The Effects of School Ventilation on Educational Outcomes: Evidence From A Large Field Study in Primary Schools

Nicolás Durán[1,2], Piet Eichholtz[2], Nils Kok[2], and Juan Palacios[3]

[1]University College London, Bartlett Real Estate Institute

[2]Maastricht University, School of Business and Economics

[3]Massachusetts Institute of Technology, Center for Real Estate

August 3, 2022

Abstract

Governments devote a large share of public budgets to construct, repair and modernize school facilities. However, little is known about whether investments in the physical condition of schools translate into student achievements. In this study, we report the results of a large field study, providing quasi-experimental evidence on the implications of poor air quality conditions inside classrooms − key performance measure of school infrastructure, and a common indicator guiding investments in school facilities. We continuously monitor the air quality conditions (i.e. CO_2, coarse and fine particles, temperature, humidity) in the classrooms of 5,500 children over two school years, and link them to their scores in over 38,000 nationally standardized tests. Using a fixed-effects strategy, relying on within-pupil changes in air quality conditions, we find that exposure to poor indoor air quality during the school term preceding the test is associated with significant performance drops. We document that a one standard deviation increase in average daily peak CO_2 during the school term leads to a 0.10 standard deviation decrease in test scores. Our results add to the ongoing debate on the determinants of student human capital accumulation, highlighting the role of physical capital in affecting learning outcomes.

Keywords: Academic Performance, Human Capital, Indoor Air Quality, School Infrastructure, Education.

JEL classification: I21, Q53.

1 Introduction

Governments invest billions of dollars in the construction and modernization of school facilities every year. In the U.S. alone, school infrastructure receive USD 490 billion, the second-biggest public investment in the country (G.A.O., 2020). In addition, the 2020 "Reopen and Rebuild America's Schools Act" allocated an extra USD 130 billion to the renovation, modernization and construction of schools across the country (Cochrane, 2021). Yet, many schools across the country are in some state of disrepair, with heating, ventilation and air-conditioning (HVAC) systems having the highest priority to receive investments. Over 40% of schools in the country rely on outdated HVAC systems that need to be updated or replaced, a significant concern for facilities where children spend eight hours a day (G.A.O., 2020).

Despite the importance of capital spending, little is known about the impacts of deficient HVAC systems on educational outcomes. Deficient ventilation systems produce deficient air exchange into buildings, and are a source of health risks to occupants, such as infection risk to airborne diseases, school absences, or cognitive impairments.[1] Lab studies provide initial evidence on the detrimental consequences of short term exposure to poorly ventilated rooms for human cognition and decision-making (Seppanen et al., 2006; Fisk, 2017; Du et al., 2020). Subjects in lab studies exposed for 3 hours to poorly ventilated rooms display lower performance in higher order cognitive functions (basic activity level, applied activity level, focused activity level, crisis response, information usage, breadth of approach, strategy) (Allen et al., 2016). In children, experimental evidence shows that kids in poorly ventilated classrooms struggle to pay attention, and display lower performance in memory and concentration tests (Bakó-Biró et al., 2012). However, there is a lack of large, long-term field studies that are able to evaluate the causal impact of poor indoor air quality on educational outcomes.

This paper provides the first causal evidence on the implications of variation in indoor air quality conditions on academic achievement in a large sample of primary school students. Primary school students spend most of their schooldays in the same classroom, experiencing prolonged exposure to the indoor conditions of that specific room. We report the results of a large, pre-registered field study in which we deploy a network of sensors, continuously monitoring the indoor environmental conditions in 225 classrooms across 27 primary schools over a period

[1] Since the beginning of the COVID-19 pandemic, there has been an increasing focus on the implications of poor ventilation for the spread of airborne diseases. Poorly ventilated rooms represent a public health risk, given the high exposure risk to occupant's droplets. In response, governments in countries like Germany, the Netherlands and the US are increasing public spending to upgrade ventilation systems to reduce the risk of SARS-COV-2 transmission in schools (BBC, 2020; Rijksoverheid, 2020).

of five academic semesters.[2] Each sensor collects high-frequency measurements on a range of indoor environmental variables – i.e. CO_2, coarse and fine particles, temperature, humidity, noise levels and light intensity. To estimate the impact of a classroom's air quality on the cognitive development of students, we link daily measures of indoor air quality in classrooms providing space to more than 5,500 primary school students (aged 6 to 12) to their scores on standardized tests. During the sample period, each student took an average of 7 tests across a range of subjects, including mathematics, spelling, reading and vocabulary, resulting in more than 38,000 unique test outcomes. All tests in our sample are designed by a national examination center (i.e. not by the teacher), as part of a national tracking system to monitor the development of students throughout their primary school education.

Our primary measure of air quality in classrooms is based on the concentration of carbon dioxide (CO_2) in the room. Humans produce and exhale CO_2, which is typically removed from the room by either mechanical systems (i.e. HVAC) or natural ventilation systems (i.e. windows), that exchange indoor with outdoor air. CO_2 is a widely used indicator by engineers and scientists to measure how much fresh (outdoor) air is brought into a room, and by public officials to set guidelines and evaluate the performance of ventilation systems in (public) buildings. As the ventilation rate (i.e., the replacement rate of indoor air with fresh, outdoor air supply) decreases, the CO_2 concentration in the room increases. To account for potential confounders, we rely on the panel structure of the data to estimate models with subject and student-fixed effects. Fixed effect regressions identify the impact of indoor air quality (i.e. CO_2) during the prior school term by leveraging within-student-subject variation in air quality over multiple school terms.

Our identification strategy relies on the premise that variation in air quality over successive terms for a given student is uncorrelated with unobserved determinants of learning. We therefore include controls for classroom infrastructure attributes by including classroom fixed-effects, and other time-varying factors that could be contemporaneous and correlated with CO_2, such as air particles, temperature, noise, and relative humidity. Nevertheless, it is still possible that other unobserved factors that are correlated with both CO_2 and test scores remain present. In particular, kids' activity patterns or teacher behavior can alter the generation levels of CO_2 in the classroom and therefore hinder a causal interpretation of our estimates. We provide a number of robustness checks to ensure that the impact of CO_2 on student performance is a direct impact of poor ventilation and not a by-product of changes in student or teacher behavior. First, we construct measures of pupil activity patterns from the minute-by-minute readings of the noise

[2] For the pre-registration of the study see Palacios et al. (2020)

sensor (e.g. number of minutes above certain decibel's (dBA) thresholds and daily averages of dBA's during the school term). Second, we use the variation in CO_2 levels associated with breakdowns in ventilation systems, linking the number of days a system is failing with student test scores.

The main results show that children who were exposed to high concentrations of CO_2 during the learning period perform worse on subsequent standardized tests. In our preferred specification, including a rich set of fixed effects, a one standard deviation increase in CO_2 level during the school term leads to a 0.11 standard deviation drop in test scores. The effects are strongest in mathematics, where a one standard deviation increase in CO_2 level during the school term is associated with an 0.21 drop in test scores. Evidence from a heterogeneity analysis suggests that poor ventilation impairs learning outcomes of students specially when they are between eight and twelve years old.

A battery of (robustness) tests sheds light on the role of three key sources of CO_2 concentrations in classrooms in our sample: (1) changes in student activity patterns, (2) teacher behavior, and (3) ventilation infrastructure. First, we control for changes in activity patterns of students in the classroom. The noise sensor is able to capture changes in a number of student behaviors that are related to the production of CO_2, such as screaming or moving in the class during teaching hours. The noise sensor also picks up the background noise associated with the opening of windows. In an analysis that tests over one hundred specifications of our main regression analysis, we show that the magnitude and significance of the coefficient associated with CO_2 levels remains unchanged when including several combinations of IEQ and noise indicators. Second, we check whether teacher quality correlates with CO_2 concentration in the classroom. Using variation of teacher's exposure to CO_2 across academic terms, we show that it does not (negatively) correlate with variation across time of their students' test scores. Finally, we exploit the plausibly exogenous variation in CO_2 concentration associated with failures in schools' ventilation systems, in the subsample of schools that are mechanically ventilated. The failure of a mechanical ventilation system in a given day results in sustained and abnormally high levels of CO_2 in the classroom (20-40% higher than in normal ventilation days). The results of an instrumental variable strategy using ventilation breakdowns show a significant drop of 0.25 to 0.4 standard deviations in test scores depending on the specification used.

This study is the first to show that exposure to poor air quality *inside* the classroom can hinder student performance, which speaks to the long-standing debate on the relationship between investments in school infrastructure and academic achievements (see Hanushek (2003)). Existing evidence shows a positive impact of school construction projects in contexts where school facilities were either in extremely poor condition or just non-existent, which suggests that new school construction projects generally are positively associated with student outcomes (Duflo, 2001; Aaronson and Mazumder, 2011; Neilson and Zimmerman, 2014). Similarly, another stream of quasi-experimental studies investigates the link between (general) school spending or school investment campaigns for school infrastructure and academic outcomes (Martorell et al., 2016; Jackson et al., 2016). Finally, Stafford (2015) provides evidence that public funding campaigns targeting mold reduction and ventilation improvements have a positive impact on student performance in elementary schools. This study departs from the existing literature by investigating actual air quality in the classroom, rather than just broad, monetary indicators of changes in school infrastructure, providing insights based on objective, high-frequency measures of indoor environmental quality. Our outcome-based approach of school quality can facilitate for more precise estimates as compared to a purely input-based approach (see Hanushek (2003) for a discussion of misallocation of resources in school investments).

In addition, this paper contributes to the nascent literature that explores the role of environmental factors (i.e., air pollution and extreme heat) in cognitive performance and human capital development. Over the last decade, there have been a number of studies providing quasi-experimental evidence on the negative effects of exposure to extreme temperatures or ambient air pollution on human health and human capital accumulation (see, for a review, Graff Zivin and Neidell (2013) and Graff Zivin and Neidell (2018); Roth (2017)). Prolonged exposure to high levels of air pollution has been associated with numerous respiratory problems in early life (e.g. asthma), ultimately affecting school absence (Currie et al., 2009; Currie and Walker, 2011; Knittel et al., 2016) and infant mortality (Chay and Greenstone, 2003; Currie and Neidell, 2005).[3]

Beyond the health damage, there is increasing evidence on the direct consequences of exposure to air pollution on the human brain and cognitive performance (Zhang et al., 2018). An increasing number of studies show that exposure to air pollution harms student performance. Numerous studies have linked local levels of air pollution on testing days (i.e. high levels of $PM_{2.5}$) to lower performance of young adults in high-stakes examinations (Ebenstein et al., 2016; Roth, 2018; Graff Zivin et al., 2020).[4] In the medium term, accumulated exposure to traf-

[3] There are also numerous studies showing the effects of elevated concentrations of fine particles on mortality rates in adult populations (Liu et al., 2019). At the macro level, the impact of air pollution on human health is staggering: the World Health Organization (WHO) estimates seven million premature deaths due to poor air quality (WHO, 2014).

[4] Air pollution also affects labor market outcomes. In particular, the literature provides evidence of air pollution

fic or industry-induced pollution during an academic year is correlated with lower test scores in subsequent exams, and with behavioral incidents during high school (Persico and Venator, 2019).

Our results contribute to the existing literature in multiple ways. First, the overwhelming majority of studies use *outdoor* climate measurements to assess students' exposure, often using data from air quality or weather stations located miles away from the schools where the pupils are learning and taking their tests. [5] We collect data on air quality and other environmental metrics inside the classrooms, where our subjects are learning and taking exams, overcoming the challenge of measurement error which could result from miss-assigning environmental conditions to individuals (Moretti and Neidell, 2011; Roth, 2018). The access to rich data retrieved from our large sensor network allows controlling for a rich set of factors often neglected in the literature (e.g. noise) and explore channels through which poor air quality affects student performance. Second, we provide evidence on the impact of environmental conditions on children in elementary schools, a cohort in which the implications of exposure to poor air quality or extreme temperatures are still largely unexplored. The current evidence mostly relies on samples of high-school or university students.[6] Our estimates are based on individual standardized tests taken by children twice a year throughout all elementary school years (age 6 to 12), a critical age range for cognitive and human capital development (Howard-Jones et al., 2012; Heckman, 2006).

The remainder of the paper is organized as follows. In Section 2, we describe the study design and descriptive statistics of the main variables in the study. Section 3 describes the empirical strategy used to link indoor environmental conditions to student academic performance. Section 4 present the estimation results, and Section 7 concludes.

