analyses, grade-school-year observations are weighted by the number of students in the relevant demographic category.

Columns 2 and 3 show the results for females and males, respectively. It is helpful to consider each gender separately as the healthy fitness zone standards differ specifically by gender. The estimates for ozone are statistically significant for both genders, with slightly larger point estimates for females. Columns 4 and 5 show the effects for Hispanic and economically disadvantaged students, respectively. The results for ozone are large and statistically significant for both groups students. Focusing on panel A, if all Hispanic students take the test on a day with ozone above the EPA threshold, poor aerobic capacity increases by 5.4 percentage points, or 12 percent from the mean for Hispanic students. For economically disadvantaged students taking the test on a day with ozone above the EPA threshold, there is a 5.3 percentage point, or 12 percent, increase in poor aerobic capacity. Across all these subgroups, the effects of ozone are significant and similar in magnitude.

[11] Some schools do not report results for each demographic category, either by choice or because there are 10 or fewer identified students in the grade that fall within the category. Both schools and students can choose not to report any fields related to demographics. To protect the confidentiality of students, the California Department of Education suppresses PFT scores when the number of students tested is 10 or fewer. Table A7 in the appendix shows the results for the unrestricted sample of schools, which are very similar. A smaller number of schools report additional subgroup data, but these data are not consistently reported across the majority of schools in the sample, making it difficult to interpret these results.

5 Robustness

First, I begin by exploring the possibility that there may be avoidance through school absences. In general, parents and students are informed of test dates in advance. However, the information made available to parents/students and the policies for absences on the test day may vary from one local educational agency to another. The estimates of the impact of air pollution on aerobic capacity may be underestimated if students with greater susceptibility to air pollution, such as asthmatic children, miss school on the test day when air pollution is elevated.[12] Using information on the individual dates of test administration for each student in each grade in each school for each year, I am able to explore avoidance of testing on high pollution days directly in Table 3. I observe the percent of students in a grade-school-year cell that take the PFT test on the most common test date among all students (or the "scheduled" test date) as well as the percent of students taking the test before or after the scheduled date. If students avoid attending school when the physical fitness test is conducted on a high pollution day, we would expect high pollution on the scheduled test date to increase the percent of students taking the test after the scheduled date.

The outcomes in columns 1-3 of Table 3 are the percent of students taking the PFT on the scheduled date, the percent taking the test later than the scheduled date, and the percent taking the test earlier than the schedule date, respectively. For all regressions, pollution is measured as pollution on the scheduled test date, and each regression includes school fixed effects, indicators for the year-month of the scheduled test date, grade dummies, and weather controls for the scheduled test date. Column 1 shows there is no evidence that higher pollution decreases the percent of students taking the test on the schedule date. Columns 2 and 3 further support this claim by showing there is no significant relationship between higher pollution and the percent of students taking the test late or early. It is possible, however, that avoidance may be more prevalent among certain demographic groups. Table A8 in the appendix shows there are no differential effects by race, gender, economic status, or body composition in the percent of students who test late when the scheduled test date is on a high ozone day. These results provide some evidence that selective absences on test dates are not substantially biasing the main results.

Next, Table 4 tests the robustness of the main results to alternate specifications with the main results replicated in column 1 for comparison. Given that fitness standards and test administration both differ across grade, school fixed effects may not be sufficient. Column 2 includes school-grade

[12]Note that these school absences could be due to either avoidance behavior or poor student health.

specific fixed effects and the results are almost identical to the first column.

Second, it is possible that underlying demographic characteristics of the student body change across years and these changes may be correlated with aerobic capacity. Therefore, column 3 includes demographic controls for the percent of the grade by gender and race/ethnicity. The main results are robust to these controls. In addition, it is important to note that the fitness standards were updated in 2011 and 2014, with the largest changes occurring in 2011. As these changes differentially impact students by gender and age, year fixed effects may not sufficiently capture the shifts in aerobic capacity over time driven by changing standards. To account for the differential impacts by gender and age, column 4 includes year-by-gender, year-by-grade, and year-by-race controls, and the results are consistent with the main findings.