2 Data

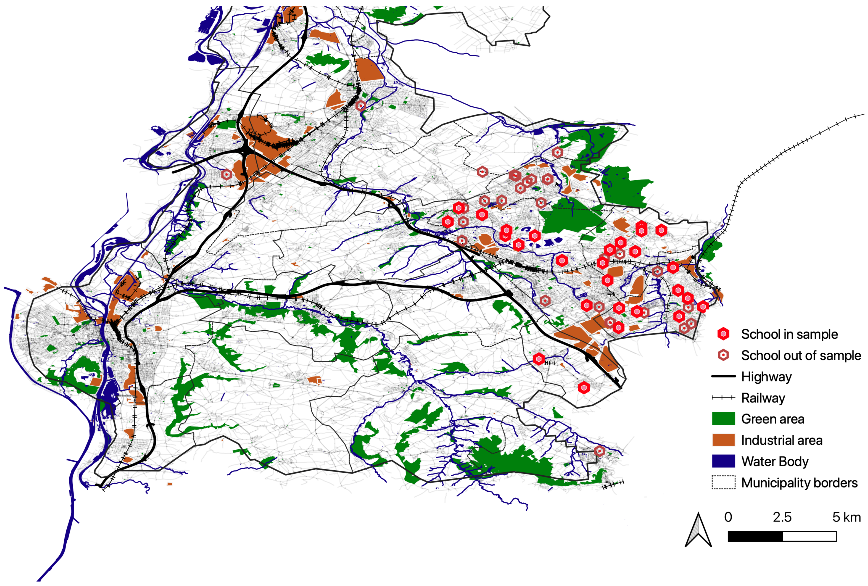
This study exploits data from a large-scale network of sensors, deployed in 225 classrooms across 27 schools in the Netherlands. In each year of the sample period, the schools have an aggregate

affecting the productivity of agricultural workers (Graff Zivin and Neidell, 2012), the productivity of factory workers (Chang et al., 2016), and soccer players (Lichter et al., 2017). Importantly, the effects of outdoor ambient air pollution also have implications for indoor labor, affecting call center productivity (Chang et al., 2019), trading activity (Meyer and Pagel, 2017), decision time and quality of judges (Kahn and Li, 2020), and the performance of chess players (Künn et al., 2019).

[5] A notable exception is Roth (2018), who deploys indoor sensors to measure the level of air particles (PM_{10}) during the exams of university students.

[6] Persico and Venator (2019) is a notable exception, investigating the impact of proximity to industrial sites or busy highways on the performance of elementary school students, studying the learning performance of children from grade 5 ($10-11$ years old).

Figure 1: Map of School Locations and Average Income



Note: Figure 1 shows the location of each school in the sample (bright red). Dots in dark red represent schools in the area not selected for the study. Schools were selected based on a random sample of a set of 47 schools belonging to the largest school board in the region. All schools in the sample below to the same metropolitan area, with similar household income, outdoor temperature and outdoor level of air quality. The metropolitan area has a total of 257.499 inahbitants, spread over 6 municipalities (Wikipedia, 2022).

enrollment of more than 5,500 students.

2.1 School Characteristics

Our sample consists of 27 schools randomly selected from 47 schools managed by the largest school board in province of Limburg, the most southern province in the Netherlands.[7] The bright red polygons in Figure 1 show the location of each school. All schools are located in the same metropolitan area, exposed to similar levels of outdoor temperature and air quality. The metropolitan area in which the schools are located is generally considered a lower-SES part of the Netherlands, with median net household incomes varying from €21.9-25.6k, compared to the national median household income of €25.8k.

The schools in our sample all follow the same teaching curriculum, and children are evaluated following the same set of standardized exams (Section 2.3 describes the national testing system in depth). Panel A of Table 1 displays the characteristics of the schools in the sample. The average school has about 10 classrooms, with an average class size of 28 students. The youngest students are four years old, and the oldest students are thirteen years old. In our study, we focus

[7]See Palacios et al. (2020) for the study protocol, including a detailed description of the sample and school typology, pre-analysis plan, and an extensive discussion of sensor placement and calibration.

Table 1: Summary Statistics: Schools and Groups

Panel A: Schools	Mechanically Ventilated				Naturally Ventilated			
	Mean	St. Dev.	Min.	Max.	Mean	St. Dev.	Min.	Max.
Schools	23	0	23	23	4	0	4	4
Classrooms per school	8	2	5	13	7	2	5	10
Age of Buildings	25	20	2	88	48	9	38	60
Age of Ventilation System	8	6	0	21				

Panel B: Groups	Mechanically Ventilated				Naturally Ventilated			
	Mean	St. Dev.	Min.	Max.	Mean	St. Dev.	Min.	Max.
Age	9	2	5	13	10	2	5	13
Group Size	28	10	15	63	30	11	15	55
Years of Proficiency	3	1	0	8	3	1	0	5

on students who are older than six, the age at which the standardized test system starts for all core subjects in the curriculum.

Schools in our sample are 25 years old, on average. The majority of schools in our sample (85 %) have a mechanical ventilation system, in the other schools, the ventilation system is "natural" (i.e. by opening and closing windows). Note that none of the schools have an air-conditioning (cooling) system, given the relatively mild temperatures in Summer.

2.2 Environmental Conditions in Classrooms

For each school in the sample, the environmental conditions in each classroom with students aged 6 and above are continuously monitored throughout the sample period, using advanced environmental sensors. We use wall-mounted stationary sensors from the sensor company Aclima, Inc., that monitor the levels of CO_2 (ppm), coarse and fine particles (PM_{10}, counts/L), temperature (° C), relative humidity (rH), light intensity (lux) and noise (dBA). The sensors capture raw data with a frequency of 1 to 30 seconds, transmitting all data to a cloud-based server. The data is aggregated at the minute level, and becomes available to the research team at the end of every day. The deployment of sensors took place between January 2018 and December 2018.[8]. Supplementary Figure C.3 describes the daily statistics of sensor coverage per date as well as the time period covered by the sensors. The Figure shows that the sensor coverage reaches full coverage in January 2019, upon completion of the sensor deployment, which has been fully operational since then.

We assess the quality of indoor air in each classroom based on the levels of CO_2 concentra-

[8]Each sensor is plugged into the wall for electricity and is connected to the local WiFi network for secure data transmission. During some days there are sensors that do not deliver any data (typically the result of sensors that are unplugged during cleaning, etc.)

Table 2: Summary Statistics for Environmental Conditions in Classrooms

Panel A: Room Ventilation	Mechanically Ventilated				Naturally Ventilated			
CO2 Concentration (ppm)	Mean	St. Dev.	Min.	Max.	Mean	St. Dev.	Min.	Max.
Daily Average (in ppm)	957	324	485	2783	1249	321	604	2024
Daily Peak (in ppm)	1430	601	737	4665	2036	547	928	3189
Peak > 1000 (% days with class)	75	30	0	100	95	12	53	100
Peak > 2000 (% days with class)	12	24	0	100	46	25	0	100
Peak > 1000 (% lesson time)	65	32	0	100	92	14	43	100
Peak > 2000 (% lesson time)	9	21	0	96	37	23	0	85

Panel B: Environmental Parameters	Mechanically Ventilated				Naturally Ventilated			
Daily Average	Mean	St. Dev.	Min.	Max.	Mean	St. Dev.	Min.	Max.
Temperature (in °C)	21	1	19	25	21	1	18	23
Humidity (in rH)	43	6	28	57	45	5	35	55
Noise (in dBA)	56	2	47	64	55	1	52	60
PM_{10} (in Count/L)	1045	554	72	2950	1608	512	814	3465
Daily Peak	Mean	St. Dev.	Min.	Max.	Mean	St. Dev.	Min.	Max.
Temperature(in °C)	22	1	20	27	22	1	21	25
Humidity(in rH)	48	6	31	66	50	5	39	60
Noise(in dBA)	77	3	66	85	75	2	69	82
PM10(in Count/L)	2522	1256	334	8497	3508	1240	1697	9758

tion, as a direct measure of the ventilation rate in the room. CO_2 is a widely used indicator by building facility managers and policymakers to monitor and regulate ventilation rates in buildings (ASHRAE, 2022). Occupants exhale CO_2, which stays in the room until a mechanical or natural ventilation system removes and exchanges it with outdoor air. The recommendation from the main institution setting standard on building ventilation, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE, Standard 62.1-2013) highlights that building ventilation rates should keep indoor CO_2 concentrations at a maximum of 700 parts per million (ppm) above outdoor concentrations, in order to ensure occupant satisfaction and comfort. The levels of outdoor CO_2 are always an order of magnitude lower than those in occupied rooms – the global average atmospheric carbon dioxide in 2020 was 412.5 ppm (NOAA, 2021).[9]

Table 2 provides the summary statistics of the indoor environmental quality variables obtained from the sensors, while the children in our sample are inside the classroom. We restrict the sample period to those official school days that the classroom is occupied, which can be inferred by the levels of CO_2 and noise in the room. The sensors monitor the levels of the six indoor environmental quality variables continuously throughout the year. Panel A in Table 2 describes

[9]In addition, the onset of COVID-19 has triggered the development of new guidelines to support strategies to use ventilation to tackle transmission risk of airborne diseases (e.g. EPA, 2022).

the distribution of the CO_2 in the study. Consistent to a recent review of the literature of CO_2 in schools Fisk (2017), all students in our sample are exposed to CO_2 levels above the recommended threshold. The distribution of daily CO_2 peaks indicates that the average student in our sample is exposed to 1,436 ppm on a given school day, with a range spanning from 700 ppm to 3,950 ppm (i.e. almost four times the limit recommended by ASHRAE). The comparisons between the right and left side of the table shows how the levels of CO_2 are significantly influenced by the presence of a ventilation system in the school. In naturally ventilated classroom, which rely on opening windows rather than a mechanical HVAC system to ventilate rooms, the average levels of CO_2 are substantially larger – i.e. 606 ppm (42%) higher than those in mechanically ventilated schools.

Panel B in Table 2 provides the summary statistics of the remaining indoor environmental quality variables collected by the sensors. The results show that, on average, the levels of temperature, humidity, particles and noise are within the healthy and comfortable levels proposed by regulators. The average thermal conditions, measured by the relative humidity and temperature in the classroom, are within the comfortable levels. The average temperature levels are 21°C (69.8°F), within the comfortable range for humans (17–24°C; 63-75°F) and below the temperatures considered harmful for human health (Asseng et al., 2021). The maximum daily peak in our sample is 27°C (80 °F), within levels recently estimated to impact student performance (Park et al., 2020).[10] The levels of noise in our classroom are also within the levels to ensure human health (Hammer et al., 2014). [11]

2.3 Student Performance Data: Nationally Standardized Tests

In the Netherlands, student performance in primary schools is tracked through biannual, nationally standardized tests taking place halfway through the school year (January-February) and at the end of the school year (May/June). The tests cover a wide range of education domains, including Mathematics, Reading, Spelling, and Vocabulary, and apply to students from Kindergarten through 6th grade. [12].

For each student in our sample, we collect scores for all tests over the entire primary school education. The source of this data is the *OnderwijsMonitor Limburg* – a collaboration be-

[10]In our regression, we control flexibly for temperature to avoid any confounding effects in our estimates associated with CO_2.

[11]We also tested for background levels of noise with the distribution of noise in unoccupied classrooms. On average, all classrooms in our sample display noise levels below 35 dBA – i.e. the recommendations by the American National Standards Institute and the Acoustical Society of America (Spratford et al., 2019).

[12]In the Netherlands, grades correspond to groups, where Kindergarten is group 1, for 4 year-old children, and 6th grade is group 8, for 12 year-old children

Table 3: Summary Statistics for Test Scores

	Mechanically Ventilated					Naturally Ventilated				
By Age Bracket	N	Mean	St. Dev.	Min.	Max.	N	Mean	St. Dev.	Min.	Max.
[5 − 7]	8810	0.00	0.45	-3.41	3.69	945	0.01	0.40	-2.09	0.77
[8 − 9]	11748	-0.02	0.39	-2.64	2.82	1318	0.01	0.42	-3.41	3.35
[10 − 13]	12886	-0.08	0.38	-2.57	1.99	1697	0.01	0.36	-2.10	1.04
By Test Domain	N	Mean	St. Dev.	Min.	Max.	N	Mean	St. Dev.	Min.	Max.
Mathematics	9236	-0.03	0.42	-3.41	3.69	1122	0.01	0.33	-1.06	2.27
Spelling	9491	-0.04	0.41	-2.47	3.15	1043	-0.03	0.41	-1.79	3.35
Reading	7504	-0.03	0.38	-2.16	2.10	843	-0.08	0.39	-2.66	1.28
Vocabulary	2393	-0.12	0.38	-2.27	2.61	250	0.31	0.42	-3.41	0.79
Other	4820	-0.01	0.41	-2.04	2.82	702	0.07	0.38	-1.53	2.82
By Test Period	N	Mean	St. Dev.	Min.	Max.	N	Mean	St. Dev.	Min.	Max.
Jan-Feb 2018	1364	-0.17	0.29	-2.41	0.67					
May-Jun 2018	1229	-0.18	0.38	-2.46	1.02					
Jan-Feb 2019	8431	-0.04	0.42	-1.85	3.69	522	-0.05	0.31	-0.90	0.56
May-Jun 2019	9504	-0.01	0.41	-3.41	2.22	1644	0.03	0.39	-3.41	2.05
Jan-Feb 2020	12916	-0.02	0.39	-2.57	3.57	1794	0.01	0.40	-2.10	3.35

tween Maastricht University and the elementary schools, school boards, and municipalities in the province of Limburg (for more information, see Borghans et al. (2015)). For the purpose of this study, we exclude testing data in Kindergarten (groups 1 and 2), given the limited testing taking place for these grades and the limited comparability of test results relative to subsequent grades.

The tests are designed by a national examination center, administered at each individual school, and graded by the teachers using a standardized grading scheme. The raw results are transformed into percentiles by the testing center. The rules for transformation are constant over time, such that results for the same student can be compared between periods. For each test period, we construct a comparable scale for each domain, standardizing the variable to have a mean of zero and a standard deviation of one within a given test in a test period. The main outcome variable is thus the standardized score for each student, in each test period and domain (we only consider students that take the same test domain at least twice across different test periods).

3 Empirical Approach

In our identification strategy, we exploit the fact that we observe students that are tested multiple times during the sample period, with exposure to varying levels of indoor air quality during the school term preceding the test (i.e. the learning period). The data allows us to test whether students score lower following a term in which their classroom was poorly ventilated, relative to their own score following a school term in a classroom with “good” air quality.

We estimate a fixed-effects model, that removes the influence of confounding factors driven by ex-ante differences in student skills or socio-economic background, classroom infrastructure, and general trends in test scores in our sample for each testing domain. More formally, we estimate the following specification:

$$Score_{idt} = \beta CO2_{ct} + \alpha_{id} + \alpha_t + \alpha_c + \alpha_l + \Gamma X_{ct} + \varepsilon_{idt} \quad (1)$$

where $Score_{idt}$ is the standardized test score of student i, in domain d (e.g., mathematics, vocabulary, etc.), and testing period t (e.g. February 2020). We define cumulative CO_2 exposure, $CO2_{ct}$, as the average daily peak CO_2 experienced during school days in the school term prior to the test for all students in classroom c taking the test in testing period t. We standardized the parameter, such that the coefficient of interest β can be interpreted as the standard deviation impact on a student's test score associated with a standard deviation increase in the average daily peak of CO_2 in the classroom where students took classes during the term.

The model includes fixed effects for each student i by test domain d (α_{id}) to capture changes in idiosyncratic abilities of students in different education domains (e.g. mathematics, reading, etc.). The inclusion of testing-period t fixed effects (α_t) controls for common factors affecting all pupils taking a test in the same testing period. The classroom fixed effects (α_c) control for all time-invariant characteristics of a classroom, such as views, or any relevant major (time invariant) differences in teaching infrastructure in the room (e.g. digital boards, furniture, etc.).

Finally, X_{ct} is a vector of IEQ, group, and individual controls containing average daily peak measures of PM_10, Temperature, and Humidity observed in the classroom during the learning period, a linear and quadratic term for group size, two dummies indicating the age of the student, i.e. 8 to 10, and 10 to 13 years old with reference 5 to 8 years old, and the average level of noise in the classroom during the learning period. In addition, we include a a series of dummmies describing the number of terms that students have received lessons on the specific subject by the time they take the test in their school time. These control for heterogeneity in test results that might arise from students repeating the taking of a test, for example.

Standard errors (ε_{ict}) are clustered at the classroom by period level to control for correlation among tests results for students learning in the same classroom at the same time, and following the suggestion by (Abadie et al., 2017) to cluster at the level of treatment variation. In addition, all regressions are weighted by the number of days that a sensor records valid data in each school term, to take into consideration the sample size from which we derive our exposure measures during each school term.

The identifying assumption of our empirical approach is that the variation in standardized test scores for each student is independent of other variables that might be correlated with CO_2 levels. The key identifying assumption is therefore that, conditional on the rich set of fixed effects, there is no unobservable (environmental) factors correlated with classroom CO_2 and test scores. To ensure the robustness of our results, we implement the following strategies. First, the access to a rich set of environmental parameters from the environmental sensors allows controlling for a number of factors often neglected in the literature. All specifications include a rich set of environmental controls (X_{ct}) that enable us to isolate the impact of indoor air quality (CO_2) from all other key environmental conditions in the classroom – i.e. average exposure to a classroom's coarse and fine particles (PM_10), temperature, relative humidity and noise intensity.[13] To test the robustness of our results to different specifications, we implement a specification curve where we test the changes in our main parameter of interest under multiple transformations of the CO_2 measure and the environmental controls.

Second, in order to further isolate the impact of ventilation on learning outcomes, we exploit the plausibly exogenous variation in CO_2 associated with failures in schools' ventilation systems. A selection of schools in the sample are mechanically ventilated. The failure of a mechanical ventilation system on a given day results in sustained and abnormally high levels of CO_2 in the classroom. We constructed a data-driven algorithm where we detect days when there are abnormal levels of CO_2 for a sustained amount of time in the classrooms of our sample during our sample period. The algorithm employs a fixed effects specification to predict each classroom's steady peak and average CO_2 levels (conditional on the time of the day, date, academic term, and classroom), and indicates the days when both measures of CO_2 in a classroom are 15000 ppm above them.

Finally, it is important to note that in our setting, classroom equipment and teacher quality are unlikely to co-move over time with the indoor air quality in the classroom during our sample period. A series of focus interviews between the research team and the board of the schools confirmed the lack of major changes in school furniture or equipment during our sample period. In our sample of schools, the teaching material and equipment is procured at the school level

[13]We construct these variables following the same procedure as our main treatment variable. The measurements are standardized to facilitate comparisons across the different environmental factors.

or at school board level. In a robustness check, we include school-by-year fixed effects α_{st} to control for the present of any school-level shocks. [14] In addition, an ultimate reason of concern is that teachers sort into rooms based on indoor environmental quality, with high-quality teachers selecting classrooms with healthier conditions. In our sample of schools, it is the school principal who determines classroom allocation to teachers and student groups, rather than individual teachers. Furthermore, the classroom fixed effects of the fixed-effect model partly control for this self-selection into classrooms, since those teachers that always teach in their favorite classroom will be part of the time invariant aspects of the classroom – this is absorbed by the classroom fixed effects α_c.

Nevertheless, we collect data on the allocation of teachers to classrooms at each term and for each school in our sample to test how air quality and teacher quality correlate. We construct the same environmental variable as for the main analysis for air quality using the average daily peak exposure to CO_2 for each teacher. Similarly, we compute the average score achieved by each teacher's students in each academic term. Regressing separately average scores and average CO_2 daily peaks on fixed effects indicating each teacher, we look at how their estimated coefficients correlate and show that our assumption is plausible.

4 Results

4.1 Indoor Air Quality and Student Performance

This section presents our main estimation results liking the variation of CO_2 in the school term to student performance in standardize tests, together with a series of heterogeneity tests, and robustness checks.

4.1.1 Main Results

We first analyze how changes in average exposure to CO_2 concentrations during the school term affect the test scores of individual students.

Table 4 displays the estimated coefficients in our main specification (Eq. 1), with different versions that introduce sequentially the set of controls to test the sensitivity of our main effects. The results indicate that high concentrations of CO_2 in the classroom during the school term clearly lowers the performance of students in the test. The coefficient is highly statistically significant. After controlling for the environmental conditions in the classroom (Column (2)),

[14] In addition, the research team undertook visual inspection to a random set of schools every school term, confirming the lack of changes in the classrooms during the study period.

Table 4: Average Daily Peak CO2 Concentration During Learning Period on Standardized Test Scores

	Dependent Variable: Standardized Test Scores				
	(1)	(2)	(3)	(4)	(5)
CO_2 (z-score)	-0.066*	-0.094**	-0.115***	-0.108**	-0.110***
	(0.037)	(0.041)	(0.043)	(0.045)	(0.038)
PM_{10} (z-score)		0.016	0.010	0.044	0.047
		(0.062)	(0.058)	(0.057)	(0.071)
Temperature (z-score)		-0.014	-0.011	-0.006	-0.016
		(0.029)	(0.029)	(0.025)	(0.031)
Humidity (z-score)		0.058	0.066	0.055	-0.029
		(0.048)	(0.053)	(0.053)	(0.067)
Age [8-9]			-0.030	-0.024	-0.032
			(0.039)	(0.039)	(0.038)
Age [10-13]			-0.022	-0.019	-0.019
			(0.025)	(0.025)	(0.024)
Class Size				0.021***	0.024***
				(0.008)	(0.008)
Class Size Sq.				-0.000**	-0.000**
				(0.000)	(0.000)
Avg. Noise (z-score)				-0.039	-0.055
				(0.031)	(0.040)
Fixed Effects					
Student by Domain	Y	Y	Y	Y	Y
Period	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y
Proficiency	N	N	Y	Y	Y
School by Period	N	N	N	N	Y
Obs.	37,451	37,451	37,451	37,451	37,451
Adj. R^2	0.741	0.741	0.750	0.750	0.754

All models relate average daily peak of CO2 concentration in the classroom with standardized test scores. Column (1) provides results for a model with fixed effects for students by each domain and periods. Column (2) provides results for a model that controls for observed and unobserved physical conditions within the classroom using average daily peaks of other IAQ variables and classroom fixed effects. Column (3) gives results for a model that adds characteristics of the groups and students, i.e. group size, age, and the average daily mean noise within the classroom. Finally, the last column adds school by period fixed effects to use only variation within each specific school and rule out unobserved heterogeneity at the school level. Clustered standard errors at the classroom by period level are shown in brackets. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

our estimates describe that the exposure to one standard deviation extra in average CO_2 levels during the term lowers student performance by 0.094 standard deviations. This result is very precisely estimated and robust to a variety of controls for potential confounding factors.

That students score declines following terms with poor air quality relative to their own scores following terms with good air quality in the classroom does not appear to be driven by other channels potentially correlated with CO_2 in the school term leading up to the test, as seen in the next three columns of Table 4 (Column (3)-(5)). Controlling for changes in student age, and for promotions of students within the curriculum to higher level groups (or students repeating the same grade) has nearly no effect on the point estimate (Column (3) in Table 4). Similarly, controlling for changes in class characteristics (i.e. class size, and noise levels capturing changes in behavior of students) leaves our estimate nearly unchanged, implying that we are measuring the direct impact of poor ventilation in the classroom and not changes in class characteristics (Column (4) in Table 4). Finally, Column (5) presents the results of a regression including school-period fixed effects, that control flexibly for any changes in the school across periods, such as outdoor environmental conditions (e.g. air pollution or outdoor temperature), investments in equipment, or changes in the leadership of the school 4. Again, the results remain largely unaltered, supporting the hypothesis that our effects are reflecting the direct effects of CO_2 concentrations in the classroom during the school term, and not other confounded changes in the school setting.

The estimates associated with other environmental parameters considered in the study (i.e. PM_{10}, temperature, humidity) indicate that these factors have, on average, no significant impact on test scores in our sample, conditional upon the inclusion of fixed-effects. It is important to note that indoor temperature or particles have a lower range of variation in our data set, due to the thermal control infrastructure in the schools (i.e. heating systems). In addition, the geographical density of the schools in our sample limits the cross-sectional variation of coarse particles, that is our proxy for outdoor particulate matter. The changes over time in those conditions are controlled for by the period fixed effects and school-period fixed effects.

4.2 Heterogeneity

This section explores differences in the damage functions associated with poor ventilation. We wish to identify whether there are subpopulations or mental tasks that may be particularly responsive to poor indoor air quality. Second, this may help to identify mechanisms for the observed reduced form relationship between indoor air quality and student performance.

Table 5 presents the estimates of our main regression for the different sub samples, including the full set of fixed effects. Column (1) to (3) in Table 5 show the estimated impact of CO_2 for spelling, mathematics, and reading separately. The Table shows that the results in the pooled estimation, including all tests in the sample, are mainly determined by mathematics tests and reading comprehension tests. Columns (4) to (6) in Table 5 present the main estimates for three age groups (5 to 7 years old, 8 to 10 years old, and older than 10 years old). The results indicate that the impact of CO_2 is strongest for the two oldest cohorts.