Column 5 addresses the concern that changing economic conditions may be correlated both with air pollution and child aerobic capacity. Column 5 includes the county-level unemployment rate and the percent of students eligible for free or reduced price school lunch as additional controls. The results are robust to these controls.

Next, columns 6 and 7 consider alternate choices of weights from the main results which weight by number of students in each cell. Column 6 shows the results without any weights. Column 7 weights the results by inverse distance to the nearest pollution monitor to account for the fact that there will be greater measurement error for pollution if the nearest pollution monitor is far away from the school. The coefficients on ozone remain statistically significant for both columns 6 and 7, with larger coefficient estimates when weighting by inverse distance to the nearest pollution monitor.

Column 8 shows the results when the sample is limited to test dates within the official testing window in which schools are supposed to administer the test (February to May). Over 99 percent of tests are administered within the window, and the results are very similar to the baseline specification in column 1. Finally, one may be concerned that newly placed monitors may underestimate pollution hotspots in disadvantaged neighborhoods in attainment counties due to endogenous monitor placement (Grainger and Schreiber, 2019). Column 9 addresses the concern by excluding any new monitors introduced during the sample period. Across both specifications, the results are very similar and continue to suggest that aerobic capacity is lower for levels of ozone even below the EPA's threshold levels.

Table A9 shows the main results for pollution measured at alternate distances. Columns 1 to 4 show the effect of inverse distance squared weighted pollution from monitors within 3km, 5km, 8km, and 10km of the school, averaged over all testing dates. The main specification uses 8km, or approximately 5 miles, which is replicated in column 3. The results based on 10km in column 4 are very similar to the main results, although the estimates in panel C are slightly smaller in magnitude. For narrower distance bands shown in columns 1 and 2, the point estimates in panel C are larger in magnitude but imprecisely measured. Overall, estimates based only on monitors within 3 or 5km are less precise and rely on much smaller samples. Table A10 shows these results are very similar when based on the nearest monitor within these distance bands as well.

Next, one potential threat to identification would be if changes in pollution are correlated with changes demographic characteristics within schools. For example, if a particular area decided to reduce pollution due to new political leadership or a sudden influx of wealthy families, and the resulting change in pollution accompanied changes in the demographics of the student body. To test if this is driving the results, Table 5 repeats the main specification but with school-grade demographic characteristics as the outcome variables, including the percent receiving free or reduced price lunch and the percent in each race group. Almost all of the coefficients on ozone are statistically insignificant and small in magnitude, and there is no strong pattern of results. This indicates that the main findings are not driven by demographic changes in the underlying student population over time.

Finally, Table 6 shows the effect of air pollution on other physical fitness test measures. Some of these fitness tests can be thought of as placebos, such as body composition and flexibility, since they should not be impacted by air pollution. However, there may be some impact on other measures of fitness, such as abdominal strength, trunk strength, and upper body strength. Column 1 replicates the main results for aerobic capacity, showing significant impacts of ozone exposure. Columns 2-6 show the effect of air pollution on other physical fitness measures from the PFT: body composition, flexibility, abdominal strength, trunk extensor strength, and upper body strength. Each regression shows the effect of pollution at various levels on the fraction of students not in the healthy fitness zone for each test. Columns 2 and 3 show there is no significant impact of ozone on body composition or flexibility, as expected. This is reassuring since we would not expect higher ozone on the testing date to impact body composition or flexibility. Columns 4, 5, and 6, present the results for each of the three strength measures. Columns 4 and 6 shows that there may be some negative impact of ozone on testing days for abdominal strength and upper body strength, but there is no evidence of a significant impact on trunk extensor strength in column 5.

6 Discussion & Conclusion

Results shown above indicate that important air pollutants are harmful to child health at levels even below the EPA's thresholds. Using a more sensitive measure of respiratory health, aerobic capacity (VO_2max), not only does ozone above the EPA threshold increase the fraction of students with poor aerobic capacity, but so does ozone at lower levels, even as low as 25-50 percent of the threshold. The magnitude of the effects suggests a nonlinear impact on the binary measure of poor aerobic capacity with the steepest slope between ozone below 25 percent and ozone 25 to 50 percent of the threshold.