In addition, we explore differences in impacts associated with high concentrations of CO_2

Table 5: Average Daily Peak CO2 Concentration on Standardized Test Scores

	by Domain			by Age		
	(1) Spelling	(2) Math	(3) Reading	(4) Age[5-7]	(5) Age[8-9]	(6) Age[10-13]
CO_2 (z-score)	-0.022	-0.215***	-0.116**	0.109	-0.173***	-0.185***
	(0.061)	(0.057)	(0.059)	(0.072)	(0.052)	(0.064)
Fixed Effects						
Student	Y	Y	Y	Y	Y	Y
Student by Domain	N	N	N	Y	Y	Y
Period	Y	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y	Y
Proficiency	Y	Y	Y	Y	Y	Y
Obs.	10,627	10,437	8,434	9,553	13,164	14,734
Adj. R^2	0.7565	0.7836	0.7503	0.7200	0.7749	0.7624

All models relate average daily peak of CO2 concentration in the classroom with standardized test scores. Column (1) provides results for a model with fixed effects for students by each domain and periods. Column (2) provides results for a model that controls for observed and unobserved physical conditions within the classroom using average daily peaks of other IAQ variables and classroom fixed effects. Column (3) gives results for a model that adds characteristics of the groups and students, i.e. group size, age, and the average daily mean noise within the classroom. Finally, the last column adds school by period fixed effects to use only variation within each specific school and rule out unobserved confounders at the school level. Cluseterd standard errors at the classroom by period level are shown in brackets. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

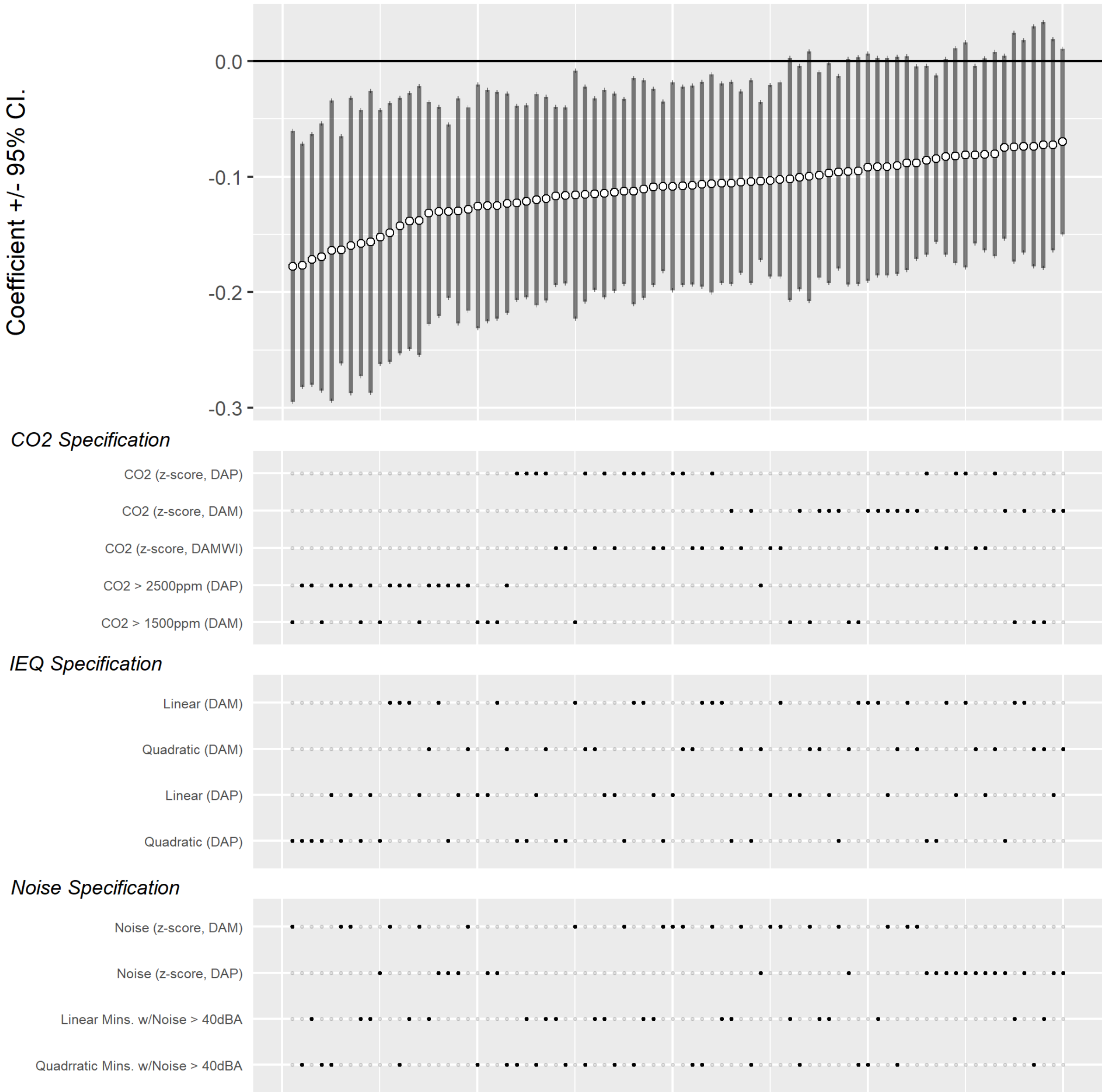
across school week days. Schools design their curriculum in a way that content and skill development training is distributed over the school term and week. Here, we construct exposure measures for each day of the week and re-estimate our main specification 1. Supplementary Table A.3 displays the coefficients associated with the average concentration of CO_2 in different school days. The results indicate that poor environmental conditions are equally important across school days, except for Wednesdays, when students have only half day of class in Dutch primary schools. That estimates associated with the exposure the day of the week that there is only half instruction time highlights the role of exposure time in our study.

5 Robustness Checks

5.1 Specification tests

In our main results, we include all regressors as the average of daily maximum over the school-term. This subsection presents the results of a specification curve analysis (Simonsohn et al., 2020), where we test the sensitivity of our main results to the functional form of our environmental exposure measures. In particular, we run a variety of specification tests, where we include in the regression model different forms of our treatment measure and indoor environmental

Figure 2: Specification Sensitivity Curve



controls.

Specification curve analysis of environmental parameters. Figure 2 displays the coefficients and confidence intervals associated with the concentration of indoor CO_2 in the school term preceding the test for a combination of dozens of different forms of the variable itself and the environmental controls. In our main specification, our measure of exposure during the term is based on the average daily peaks across all teaching days in the term (denoted in Figure 2 as CO_2(z–score DAP)). The figure shows the sensitivity of the estimate to specifying the school-term concentrations of CO_2 based on average daily average concentrations (CO_2(z–score DAM)), average daily average concentrations after correcting the algorithm to consider only the minutes that children were inside the classroom (CO_2(z–score DAMWI)) [15], a dummy variable that indicates that the school-term average of daily peaks of CO_2 were above 2500 ppm ($CO_2 > 2,500$ ppm (DAP)), or a dummy variable that indicates that the school-term average of daily averages of CO_2 were above the 1500 ppm ($CO_2 > 1,500$ ppm (DAM)). Similarly, we test different specifications of the environmental controls based on the school-term daily averages (Linear (DAM)) and their quadratic form (Quadratic (DAM)), and the school-term daily peaks (Linear (DAP)) and their quadratic form (Quadratic (DAP)).

Figure 2 shows that our results are robust to the different specifications of the treatment and environmental control variables. The coefficient shows high stability in terms of its statistical significance. The estimates in our main specification are within the confidence intervals of the CO_2 coefficient in each specification in the curve. The only few specifications ($< 2\%$ of specifications) that are only statistically significant at 10%are associated with the school-term average of daily mean, without the adjustment of the algorithm to include only the minutes that children are in the classroom (CO_2 (z–score) DAM). The occupancy patterns in the classroom affect the levels of CO_2 and therefore generate intraday variation in the CO_2 measurements that introduce measurement error in our estimation, and reduce the size of our coefficient. The consistency of our estimates across specifications support the robustness of the findings, and indicate that our results are not driven by one single form of computing the exposure of kids to poor indoor air quality.

Noise measures of activity patterns in the classroom. Alterations in children activity patterns might generate changes in CO_2 levels. As the physical activity increases, the exhalation rate increases, producing a subsequent increase in the production of CO_2 and ultimately increasing its concentration in the classroom. To test the role of these channels in our setting, we use the noise sensor to construct multiple indicators of activity patterns among children in the classroom. In particular, we include the school-term average of daily peaks and daily averages, and a variable capturing the number of minutes in the school-term where the sensor in the room captured a signal that was above 40 dBA,[16] as a proxy of the number of minutes where occupants were actively moving or speaking. Figure 2 shows the robustness of our results to changes in the specification of the noise in the room. This suggests the lack of influence of activity patterns in the room on our estimates, and therefore indicates that alterations of activity patters in the

[15]See Supplementary Section C.3 for a description of the algorithm to detect the presence of students in the classroom

[16]We tested the sensitivity of the results to different thresholds showing no influence in our estimates.

classroom are not a source of meaningful variation in CO_2 levels in our study.

Falsification test: miss-assignment of sensor to student. Finally, we perform a set of falsification tests where we misspecified the exposure of CO_2 of the student in the school term with a sensor different from the one deployed in the classroom of the student. We estimate the relationship between indoor CO_2 in a classroom *other* than the actual classroom where the student was allocated for the school term. In Supplementary Table A.4, we display the relationship between test scores and indoor air quality collected from a sensor different than the sensor deployed in the classroom where the student learned during the school term. None of the results in the table are statistically different from zero. This is the case for both, classrooms that belong to the same school of the assigned classroom for the student (Panel b), and a classroom belonging to a different school than the student was truly assigned (Panel a). The lack of a significant effect in these placebo tests is reassuring that our results are not driven by a spurious correlation.

5.2 Instrumenting CO_2 concentrations in the classroom with malfunctioning ventilation systems

In this subsection, we explicitly test for the role of the school infrastructure (i.e. HVAC system) as a key driver of main effects. We develop a data driven algorithm to detect infrastructure failures based on the presence of jumps (i.e. sudden changes to transitory and abnormally high levels) in the time series of daily peaks of CO_2. Then, we implement a two stage strategy where the school-term average concentration of CO_2 is instrumented with the number of days that the algorithm detects that the HVAC system supplying fresh air in the classroom is broken.

In our sample 85% of schools are mechanically ventilated, with an HVAC system refreshing the air in the room with outdoor air. One of the most common consequences of an HVAC malfunction generates is the reduction of air exchange rates in the room, limiting the refreshment of indoor air with outdoor air and letting the level of CO_2 accumulate to very high levels. The presence of high concentrations of CO_2 is one of the main indicators used by technicians to detect failures in ventilation systems. The failure of engines or blockage of the pipes in the system limits are common failures that generate abnormally high concentrations of CO_2 in the classroom, that remain high until the system is repaired. Ventilation failures are unlikely to be correlated with teaching quality, as they are the consequence of an infrastructure failure that teachers cannot predict.

We use our data from our sensors to detect spikes in the daily concentration of CO_2 in the classroom. We design a simple algorithm to detect outliers in the time series of daily peaks for each classroom. In particular, we infer from the data that there has been a ventilation failure by regressing the daily maximum and daily average levels of CO_2 observed in each classroom on a series of fixed effect that control for regularities (i.e. flexible trend) in observed CO2 levels during lessons and specific dates as shown below:

$$CO2^s_{ch\tau} = \alpha_{cht} + \alpha_{cwt} + \alpha_\tau + \varepsilon_{ch\tau},$$

where $CO2_{ch\tau}$ is the maximum level of CO_2 observed in classroom c, during the lesson that started at time h on date τ, α_{cht} is a classroom c by hour of the school day h during learning period t, α_{cwt} is a classroom by day of the week w during learning period t, and α_τ is a date fixed effect that controls for common environmental factors affecting all classrooms at the same time, such as weather conditions.

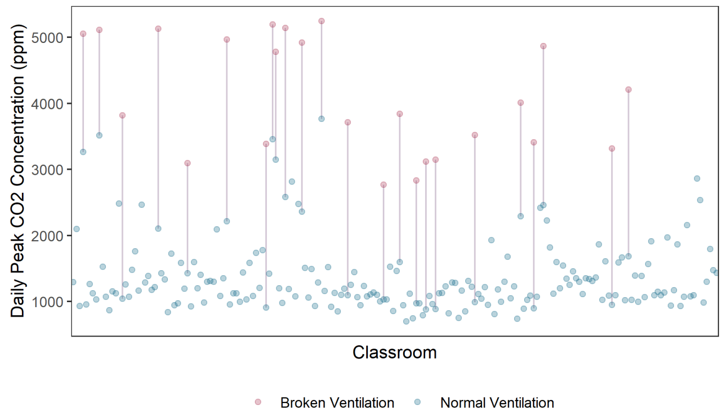
To determine whether a specific classroom presents a ventilation system breakdown during a specific day, we look at the residual elements resulting from the previous regressions $\hat{\varepsilon}_{ch\tau} = \{\hat{\varepsilon}_{ch1}, \ldots, \hat{\varepsilon}_{ch\tau}, \ldots, \hat{\varepsilon}_{chT}\}$. We consider that a ventilation system breakdown has taken place at classroom c during time of the day h in date τ if $\hat{\varepsilon}_{ch\tau} > 1,500$ ppm for both, average and maximum levels. The key rational of considering jointly the maximum and the mean values in the analysis, rather than one of them separately, is to ensure that the abnormally high levels of CO_2 are sustained over a substantial amount of time in the day, rather than just being a spike in CO_2 in a few minutes during the entire school day.