These results are shown to be robust to many alternate specifications, such as school-grade fixed effects, including additional demographic and economic controls, and alternate pollution measures. In addition, there is no evidence of avoidance through absences on high pollution test days, and there is no evidence of changing demographic characteristics that correspond to pollution changes. Importantly, the effects of ozone are driven through changes in aerobic capacity rather than other test areas that can be considered placebos, such as body composition or flexibility. These findings are reassuring and suggest that the main results operate through the effect of air pollution on respiratory capacity, rather than through another channel.

Not only does pollution have a large impact on overall respiratory performance, but these effects may be exacerbated by educational consequences in some districts as well. Failure to achieve a healthy fitness level in 5 of the 6 test categories for students in the Los Angeles Unified School District, for example, means these students are required to continue taking physical education courses until they pass or graduate.[13] Similar to findings of Ebenstein et al. (2016) who show that pollution exposure during high-stakes exams has long-run effects on educational attainment and earnings, it is possible that poor air quality on the fitness test date may have an impact on pass rates and may preclude some students from taking other courses needed for graduation. As disadvantaged students tend to have higher exposure to poor air quality overall, poor air quality on the test date could widen existing educational gaps and have other important long-run consequences.

[13] In 2017, Los Angeles Unified had a lower graduation rate, 76 percent, and a higher percent of students with poor aerobic capacity, 47 percent, as compared to the state average of 83 percent and 41 percent, respectively. Only about 26 percent of ninth grade students in Los Angeles Unified schools passed 5 of 6 test components, meaning the majority of students, 74 percent, were required to retake physical education classes until they pass or graduate.

References

Arceo, Eva, Rema Hanna, and Paulina Oliva, "Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City," *The Economic Journal*, 2016, *126* (591), 257–280.

Archsmith, James, Anthony Heyes, and Soodeh Saberian, "Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation," *Journal of the Association of Environmental and Resource Economists*, 2018, *5* (4), 827–863.

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

Barnes, Peter J, "Air pollutionand asthma: molecular mechanisms," *Molecular medicine today*, 1995, *1* (3), 149–155.

Brunst, Kelly J, Patrick H Ryan, Cole Brokamp, David Bernstein, Tiina Reponen, James Lockey, Gurjit K Khurana Hershey, Linda Levin, Sergey A Grinshpun, and Grace LeMasters, "Timing and duration of traffic-related air pollution exposure and the risk for childhood wheeze and asthma," *American journal of respiratory and critical care medicine*, 2015, *192* (4), 421–427.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell, "Particulate pollution and the productivity of pear packers," *American Economic Journal: Economic Policy*, 2016, *8* (3), 141–69.

Chang, Tom Y, Joshua Graff Zivin, Tal Gross, and Matthew Neidell, "The effect of pollution on worker productivity: evidence from call center workers in China," *American Economic Journal: Applied Economics*, 2019, *11* (1), 151–72.

Chay, Kenneth and Michael Greenstone, "Air quality, infant mortality, and the Clean Air Act of 1970," *No. w10053. National Bureau of Economic Rsearch*, 2003.

Chay, Kenneth Y and Michael Greenstone, "The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession," *The quarterly journal of economics*, 2003, *118* (3), 1121–1167.

Chen, Zhanghua, Muhammad T Salam, Sandrah P Eckel, Carrie V Breton, and Frank D Gilliland, "Chronic effects of air pollution on respiratory health in Southern California children: findings from the Southern California Childrens Health Study," *Journal of thoracic disease*, 2015, *7* (1), 46.

Currie, Janet and Matthew Neidell, "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?," *The Quarterly Journal of Economics*, 2005, *120* (3), 1003–1030.

_ , Eric A Hanushek, E Megan Kahn, Matthew Neidell, and Steven G Rivkin, "Does pollution increase school absences?," *The Review of Economics and Statistics*, 2009, *91* (4), 682–694.

_ , Matthew Neidell, and Johannes F Schmieder, "Air pollution and infant health: Lessons from New Jersey," *Journal of health economics*, 2009, *28* (3), 688–703.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth, "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution," *American Economic Journal: Applied Economics*, 2016, *8* (4), 36–65.

Esposito, Susanna, Rossana Tenconi, Mara Lelii, Valentina Preti, Erica Nazzari, Silvia Consolo, and Maria Francesca Patria, "Possible molecular mechanisms linking air pollution and asthma in children," *BMC pulmonary medicine*, 2014, *14* (1), 31.

Giles, Luisa V and Michael S Koehle, "The health effects of exercising in air pollution," *Sports Medicine*, 2014, *44* (2), 223–249.

Grainger, Corbett and Andrew Schreiber, "Discrimination in Ambient Air Pollution Monitoring?," in "AEA Papers and Proceedings," Vol. 109 2019, pp. 277–82.

Green, Rochelle S, Svetlana Smorodinsky, Janice J Kim, Robert McLaughlin, and Bart Ostro, "Proximity of California public schools to busy roads.," *Environmental Health Perspectives*, 2004, *112* (1), 61.

Guarnieri, Michael and John R Balmes, "Outdoor air pollution and asthma," *The Lancet*, 2014, *383* (9928), 1581–1592.

Hales, Nicholas M, Caleb C Barton, Michael R Ransom, Ryan T Allen, and C Arden Pope III, "A quasi-experimental analysis of elementary school absences and fine particulate air pollution," *Medicine*, 2016, *95* (9).

Hanna, Rema and Paulina Oliva, "The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City," *Journal of Public Economics*, 2015, *122*, 68–79.

He, Jiaxiu, Haoming Liu, and Alberto Salvo, "Severe air pollution and labor productivity: Evidence from industrial towns in China," *American Economic Journal: Applied Economics*, 2019, *11* (1), 173–201.

Heissel, Jennifer, Claudia Persico, and David Simon, "Does pollution drive achievement? The effect of traffic pollution on academic performance," Technical Report, National Bureau of Economic Research 2019.

Ignatius, TS, Tze Wai Wong, and Hong Jie Liu, "Impact of air pollution on cardiopulmonary fitness in schoolchildren," *Journal of occupational and environmental medicine*, 2004, *46* (9), 946–952.

Jans, Jenny, Per Johansson, and J Peter Nilsson, "Economic status, air quality, and child health: Evidence from inversion episodes," *Journal of health economics*, 2018, *61*, 220–232.

Knittel, Christopher R, Douglas L Miller, and Nicholas J Sanders, "Caution, drivers! Children present: Traffic, pollution, and infant health," *Review of Economics and Statistics*, 2016, *98* (2), 350–366.

Liu, Haoming and Alberto Salvo, "Severe air pollution and child absences when schools and parents respond," *Journal of Environmental Economics and Management*, 2018, *92*, 300–330.

Lleras-Muney, Adriana, "The needs of the army using compulsory relocation in the military to estimate the effect of air pollutants on childrens health," *Journal of Human Resources*, 2010, *45* (3), 549–590.

McConnell, Rob, Talat Islam, Ketan Shankardass, Michael Jerrett, Fred Lurmann, Frank Gilliland, Jim Gauderman et al., "Childhood incident asthma and traffic-related air pollution at home and school," *Environmental Health Perspectives*, 2010, *118* (7), 1021.

Mccoy, Shawn J and Xiaoxi Zhao, "Wildfire and infant health: a geospatial approach to estimating the health impacts of wildfire smoke exposure," *Applied Economics Letters*, 2021, *28* (1), 32–37.

Moretti, Enrico and Matthew Neidell, "Pollution, health, and avoidance behavior evidence from the ports of Los Angeles," *Journal of human Resources*, 2011, *46* (1), 154–175.

Mullins, Jamie T, "Ambient air pollution and human performance: Contemporaneous and acclimatization effects of ozone exposure on athletic performance," *Health economics*, 2018, *27* (8), 1189–1200.

Neidell, Matthew, "Information, Avoidance behavior, and health the effect of ozone on asthma hospitalizations," *Journal of Human Resources*, 2009, *44* (2), 450–478.

Neidell, Matthew J, "Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma," *Journal of health economics*, 2004, *23* (6), 1209–1236.