Panel (a) in Figure 3 shows average peak CO_2 concentration levels in each classroom during the school term in days with (in red) and without (in blue) broken ventilation. The figure shows how the days detected as failing HVAC system are sudden increases in CO_2. Panel (b) in Figure 3 describes the coefficients from an event study estimation that describe the changes in CO_2 in the days immediately preceding and following a failure in ventilation systems using an event study approach. The estimates show that on the date where the ventilation system is broken, CO_2 levels in the room increase by about 40% compared to the average of the previous 5 days, and remain high (about 20% higher) in the day immediately after the brake down was detected, suggesting that sometimes HVAC systems cannot be fixed on the same day.

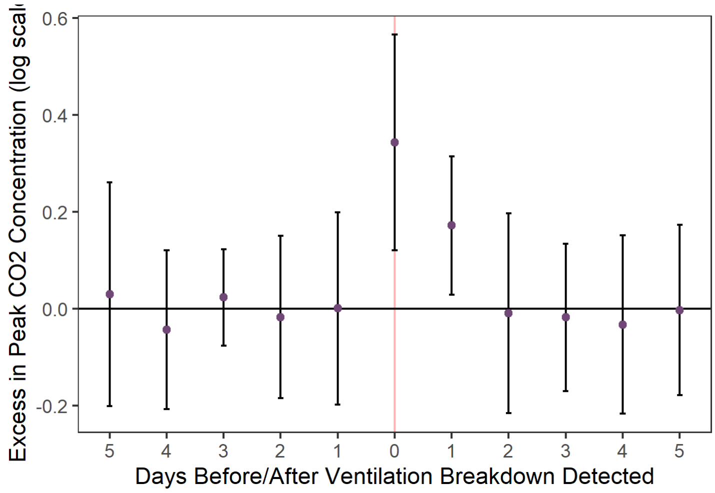
We then create dummy variables $Days\_Broken\_Vent_{ctv}$ corresponding to bins $v = (0, 1-2, 3-4, 5-6, >6)$ indicating the number of days with broken ventilation detected, in a regression

Figure 3: Average Peak CO2 Concentration and Mean Noise in classrooms on days with and without broken ventilation

(a) Detection ventilation failures



(b) Changes in CO_2 in days before and after ventilation failure events detected



Note: Panel (a) shows daily peak CO_2 concentration levels for each classroom in each academic term when ventilation is functioning normally (blue) and when it is broken (red), as identified by our algorithm. Panel (b) shows results of an event study identifying the excess levels of peak CO_2 in days when a ventilation breakdown has been detected (day 0) and during the following 5 days, compared to the average level during the previous 5 days.

including the same fixed effects as in our main regression specification (Eq. 1):

$$\widehat{CO2_{ct}} = \sum \xi_v Days\_Broken\_Vent_{ctv} + \Gamma X_{ct} + \alpha_{id} + \alpha_c + \alpha_t + \alpha_l + \varepsilon_{idt}, \tag{2}$$

Second stage:

$$Score_{idt} = \beta \widehat{CO2_{ct}} + \Gamma X_{ct} + \alpha_{id} + \alpha_c + \alpha_t + \alpha_l + \varepsilon_{idt}. \tag{3}$$

Table 6 shows the coefficients in our main regression, using the instrumented school-term average concentrations CO_2 as regressors. At the bottom of the table, we include the coefficients associated with the set of dummies describing the failures of the ventilation system in the first

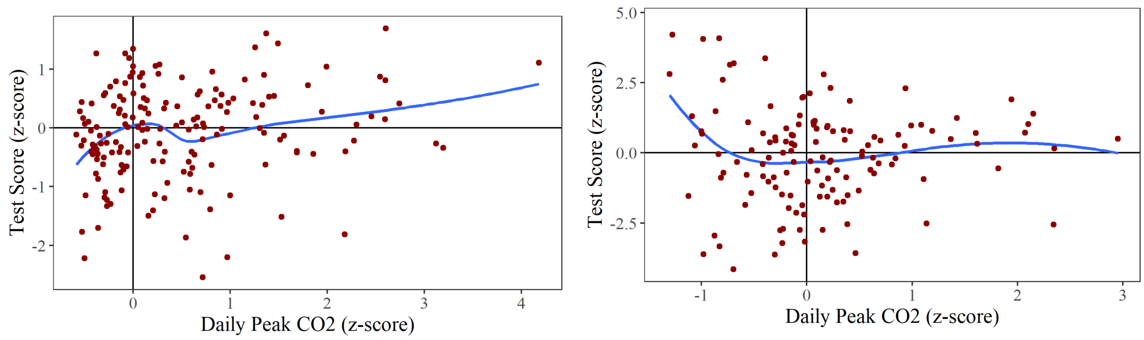
Table 6: Average Daily Peak CO2 Concentration on Standardized Test Scores

	Peak (z-score)	Average (z-score)	Peak> 2500ppm	Average> 1500ppm
CO2 (IV)	-0.247^{***}	-0.338^{***}	-0.392^{***}	-3.366
	(0.047)	(0.103)	(0.082)	(3.342)
Fixed Effects				
Student by Domain	Y	Y	Y	Y
Period	Y	Y	Y	Y
Classroom	Y	Y	Y	Y
Proficiency	Y	Y	Y	Y
Controls				
IEQ Parameters	Y	Y	Y	Y
Age	Y	Y	Y	Y
Class Size	Y	Y	Y	Y
Obs.	32,442	32,442	32,442	32,442
Adj. R^2	0.745	0.743	0.745	0.675
First Stage				
1-2 days broken	0.618^{***}	0.564^{**}	0.008	0.050
	(0.222)	(0.225)	(0.077)	(0.087)
3-4 days broken	2.097^{***}	1.122^{***}	1.118^{***}	0.136
	(0.178)	(0.196)	(0.057)	(0.115)
> 6 days broken	1.077^{***}	1.381^{***}	0.181	0.038
	(0.285)	(0.279)	(0.113)	(0.098)
F-statistic	44.014	17.568	166.411	0.363
p-val	0.0000	0.0000	0.0000	0.8349

Note: This table presents results for our main specification when instrumenting CO_2 concentration levels during the learning period using dummies indicating the number of days that the ventilation in the school was detected as broken by our algorithm. Each column shows results from instrumenting a different measure of CO_2 concentration as indicated in the column title. Significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

stage. The high value associated with the F-statistic suggests the high predictive power of the ventilation breakdown events on the school-term average CO_2 concentrations in the classroom, after controlling for classroom fixed effects and the remaining factors listed in our main regression (Eq. 1). The estimates associated with the instrumented CO_2 concentrations are similar in magnitude and statistical significance as in our main results, indicating that our main effects are mainly driven by ventilation breakdowns, identified as sudden jumps in the series of CO_2 in classrooms. Similarly, Supplementary Table B.3 replicates the heterogeneity analysis.

Figure 4: Correlation Between Estimates of Teacher Fixed Effects on Average Scores and Daily Peak CO_2



(a) Only Teacher FEs

(b) Teacher, School, Domain, Period, and Proficiency FEs

Notes: The figure describes the correlation between the estimated teacher fixed effects in Equation 4 and the estimated teacher fixed effects in Equation 5. Panel A. includes no controls in the regression models. Panel B. controls for school, test domain, period and the student proficiency. Fitted lines were estimated using a locally weighted regression with a smoothing parameter $\alpha = 0.8$. The Pearson correlation coefficients between are 0.18 and -0.12, respectively, and neither was statistically significant.

5.3 Teacher Effects on Indoor Climate

A common identification challenge in the identification of impacts of school infrastructure has to do with its correlation with teacher quality. In our study, the correlation between teacher quality and air quality challenges the interpretation of our results as air quality impacts.

In this subsection, we estimate the extent to which systematic differences in teaching quality correlate with classroom's CO_2 levels. In this analysis, we link the average scores of students taught by a given teacher with the average levels of CO_2 in the classrooms allocated to the teacher over the sample period of our study. We compute the teacher-specific differences in the distribution of tests scores of students and classroom CO_2 levels from teacher fixed effects in the following two regression models:

$$CO2_{ctp} = Teacher_p^{CO_2} + \varepsilon_{itp} \tag{4}$$

and

$$Score_{ctdp} = Teacher_p^{tests} + \varepsilon_{itdp} \tag{5}$$

where $Teacher_p^{CO_2}$ in Eq. 4 describes the teacher fixed-effects, containing dummy variables describing the individual differences across teachers in the CO_2 levels in classrooms in the school terms where the teacher was allocated to the classroom. In Eq. 5, $Teacher_p^{tests}$ describes the teacher fixed-effects, containing dummy variables describing the individual differences across teachers in the test scores of their students. Figure 4 displays the link between the two sets of fixed effects, to test for the link between student test scores and classroom CO_2. The lack of relationship (or rather, a slightly positive relationship) between the two set of fixed effects indicates that teacher quality is not linked to the levels of CO_2 in the classroom. This results suggest that teachers are not a key driver of the impacts of CO_2 on the performance of students.

6 Long-run Effects: Effects on high school advice

In our final analysis for learning achievements, we relate indoor environmental conditions during the learning period with test results that students take in their final year of primary education. These tests largely determine the advice that teachers will give about the type of high school a student should attend in the future.

Students in the Netherlands can attend different types of high schools, according to their performance in primary school. The first type of high schools, namely "VMBO", require lowest performance in their primary school education, and it is more orientated towards practical set of skills. The second level, namely "HAVO", requires an intermediate academic level and focuses on teaching a mixed of applied and conceptual skills. And finally, for students with the highest level of academic achievement, "VWO" high schools will teach more conceptual subjects with most of these students entering research universities after high school completion (Supplementary Figure C.4 displays the distribution of test scores in the final test separately for each secondary school track).

All tests results since third grade of primary education provide the teacher with elements to determine the final advice about the type of high school that the student should attend. In their final year of primary education, students take a final set of tests conducted with the sole purpose of confirming the teacher's advice. Teachers calculate a comprehensive score combining all these test results and compare it with their own advice. If the advice shown by the tests is to a higher high school level than advised, then teachers are obliged to review their decision, if in turn the opposite occurs, there are no consequences and the original advice remains.

In a cross-section analysis, we relate the school-term CO_2 in the classroom preceding the final test with the result in the final test, the result in each part of the test (i.e. Language and Mathematics), and with the probability that the student is given the advice to attend a HAVO or VWO high school. Table (7) shows results for this analysis, including controls for the other IEQ variables and the class size.[17] Column (1) shows that a one standard deviation increase in

[17]There are no fixed effects possible to control for since there is only one group 8 per school, and we observe only one period for each student taking this test.

Table 7: Average Daily Peak CO2 Concentration on 8th Grade Test Scores and on Probability of Student Receiving Advice to Join VWO/HAVO Level High School

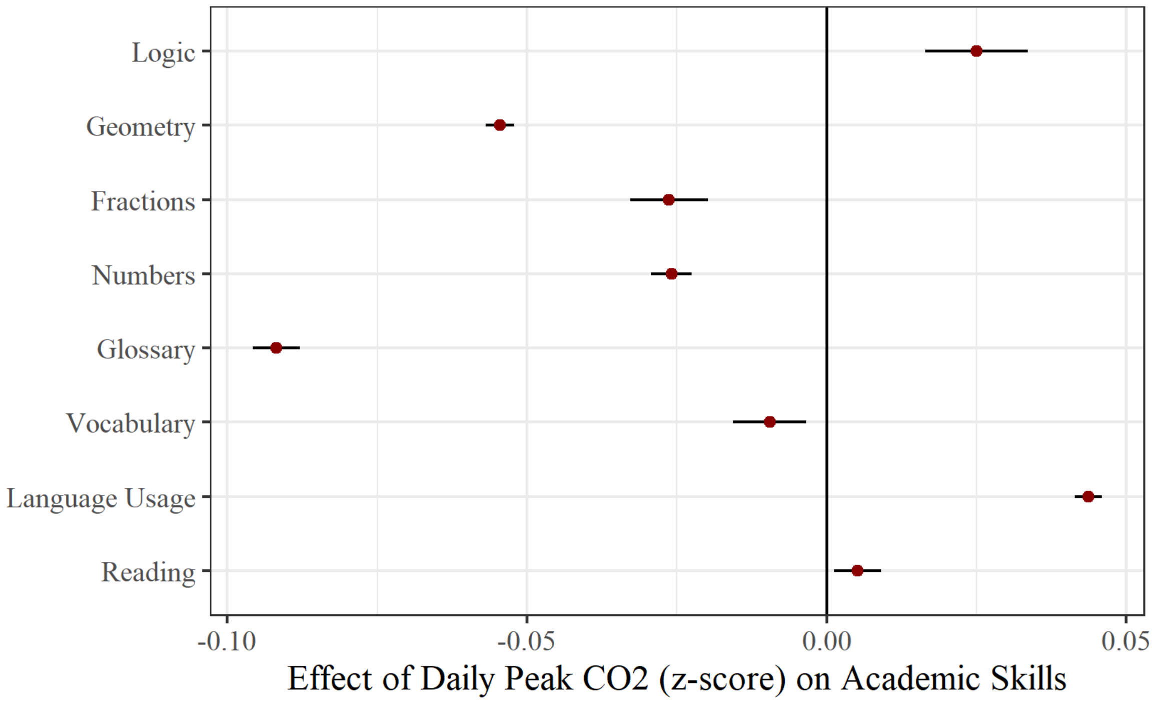
	Full Score (logs)	Language	Mathematics	Advice to VWO/HAVO (Prob.)	Skill Score (logs)
	(1)	(2)	(3)	(4)	(5)
CO2 (z-score)	-0.019***	-0.022***	-0.015***	-0.127***	-0.020***
	(0.001)	(0.000)	(0.003)	(0.003)	(0.003)
Controls					
IEQ Controls	Y	Y	Y	Y	Y
Class Size	Y	Y	Y	Y	Y
Obs.	193	193	193	193	1615
R^2	0.020	0.025	0.025	0.031	0.010