Parish, O Owen and Terrill W Putnam, "Equations for the determination of humidity from dewpoint and psychrometric data," Technical Report, National Aeronautics and Space Administration (NASA) 1977.

Persico, Claudia L and Joanna Venator, "The Effects of Local Industrial Pollution on Students and Schools," *Journal of Human Resources*, 2019, pp. 0518–9511R2.

Ransom, Michael R and C Arden Pope III, "Elementary school absences and PM10 pollution in Utah Valley," *Environmental research*, 1992, *58* (1-2), 204–219.

Romieu, Isabelle, Marlene Cortes Lugo, Silvia Ruiz Velasco, Sergio Sanchez, Fermando Meneses, and Mauricio Hemandez, "Air pollution and school absenteeism among children in Mexico City," *American Journal of Epidemiology*, 1992, *136* (12), 1524–1531.

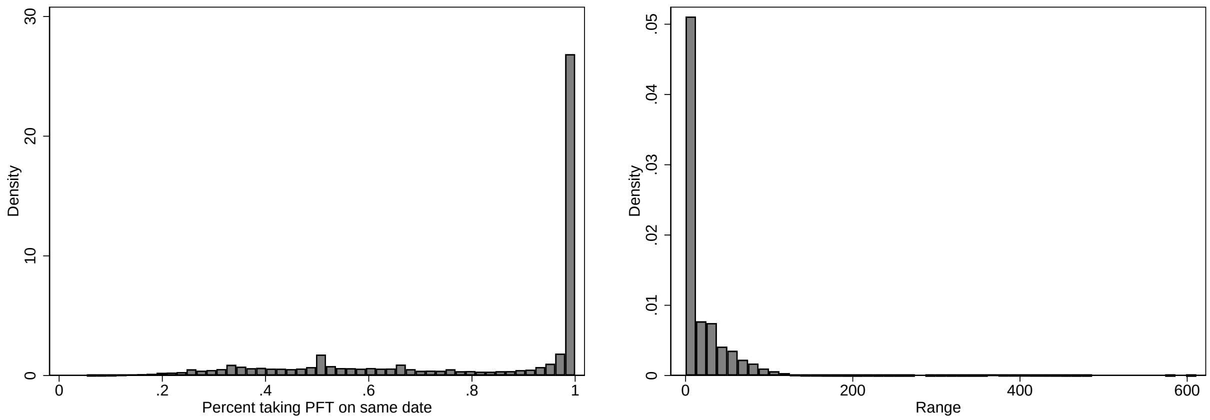
Schlenker, Wolfram and Michael J Roberts, "Nonlinear temperature effects indicate severe damages to US crop yields under climate change," *Proceedings of the National Academy of sciences*, 2009, *106* (37), 15594–15598.

_ **and W Reed Walker**, "Airports, air pollution, and contemporaneous health," *The Review of Economic Studies*, 2015, *83* (2), 768–809.

Zivin, Joshua Graff and Matthew Neidell, "The impact of pollution on worker productivity," *American Economic Review*, 2012, *102* (7), 3652–73.

7 Figures

Figure 1: Testing dates within each grade-school-year



Notes: The left panel shows a histogram of the percent of students within a grade-school-year who take the physical fitness test on the most common date for that grade-school-year. Most students in a grade take the test on the same date. On average 80 percent of students within a grade-school-year take the test on the same date. The right panel shows a histogram of the range of days (i.e. latest date - earliest date) over which students in a grade-school-year take the test. In most cases, all students in the grade take the test within a few days of each other. The average range is 17 days between the earliest and latest test date.

8 Tables

Table 1: Effect of pollution on poor aerobic capacity

	(1)	(2)	(3)	(4)
Panel A.				
O3	0.680**			0.674**
	(0.296)			(0.299)
NO2		0.00966		0.00666
		(0.0231)		(0.0236)
PM2.5			0.0196	0.0103
			(0.0395)	(0.0408)
Panel B.				
zO3	0.00682**			0.00676**
	(0.00297)			(0.00300)
zNO2		0.00129		0.000889
		(0.00308)		(0.00314)
zPM2.5			0.00125	0.000655
			(0.00252)	(0.00260)
Panel C.				
O3 25-50%	0.0433**			0.0423**
	(0.0183)			(0.0178)
O3 50-75%	0.0390**			0.0379**
	(0.0187)			(0.0183)
O3 75-100%	0.0536***			0.0526***
	(0.0206)			(0.0202)
O3 over 100%	0.0548**			0.0543**
	(0.0266)			(0.0267)
NO2 25-50%		0.00329		0.00171
		(0.00614)		(0.00638)
NO2 50-75%		0.00936		0.00753
		(0.00898)		(0.00938)
NO2 75-100%		0.00427		0.00601
		(0.0571)		(0.0579)
PM2.5 25-50%			-0.00371	-0.00440
			(0.00561)	(0.00583)
PM2.5 50-75%			0.000130	-0.00125
			(0.00777)	(0.00798)
PM2.5 75-100%			0.00350	0.000730
			(0.0123)	(0.0124)
PM2.5 over 100%			0.0119	0.00985
			(0.0193)	(0.0194)
Observations	5,788	5,788	5,788	5,788
Schools	804	804	804	804
R-squared	0.617	0.616	0.616	0.617

Notes: The outcome is the fraction of students outside the "healthy fitness zone" (HFZ). All regressions include school fixed effects, year fixed effects, grade dummies, and weather controls. Weather controls include percent of test days that fall into 7 temperature bins, percent of test days with any precipitation, and average precipitation, humidity, and wind speed on test days. Results are shown for for ozone (O3), nitrogen dioxide (NO2), and fine particulate matter (PM2.5). Panel A shows the level of each pollutant. Nitrogen dioxide and fine particulate matter are scaled by 100 in Panel A to keep coefficients readable. Panel B shows the results for each pollutant after standardizing by the mean and standard deviation. Panel C shows the results for bins of pollution relative to the EPA threshold, where the omitted category for each pollutant is 0-25% of the EPA threshold. All regressions are weighted by number of students. Standard errors clustered at the school level are in parentheses.

Table 2: Heterogeneous effects: Gender, race/ethnicity, and income

	All (1)	Female (2)	Male (3)	Hispanic (4)	Econ Disadv. (5)
Panel A. Single Pollutant Model					
O3 25-50%	0.0429**	0.0519**	0.0331*	0.0443**	0.0445**
	(0.0186)	(0.0222)	(0.0190)	(0.0192)	(0.0206)
O3 50-75%	0.0387**	0.0485**	0.0282	0.0431**	0.0368*
	(0.0191)	(0.0229)	(0.0194)	(0.0198)	(0.0212)
O3 75-100%	0.0536**	0.0623**	0.0443**	0.0583***	0.0406*
	(0.0210)	(0.0250)	(0.0212)	(0.0218)	(0.0228)
O3 over 100%	0.0516*	0.0531	0.0467*	0.0535*	0.0525*
	(0.0276)	(0.0336)	(0.0260)	(0.0320)	(0.0299)
Panel B. Multi-pollutant Model					
O3 25-50%	0.0416**	0.0511**	0.0314*	0.0443**	0.0423**
	(0.0181)	(0.0213)	(0.0189)	(0.0186)	(0.0206)
O3 50-75%	0.0373**	0.0475**	0.0263	0.0428**	0.0340
	(0.0186)	(0.0219)	(0.0193)	(0.0192)	(0.0214)
O3 75-100%	0.0524**	0.0616**	0.0424**	0.0574***	0.0370
	(0.0206)	(0.0243)	(0.0211)	(0.0214)	(0.0233)
O3 over 100%	0.0507*	0.0534	0.0446*	0.0530	0.0511
	(0.0277)	(0.0334)	(0.0261)	(0.0325)	(0.0310)
Observations	5,470	5,334	5,382	5,366	4,294
Schools	740	740	740	740	740
R-squared	0.608	0.598	0.586	0.554	0.625