Note: This table shows results of a cross-section analysis relating CO_2 concentration levels during the learning term for 193 students taking the final test that determines whether teachers will review their advice about which level of high school they should attend. The table gives the impact on the log final score obtained by the student (Column (1)), on the log scores obtained for the Language (Column (2)) and Mathematics (Column (3)) parts, and on the probability that the student is given the advice to attend to either of the two higher levels of high school (i.e. VWO and HAVO) (Column (4)). Column 5 provides an estimate for the impact of ventilation on the scores achieved for each individual skill when observations for all scores of each student are pooled together. Significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

daily average peak values of CO_2 during the learning period preceding the test reduce final test results by 2%. This result is driven by both parts of the test (i.e. Language and Mathematics) (Columns (2) and (3)). Column (4) in Table (7) shows that a one standard deviation increase in CO_2 concentration is associated with a 13% lower chance to be recommended to a higher high school level.

The data collected on these final tests allow us to look more closely into which specific skills are being relatively more affected by poor indoor air quality. Within each part of the test, different skills (i.e. logic, geometry, fractions, and numbers for mathematics; and glossary, vocabulary, language usage, and reading for the language part) are separately evaluated by assigning a score to each of them. To gain more statistical power we regress the score obtain by the student in each skill and find an average effect of a 2% decline in test scores (Table (7), Column (5)) controlling for the skill fixed effects. In addition, we individually relate the CO_2 exposure measure during the learning period to the log-scores given to each skill. Figure (5) plots coefficient estimates of these effects with corresponding 5% significance levels error bars for each skill. All mathematics skills (except logic formulations) are being affected by poor levels of air quality while the students learn these skills. A one standard deviation increase in CO_2 concentration is associated with a reduction in mathematics skills scores that range between 2.5% and 5%. However, CO_2 concentration is also associated with an increase in logic skills. For

Figure 5: Correlation Between Estimates of Teacher Fixed Effects on Average Scores and Daily Peak CO_2



Notes: The figure describes the correlation between the estimated teacher fixed effects in Equation 4 and the estimated teacher fixed effects in Equation 5. Panel A. includes no controls in the regression models. Panel B. controls for school, test domain, period and the student proficiency.

language skills, results show that the main driver of the reduction in scores is glossary with an almost 10% decrease associated to a one standard deviation increase in CO_2 concentration.

7 Discussion and Conclusion

This paper provides the first quasi-experimental evidence on how air quality conditions in the classroom affect learning outcomes. We design and implement a large field study, deploying an indoor sensing network in 27 elementary schools, continuously monitoring the indoor air quality conditions in 235 individual classrooms, covering some 4,000 children aged 6 to 12. We document large variation in indoor air quality across schools, classrooms and over time. Such variance cannot be explained (add here percentage) by classroom or school characteristics, or ambient weather conditions – strengthening the need to measure those factors indoors, and on a continuous basis. Using within-child variation in exposure to air quality conditions, we document that when children are systematically exposed to poor air quality during a prolonged learning period, their cognitive development is slowed down, as measured by the improvement on standardized tests made by the same student.

To put our findings into perspective, we can compare them to the effects documented for other school interventions, for example, the Active Living Program (Golsteyn et al., 2020) or the Early Training Program (Anderson, 2008). Of course, these intervention programs are not directly comparable, given the non-experimental setting of this paper, but the comparison of average treatment effects is nonetheless helpful in understand the magnitude of the documented results. The effects of the Active Living Program, an experiment aimed at stimulating physical activity on Dutch primary schools, showed a negative effect of 0.06 standard deviations, using a comparable sample and setting. Clearly, the effects of *passive* exposure to poor air quality conditions in the classroom trump the negative effects of a program aimed at stimulating physical activity (see Golsteyn et al. (2020)). Anderson (2008) documents average impact outcomes across programs of 0.28 standard deviations, while Deming (2009) documents an average impact of 0.23 standard deviations for the Head Start program. Although the setting of these programs is fundamentally different than ours, these simple comparisons provide some indication that the economic magnitude of the findings in this paper is quite substantial.

This study highlights the understudied role that school facility conditions play in generating educational outcomes. It also highlights the extent to which disparities in such physical environments contribute to inequality in educational outcomes. The results allow us to estimate the benefits of potential public investment, for example in school HVAC systems, that may help reduce such gaps. A variety of recent papers have examined connections between investment in school facility and academic achievements, but the mechanisms explaining that connection remain a matter of speculation. In addition, this paper complements the growing literature on the impacts of pollution, temperature, and academic performance. The existing evidence depends on measurements of outdoor temperature or ambient air quality to assess the relationship between environmental conditions and student performance, relying on strong assumptions to extrapolate outdoor measurements to indoor exposure. The use of sensors inside the classrooms where children are learning and being tested allows us to assess exposure to specific pollutants with an unprecedented accuracy.

Our results also yield several important policy lessons. The airborne transmission of the SARS-COV-2 virus has elevated the salience of indoor environmental quality as an important factor to prevent the spread of the disease. Many countries have begun preparing major investment outlays to improve ventilation, via modernization or installation of HVAC systems, or upgrading the standards of ventilation in buildings. Schools buildings are among the major targets in many nations' building portfolios, due to the high density of children in classrooms, and the long need for disrepair or installation of schools' air treatment systems. Our results suggest that upgrades in indoor air quality in schools is relevant besides reduction of the spread of viral diseases, such as COVID-19, and supports children's cognitive capacities, precisely at a critical age for human capital accumulation and skill formation.

References

Aaronson, D. and B. Mazumder (2011): "The impact of Rosenwald schools on black achievement," Journal of Political Economy, 119, 821–888.

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017): "When should you adjust standard errors for clustering?" Tech. rep., National Bureau of Economic Research.

Allen, J. G., P. MacNaughton, U. Satish, S. Santanam, J. Vallarino, and J. D. Spengler (2016): "Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments," Environmental health perspectives, 124, 805–812.

Anderson, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," Journal of the American statistical Association, 103, 1481–1495.

ASHRAE (2022): "ASHRAE Position Document on Indoor Carbon Dioxide," .

Asseng, S., D. Spänkuch, I. M. Hernandez-Ochoa, and J. Laporta (2021): "The upper temperature thresholds of life," The Lancet Planetary Health, 5, e378–e385.

Bakó-Biró, Z., D. J. Clements-Croome, N. Kochhar, H. B. Awbi, and M. J. Williams (2012): "Ventilation rates in schools and pupils' performance," Building and environment, 48, 215–223.

BBC (2020): "Coronavirus: Germany improves ventilation to chase away Covid," .

Borghans, L., B. H. Golsteyn, and A. Stenberg (2015): "Does expert advice improve educational choice?" PloS one, 10, e0145378.

Chang, T., J. Graff Zivin, T. Gross, and M. Neidell (2016): "Particulate pollution and the productivity of pear packers," American Economic Journal: Economic Policy, 8, 141–69.

Chang, T. Y., 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, 151–72.

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

Cochrane, E. (2021): "Senate Passes $1 Trillion Infrastructure Bill, Handing Biden a Bipartisan Win," The New York Times.

Currie, J., E. A. Hanushek, E. M. Kahn, M. Neidell, and S. G. Rivkin (2009): "Does pollution increase school absences?" The Review of Economics and Statistics, 91, 682–694.

Currie, J. and M. Neidell (2005): "Air pollution and infant health: what can we learn from California's recent experience?" The Quarterly Journal of Economics, 120, 1003–1030.

Currie, J. and R. Walker (2011): "Traffic congestion and infant health: Evidence from E-ZPass," American Economic Journal: Applied Economics, 3, 65–90.

Deming, D. (2009): "Early childhood intervention and life-cycle skill development: Evidence from Head Start," American Economic Journal: Applied Economics, 1, 111–34.

Du, B., M. C. Tandoc, M. L. Mack, and J. A. Siegel (2020): "Indoor CO2 concentrations and cognitive function: A critical review," Indoor air, 30, 1067–1082.

Duflo, E. (2001): "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," American economic review, 91, 795–813.

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, 36–65.

EPA (2022): "Clean Air in Buildings Challenge," .

Fisk, W. J. (2017): "The ventilation problem in schools: literature review," Indoor Air, 27, 1039–1051.

G.A.O. (2020): "K-12 EDUCATION. School Districts Frequently Identified Multiple Building Systems Needing Updates or Replacement," .

Golsteyn, B. H., M. W. Jansen, D. H. Van Kann, and A. M. Verhagen (2020): "Does Stimulating Physical Activity Affect School Performance?" Journal of Policy Analysis and Management, 39, 64–95.

Graff Zivin, J. and M. Neidell (2012): "The impact of pollution on worker productivity," American Economic Review, 102, 3652–73.

——— (2013): "Environment, health, and human capital," Journal of Economic Literature, 51, 689–730.

——— (2018): "Air pollution's hidden impacts," Science, 359, 39–40.

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

Hammer, M. S., T. K. Swinburn, and R. L. Neitzel (2014): "Environmental noise pollution in the United States: developing an effective public health response," Environmental health perspectives, 122, 115–119.

Hanushek, E. A. (2003): "The failure of input-based schooling policies," The economic journal, 113, F64–F98.

Heckman, J. J. (2006): "Skill formation and the economics of investing in disadvantaged children," Science, 312, 1900–1902.

Howard-Jones, P. A., E. V. Washbrook, and S. Meadows (2012): "The timing of educational investment: A neuroscientific perspective," Developmental Cognitive Neuroscience, 2, S18–S29.

Jackson, C. K., R. C. Johnson, and C. Persico (2016): "The effects of school spending on educational and economic outcomes: Evidence from school finance reforms," The Quarterly Journal of Economics, 131, 157–218.

Kahn, M. E. and P. Li (2020): "Air pollution lowers high skill public sector worker productivity in China," Environmental Research Letters, 15, 084003.

Knittel, C. R., D. L. Miller, and N. J. Sanders (2016): "Caution, drivers! Children present: Traffic, pollution, and infant health," Review of Economics and Statistics, 98, 350–366.

Künn, S., J. F. Palacios Temprano, and N. Pestel (2019): "Indoor Air Quality and Cognitive Performance," IZA Discussion Paper Series.

Lichter, A., N. Pestel, and E. Sommer (2017): "Productivity effects of air pollution: Evidence from professional soccer," Labour Economics, 48, 54–66.

Liu, C., R. Chen, F. Sera, A. M. Vicedo-Cabrera, Y. Guo, S. Tong, M. S. Coelho, P. H. Saldiva, E. Lavigne, P. Matus, et al. (2019): "Ambient particulate air pollution and daily mortality in 652 cities," New England Journal of Medicine, 381, 705–715.

Martorell, P., K. Stange, and I. McFarlin Jr (2016): "Investing in schools: capital spending, facility conditions, and student achievement," Journal of Public Economics, 140, 13–29.

Meyer, S. and M. Pagel (2017): "Fresh air eases work–The effect of air quality on individual investor activity," Tech. rep., National Bureau of Economic Research.

Moretti, E. and M. Neidell (2011): "Pollution, health, and avoidance behavior evidence from the ports of Los Angeles," Journal of human Resources, 46, 154–175.

Neilson, C. A. and S. D. Zimmerman (2014): "The effect of school construction on test scores, school enrollment, and home prices," Journal of Public Economics, 120, 18–31.

NOAA (2021): "Climate Change: Atmospheric Carbon Dioxide," .

Palacios, J., P. Eichholtz, M. Willeboordse, and N. Kok (2020): "Indoor environmental quality and learning outcomes: protocol on large-scale sensor deployment in schools," BMJ open, 10, e031233.

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

Persico, C. L. and J. Venator (2019): "The effects of local industrial pollution on students and schools," Journal of Human Resources, 0518–9511R2.

Rijksoverheid (2020): "360 miljoen voor verbetering ventilatie in scholen," .