Notes: For all regressions, the outcome is the fraction of students in a particular demographic group that are outside the "healthy fitness zone." The sample is restricted to a consistent set of schools with non-missing information for all subgroups. All regressions include school fixed effects, grade dummies, and weather controls. Weather controls include percent of test days that fall into 7 temperature bins, percent of test days with any precipitation, and average precipitation, humidity, and wind speed on test days. Panel A shows results for ozone from the single pollutant model in equation 2. Panel B shows results for ozone (O3) from the multi-pollutant model in equation 4, which includes controls for nitrogen dioxide (NO2), and fine particulate matter (PM2.5). For each pollutant, the omitted category is 0-25% of the EPA threshold. Regressions are weighted by number of students in the relevant demographic category for each grade-school-year cell. Standard errors clustered at the school level are in parentheses.

Table 3: Avoidance behavior: Testing before or after the scheduled test date

	% On Schedule	% Test Late	% Test Early
	(1)	(2)	(3)
O3 25-50%	-0.000700	-0.00459	0.00831
	(0.0342)	(0.0293)	(0.0239)
O3 50-75%	0.00647	-0.0241	0.0210
	(0.0345)	(0.0307)	(0.0240)
O3 75-100%	0.0158	-0.0468	0.0365
	(0.0387)	(0.0321)	(0.0280)
O3 over 100%	0.000623	-0.0655	0.0684
	(0.0601)	(0.0654)	(0.0681)
Observations	4,594	4,588	4,588
Schools	742	741	741
R-squared	0.575	0.454	0.387

Notes: For all regressions, pollution is measured as pollution on the most common test date (i.e. the "scheduled" date) across all students in a grade-school-year cell. The outcome in column 1 is the percent of students that take the test on the most common test date. The outcomes in columns 2 and 3 are the percent of students that take the test later and earlier than the most common test date, respectively. All regressions include school fixed effects, indicators for the year-month of the "scheduled" test, grade dummies, and weather controls. Weather controls include indicators for whether the temperature on the "scheduled" test date falls into one of 7 temperature bins, an indicator for any precipitation, and precipitation, humidity, and wind speed on the "scheduled" test date. Results are shown for for Ozone (O3), but include controls for nitrogen dioxide (NO2) and fine particulate matter (PM2.5). For each pollutant, the omitted category is 0-25% of the EPA threshold. Regressions are weighted by number of students. Standard errors clustered at the school level are in parentheses.

Table 4: Robustness of main results

	Baseline (1)	School-Grade FE (2)	Demo (3)	Demo×Yr (4)	Econ (5)	Weight: none (6)	Weight: Nearest Mon (7)	Feb-May (8)	No New Monitors (9)
O3 25-50%	0.0423**	0.0434**	0.0382**	0.0392**	0.0416**	0.0478*	0.125***	0.0424**	0.0402**
	(0.0178)	(0.0177)	(0.0176)	(0.0165)	(0.0181)	(0.0245)	(0.0434)	(0.0181)	(0.0175)
O3 50-75%	0.0379**	0.0387**	0.0339*	0.0343**	0.0375**	0.0484*	0.118**	0.0383**	0.0365**
	(0.0183)	(0.0182)	(0.0182)	(0.0173)	(0.0185)	(0.0249)	(0.0478)	(0.0185)	(0.0180)
O3 75-100%	0.0526***	0.0529***	0.0490**	0.0482**	0.0523**	0.0656**	0.0954**	0.0529**	0.0572***
	(0.0202)	(0.0202)	(0.0201)	(0.0191)	(0.0205)	(0.0262)	(0.0450)	(0.0205)	(0.0203)
O3 over 100%	0.0543**	0.0573**	0.0523*	0.0605**	0.0525*	0.0677**	0.110**	0.0541**	0.0412
	(0.0267)	(0.0266)	(0.0267)	(0.0267)	(0.0270)	(0.0340)	(0.0544)	(0.0269)	(0.0276)
Observations	5,788	5,757	5,787	5,787	5,764	5,788	5,788	5,787	5,363
Schools	804	800	804	804	804	804	804	804	777
R-squared	0.617	0.624	0.622	0.636	0.616	0.586	0.570	0.617	0.614

Notes: For all regressions, the outcome is the fraction of students outside the "healthy fitness zone." All regressions include school fixed effects, year fixed effects, grade dummies, and weather controls. Weather controls include percent of test days that fall into 7 temperature bins, percent of test days with any precipitation, and average precipitation, humidity, and wind speed on test days. Results are shown for for ozone (O3), but include controls for nitrogen dioxide (NO2) and fine particulate matter (PM2.5). For each pollutant, the omitted category is 0-25% of the EPA threshold. Unless otherwise noted, regressions are weighted by number of students. Standard errors clustered at the school level are in parentheses.