Roth, S. (2017): "Air pollution, educational achievements, and human capital formation," IZA World of Labor.

——— (2018): "The effect of indoor air pollution on cognitive performance: Evidence from the UK," .

Seppanen, O., W. J. Fisk, and Q. Lei (2006): "Effect of temperature on task performance in office environment," .

Simonsohn, U., J. P. Simmons, and L. D. Nelson (2020): "Specification curve analysis," Nature Human Behaviour, 4, 1208–1214.

Spratford, M., E. A. Walker, and R. W. McCreery (2019): "Use of an application to verify classroom acoustic recommendations for children who are hard of hearing in a general education setting," American journal of audiology, 28, 927–934.

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

WHO (2014): "7 million premature deaths annually linked to air pollution," World Health Organization, Geneva, Switzerland.

Wikipedia (2022): "Parkstad Limburg — Wikipedia, The Free Encyclopedia," http://nl.wikipedia.org/w/index.php?title=Parkstad%20Limburg&oldid=60826057, [Online; accessed 16-May-2022].

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

Appendix

A Main Results Using Average Mean CO2 Concentration

Table A.2: Average Daily Mean CO2 Concentration on Standardized Test Scores

	Dependent Variable: Standardized Test Scores				
	(1)	(2)	(3)	(4)	(5)
CO2 (z-score)	-0.053	-0.062	-0.088*	-0.088*	-0.120**
	(0.035)	(0.046)	(0.047)	(0.047)	(0.052)
PM10 (z-score)		-0.059	-0.042	-0.009	0.015
		(0.058)	(0.060)	(0.057)	(0.072)
Temperature (z-score)		-0.010	-0.008	-0.008	-0.027
		(0.027)	(0.028)	(0.026)	(0.040)
Humidity (z-score)		0.101	0.109	0.088	-0.013
		(0.074)	(0.074)	(0.076)	(0.105)
Age [8-10)			-0.029	-0.023	-0.031
			(0.039)	(0.039)	(0.037)
Age [10-14)			-0.022	-0.019	-0.019
			(0.025)	(0.025)	(0.024)
Class Size				0.020***	0.023***
				(0.007)	(0.008)
Class Size Sq.				-0.000**	-0.000**
				(0.000)	(0.000)
Avg. Noise (z-score)				-0.041	-0.062*
				(0.031)	(0.037)
Fixed Effects					
Student by Domain	Y	Y	Y	Y	Y
Period	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y
Proficiency	N	N	Y	Y	Y
School by Period	N	N	N	N	Y
Obs.	37,451	37,451	37,451	37,451	37,451
Adj. R^2	0.741	0.741	0.750	0.750	0.754

All models relate average daily mean CO2 concentration in the classroom with standardized test scores. Column (1) provides results for a model with fixed effects for students by each domain and periods. Column (2) provides results for a model that controls for observed and unobserved physical conditions within the classroom using average daily peaks of other IAQ variables and classroom fixed effects. Column (3) gives results for a model that adds characteristics of the groups and students, i.e. group size, age, and the average daily mean noise within the classroom. Finally, the last column adds school by period fixed effects to use only variation within each specific school and rule out unobserved confounders at the school level. Clustered standard errors at the classroom by period level are shown in brackets. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table A.3: Average Daily Peak CO2 Concentration on Standardized Test Scores by Day of the Week

	by Day of Week During Learning Period				
	Mondays	Tuesdays	Wednesdays	Thursdays	Fridays
CO_2 (z-score)	-0.075*	-0.072**	0.036	-0.065**	-0.071*
	(0.043)	(0.034)	(0.040)	(0.032)	(0.038)
Fixed Effects					
Student by Domain	Y	Y	Y	Y	Y
Period	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y
Proficiency	Y	Y	Y	Y	Y
Controls					
IEQ Parameters	Y	Y	Y	Y	Y
Age	Y	Y	Y	Y	Y
Class Size	Y	Y	Y	Y	Y
Obs.	34,675	36,710	33,764	36,564	35,344
Adj. R^2	0.7559	0.7523	0.7736	0.7503	0.7515

Each model relates standardized test scores with only the average daily peak CO_2 measured during each of the days during the learning period indicated in the column name. The last column includes daily average peak CO_2 observed in all days of the week. Significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table A.4: Falsification Test for CO2 Concentration During Learning Period and Standardized Test Scores

Panel A: Test for the effect of contemporaneous CO_2 concentration in a randomly chosen different classroom at the same school

	Peak (z-score)	Average (z-score)	Peak> 2500ppm	Average> 1500ppm
CO_2	0.017	0.032	0.051	-0.115*
	(0.036)	(0.037)	(0.075)	(0.059)
Obs.	37,451	37,451	37,451	37,451
Adj. R^2	0.745	0.745	0.745	0.745

Panel B: Test for the effect of contemporaneous CO_2 concentration in a random classroom at a different school

	Peak (z-score)	Average (z-score)	Peak> 2500ppm	Average> 1500ppm
CO_2	-0.029	-0.023	-0.153	-0.107
	(0.041)	(0.041)	(0.114)	(0.134)
Obs.	37,451	37,451	37,451	37,451
Adj. R^2	0.757	0.757	0.757	0.757
Fixed Effects				
Student by Domain	Y	Y	Y	Y
Period	Y	Y	Y	Y
Classroom	Y	Y	Y	Y
Proficiency	Y	Y	Y	Y
Controls				
IEQ Parameters	Y	Y	Y	Y
Age	Y	Y	Y	Y
Class Size	Y	Y	Y	Y

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

B Broken ventilation: robustness

B.1 Using other CO_2 thresholds

Table B.2: Average Daily Peak CO2 Concentration on Standardized Test Scores

Second Stage	(1)	(2)	(3)	(4)
CO2 (z-score, IV)	-0.322^{**}	-0.247^{***}	-0.127	-0.083
	(0.1310)	(0.0469)	(0.0981)	(0.1191)
Fixed Effects				
Student by Domain	Y	Y	Y	Y
Period	Y	Y	Y	Y
Classroom	Y	Y	Y	Y
Proficiency	Y	Y	Y	Y
Controls				
IEQ Parameters	Y	Y	Y	Y
Age	Y	Y	Y	Y
Class Size	Y	Y	Y	Y
Obs.	32,442	32,442	32,442	32,442
Adj. R^2	0.7434	0.7448	0.7457	0.7457
First Stage				
1-2 days broken	0.5601^{*}	0.6123^{**}	0.6604^{***}	0.2413^{*}
	(0.3167)	(0.3107)	(0.1831)	(0.1313)
3-4 days broken		1.2605^{***}	0.5851	0.4633^{**}
		(0.3349)	(0.5285)	(0.2340)
5-6 days broken			0.7410^{***}	0.7192^{**}
			(0.2170)	(0.3543)
$>$ 6 days broken		1.0998^{***}	0.9118^{***}	0.7738^{**}
		(0.3390)	(0.2285)	(0.3067)
F-stat	1.1572	44.0141	9.2822	7.2085
p-val	0.3291	0.0000	0.0000	0.0000

These IV estimations use 1800ppm (1), 1400ppm (2), 1000ppm (3), and 600ppm (4) as thresholds to determine the ventilation breakdown. Significance is given by $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

B.2 Temperature and noise during ventilation breakdown

Figure B.2: Average Peak Temperature at Days Before and After Ventilation Breakdown Detected

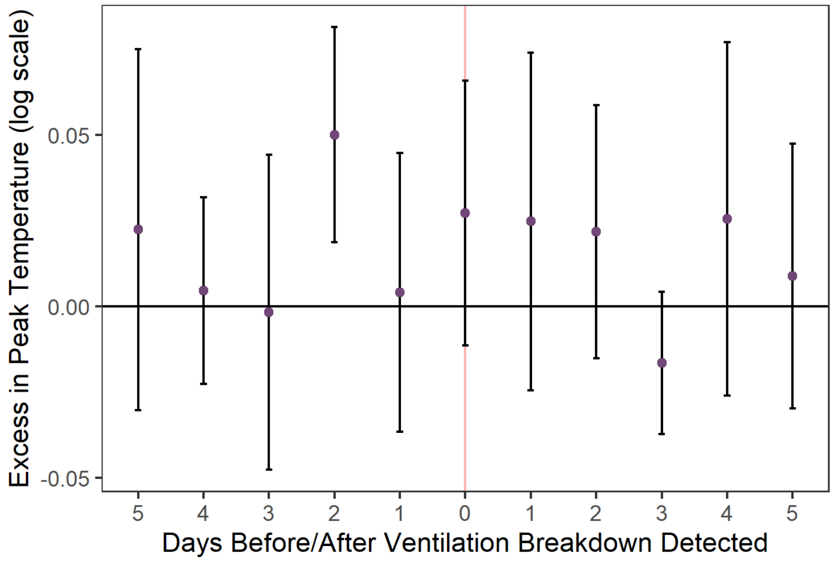


Figure B.3: Average Peak Noise at Days Before and After Ventilation Breakdown Detected

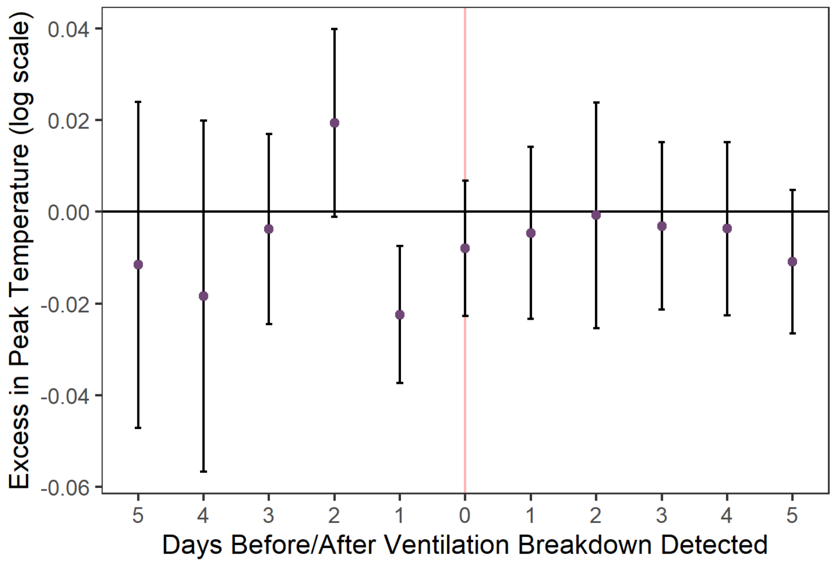


Table B.3: Average Daily Peak CO2 Concentration on Standardized Test Scores

	by Domain			by Age		
Second Stage	Spelling	Maths	Reading	[5-7]	[8-10]	[11-13]
CO2 (z-score, IV)	-0.161^{**}	-0.360^{***}	-0.308^{***}	0.081	-0.236^{***}	-0.206^{*}
	(0.080)	(0.054)	(0.077)	(0.270)	(0.051)	(0.118)
Fixed Effects						
Student	Y	Y	Y	Y	Y	Y
Student by Domain	N	N	N	Y	Y	Y
Period	Y	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y	Y
Proficiency	Y	Y	Y	Y	Y	Y
Controls						
IEQ Parameters	Y	Y	Y	Y	Y	Y
Age	Y	Y	Y	Y	Y	Y
Class Size	Y	Y	Y	Y	Y	Y
Obs.	9,282	9,028	7,361	8,249	11,674	12,519
Adj. R^2	0.752	0.784	0.743	0.709	0.777	0.763
First Stage						
1-2 days broken	0.726^{***}	0.493^{***}	0.292^{***}	0.720^{***}	0.213^{**}	0.993^{***}
	(0.196)	(0.159)	(0.091)	(0.232)	(0.103)	(0.359)
3-4 days broken	2.001^{***}	2.178^{***}	1.968^{***}	1.452^{***}	2.071^{***}	
	(0.142)	(0.179)	(0.116)	(0.188)	(0.131)	
> 6 days broken	1.285^{***}	1.045^{***}	0.691^{***}			1.406^{***}
	(0.286)	(0.273)	(0.216)			(0.477)
F-stat	50.506	42.556	90.662	15.228	67.006	2.201
p-val	0.0000	0.0000	0.0000	0.0000	0.0000	0.0697

Note: This table presents results for our main specification when instrumenting average daily peak CO_2 levels during the learning period using dummies indicating the number of days that the ventilation in the school was detected as broken by our algorithm. Each column shows results for specific subsamples as indicated in the column title. Significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

C Data Coverage and Algorithm to Detect Occupancy

C.1 Data Coverage

Figure C.2: Deployment and Coverage of Dates by Sensor

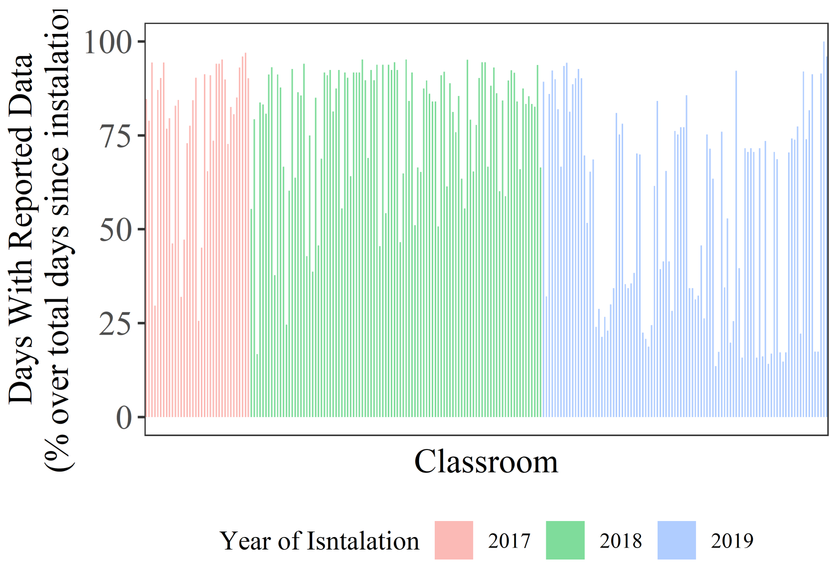


Figure C.3: Coverage of Sensors by Date

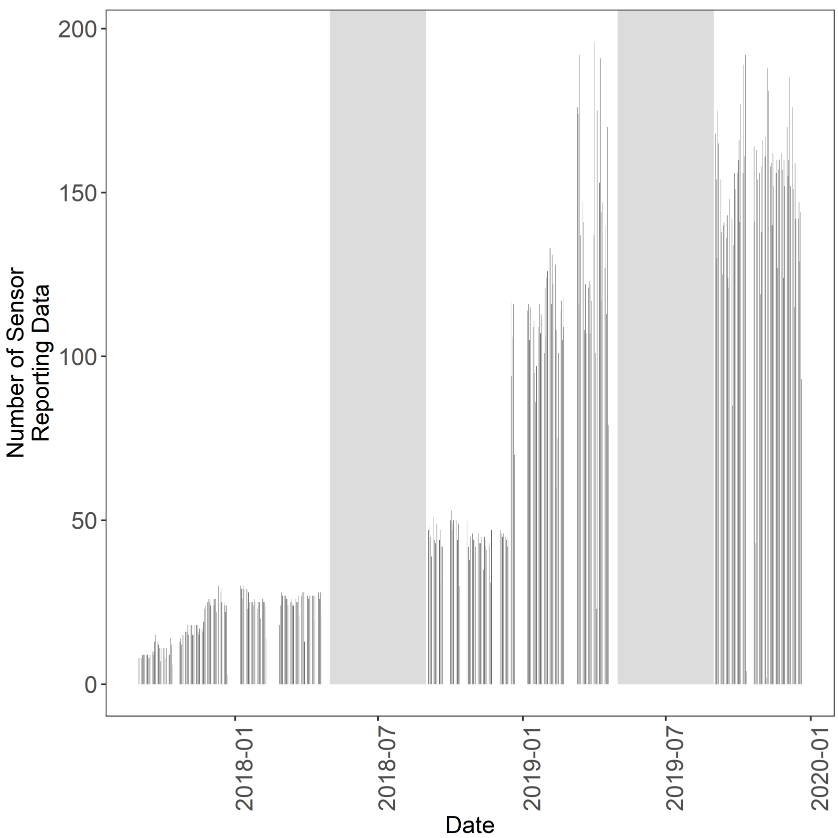
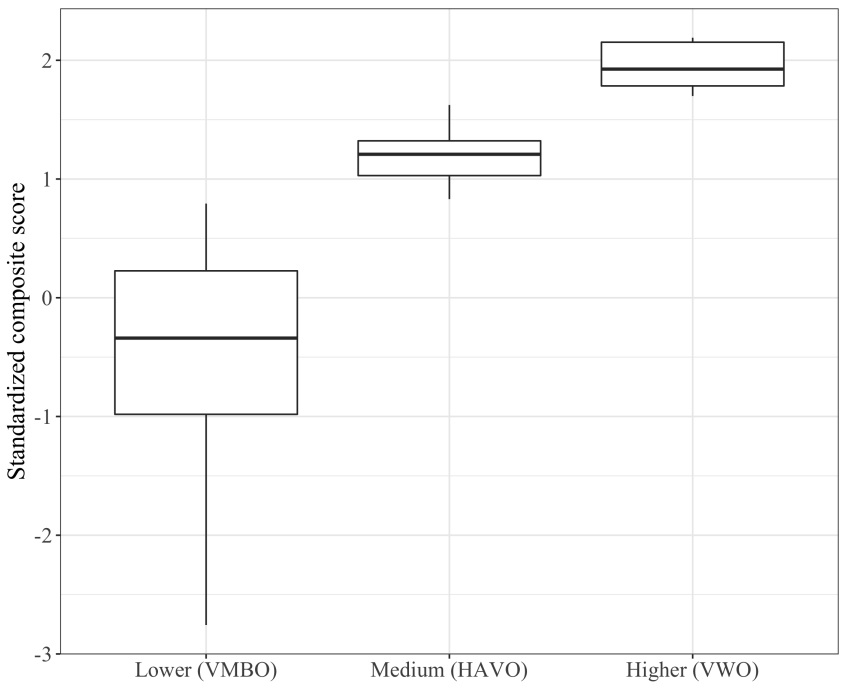


Figure C.4: Composite standard score (all tests) and high school level advice



C.2 Secondary School Advice

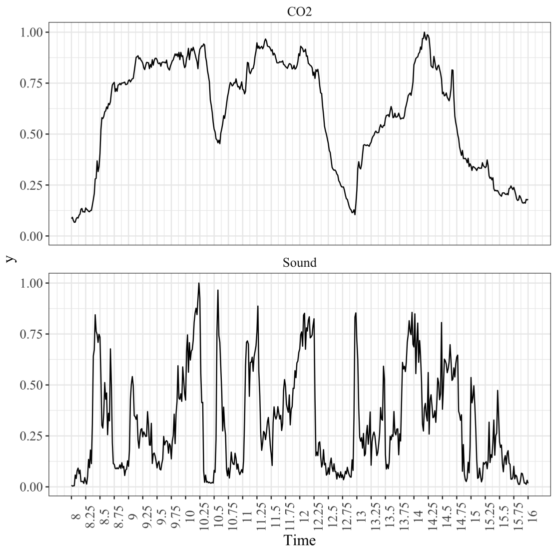
C.3 Algorithm to Detect Occupancy

The algorithm used to determine entry and exit in the classroom searches for increases in CO_2 concentration that are sustained in time while looking for a spike in the sound to detect the exact time when the children have entered the classroom. Sustained increases in CO_2 show that the room is no longer empty and that the door is closed (the rate of CO_2 generation is higher than the rate of air exchange), while a spike in sound indicates that students have entered the classroom and are (in the process of) sitting down. To detect when the classroom is empty, the algorithm searches for a sustained decrease in CO_2 concentration (the rate of CO_2 generation is now lower than the rate of air exchange, as the door is opened and there are fewer students inside the classroom), while also looking for a spike in sound (students make noise when exiting the classroom).

As described in the main text, the algorithm makes use of regularities in the behavior of both CO_2 and sound in the event of children entering or exiting the classroom. Those regularities are easily spotted in Figure C.5. The graph plots how CO_2 concentration and sound decibels move between 8am and 4pm (school hours). One can identify how accumulation of CO_2 starts and sound spikes at the morning entry between 8:15am and 8:30am (8.25 to 8.5 in the graph). For exits, the opposite occurs for CO_2, while sound also spikes as is evident during the first break at 10:15am (10.25 in the graph).

Using these regularities, the algorithm first detects all series of j consecutive minutes showing a CO_2 increase (decrease) during the school day (8am-4pm). After these series are found, we label their first minute as a candidate entry or exit if at any of those minutes in the series, we observe a spike in sound above a threshold s. Once all entry and exit candidates are labelled, the algorithm orders them by time in decreasing order, and retains the first of all consecutive entries before and exit occurs, and the first of all consecutive exits before an entry takes place, such that in order to get an exit, an entry must have been labeled before and vice-versa (except for the first entry and last exit of the day, of course).

Figure C.5: Example of observed levels of CO_2 and Sound (normalized) across a school day

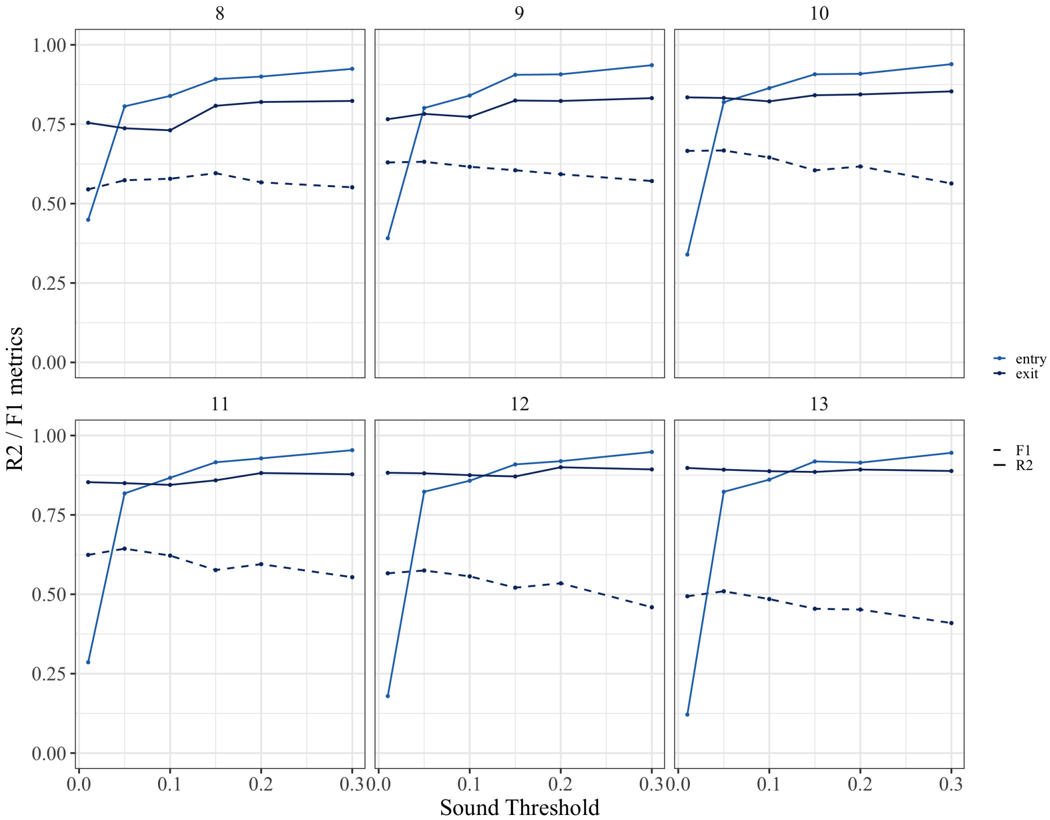


Note: This graphs describe how CO_2 (above) and sound (below) move along a school day (8am to 4pm) inside a particular classroom. It clearly shows how CO_2 starts accumulating and sound spikes when children enter the room and the opposite happens when they exit.

We assess the algorithm's sensitivity and accuracy relative to different values of j and s to determine which give the optimal result. For this purpose, we labeled the observed entries and exits during the school day for nearly 500 graphs of CO_2 and sound series in different days and schools (randomly chosen with monthly stratification). We then compare those labels to the algorithm predictions.

To measure the algorithm's performance we use two metrics: an F1 indicator and the algorithms R^2. The F1, widely used in machine learning contexts, takes the geometric mean of two ratios: (i) the number of correctly predicted entries over all predicted entries; and (ii) the number of correctly predicted entries over all observed entries. This indicator gives a sense of the algorithm's sensitivity as it assesses the proportions of false entry/exit predictions and those of unpredicted but observed entry/exit. However, this measure is silent on the algorithm's accuracy in predicting the exact time at which entry and exits took place. Hence, we assess the algorithm's prediction accuracy using an R^2 coefficient, a well known indicator to assess predictive power. Figure C.6 shows the resulting F1 and R^2 for $j = 7, 8, 9, 10, 11, 12$ minutes and for $s = 0.01, 0.05, 0.1, 0.2, 0.3$ normalized dBA.

Figure C.6: Algorithm performance (F1 and R^2)



The combination of both indicator values suggests that $j = 10$ and $s = 0.05$ predict entries and exits both most accurately as well as most frequently. The highest point achieved by the F1 indicator is at this point (upper right plot), while the highest R2 for both, entries and exits, is also achieved at the same point. We therefore construct our data set on indoor environmental quality using these parameters in the algorithm to predict when students are inside or outside of the classroom.