Table 5: Demographic changes

	Free/Reduced Lunch (1)	White (2)	Black (3)	Hispanic (4)	Asian (5)	Other (6)
O3 25-50%	0.00873	-0.00318	0.00321	0.00364	0.00203	-0.00678*
	(0.0130)	(0.00520)	(0.00224)	(0.00573)	(0.00355)	(0.00355)
O3 50-75%	0.0111	-0.00468	0.00310	0.00426	0.00169	-0.00745*
	(0.0130)	(0.00507)	(0.00233)	(0.00602)	(0.00375)	(0.00385)
O3 75-100%	0.0101	-0.00527	0.00327	0.00467	0.00267	-0.00686
	(0.0135)	(0.00577)	(0.00269)	(0.00674)	(0.00412)	(0.00429)
O3 over 100%	0.0187	-0.00740	0.00348	-0.00596	0.0130	-0.00519
	(0.0164)	(0.00729)	(0.00484)	(0.0124)	(0.00803)	(0.00443)
Observations	5,764	5,787	5,787	5,787	5,787	5,787
Schools	804	804	804	804	804	804
R-squared	0.938	0.972	0.930	0.976	0.970	0.603

Notes: Outcomes are measured as the percent of students by race and the percent of students receiving free or reduced price lunch. All regressions include school fixed effects, year fixed effects, grade dummies and weather controls. Weather controls include percent of test days that fall into 7 temperature bins, percent of test days with any precipitation, and average precipitation, humidity, and wind speed on test days. Results are shown for ozone (O3), but include controls for nitrogen dioxide (NO2) and fine particulate matter (PM2.5). For each pollutant, the omitted category is 0-25% of the EPA threshold. Regressions are weighted by number of students. Standard errors clustered at the school level are in parentheses.

Table 6: Other fitness tests

		Other Fitness Tests				
	Aerobic Capacity	Body Comp.	Flexibility	Abdom. Strength	Trunk Ext Strength	Up. Body Strength
	(1)	(2)	(3)	(4)	(5)	(6)
O3 25-50%	0.0423**	0.0119	0.0112	0.0289**	-0.00338	0.0220
	(0.0178)	(0.0107)	(0.0179)	(0.0144)	(0.0192)	(0.0179)
O3 50-75%	0.0379**	0.00551	0.0140	0.0398***	0.000916	0.0248
	(0.0183)	(0.0104)	(0.0177)	(0.0147)	(0.0194)	(0.0178)
O3 75-100%	0.0526***	0.00193	0.0173	0.0387**	0.00239	0.0268
	(0.0202)	(0.0109)	(0.0186)	(0.0165)	(0.0210)	(0.0194)
O3 over 100%	0.0543**	-0.0116	0.0155	0.0227	-0.0237	0.0573**
	(0.0267)	(0.0148)	(0.0262)	(0.0258)	(0.0301)	(0.0277)
Observations	5,788	5,788	5,788	5,788	5,788	5,788
Schools	804	804	804	804	804	804
R-squared	0.617	0.749	0.697	0.627	0.536	0.653

Notes: For all regressions, the outcome is the fraction of students outside the "healthy fitness zone." All regressions include school fixed effects, year fixed effects, grade dummies, and weather controls. Weather controls include percent of test days that fall into 7 temperature bins, percent of test days with any precipitation, and average precipitation, humidity, and wind speed on test days. Results are shown for ozone (O3), but include controls for nitrogen dioxide (NO2) and fine particulate matter (PM2.5). For each pollutant, the omitted category is 0-25% of the EPA threshold. Regressions are weighted by number of students. Standard errors clustered at the school level are in parentheses.