WORKING
PAPER
SERIES

Title

Indoor Air Quality and Learning: Evidence from A Large Field Study in Primary Schools

Authors
Juan Palacios
Nicolas Duran
Nils Kok
Piet Eichholtz

Date
December 2022

MIT Center for
Real Estate Research
Paper No.
22/13

Center for Real Estate

Indoor Air Quality and Learning: Evidence from A Large Field Study in Primary Schools
Juan Palacios, Nicolas Duran, Nils Kok, and Piet Eichholtz
MIT Center for Real Estate Research Paper No. 22/13
December 2022

Abstract

Governments devote a large share of public budgets to construct, repair, and modernize school facilities. However, evidence on whether investments in the physical state of schools translate into better student outcomes is scant. In this study, we report the results of a large field study on the implications of poor air quality inside classrooms − a key performance measure of school mechanical ventilation systems. We continuously monitor the air quality (i.e., CO2), together with a rich set of indoor environmental parameters in 216 classrooms in the Netherlands. We link indoor air quality conditions to the outcomes on semi-annual nationally standardized tests of 5,500 children, during a period of five school terms (from 2018 to 2020). Using a fixed-effects strategy, relying on within-pupil changes in air quality conditions and test results, we document that exposure to poor indoor air quality during the school term preceding a test is associated with significantly lower test results: a one standard deviation increase in the school-term average daily peak of CO2 leads to a 0.11 standard deviation decrease in subsequent test scores. The estimates based on plausibly exogenous variation driven by mechanical ventilation system breakdown events confirm the robustness of the results. Our results add to the ongoing debate on the determinants of student human capital accumulation, highlighting the role of school infrastructure in shaping learning outcomes.

Juan Palacios
Massachusetts Institute of Technology and IZA

Nicolas Duran
University College London

Nils Kok
Maastricht University

Piet Eichholtz
Maastricht University

1 Introduction

Governments invest billions of dollars in the construction and modernization of school facilities on an annual basis. In the US alone, school infrastructure receives over $60 billion annually (Cornman et al., 2022), 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 $130 billion to the renovation, modernization, and construction of schools (Cochrane, 2021). Yet, many schools, in the US and beyond, are in some state of disrepair. For example, more than 40% of schools in the US rely on outdated heating, ventilation and cooling (HVAC) systems that need to be updated or replaced, a significant concern for facilities where children spend an average of eight hours per day (G.A.O., 2020; Rijksoverheid, 2020).[1]

However, little is known about the impact of deficient ventilation systems on educational outcomes. Failing ventilation leads to an increase in concentration of indoor air pollutant concentrations in classrooms (Fisk, 2017), and increases the risk of transmission of airborne diseases.[2]. Lab studies have provided initial evidence on the detrimental consequences of short-term exposure to poorly ventilated rooms on cognitive performance (Seppanen et al., 2006; Du et al., 2020). Subjects exposed to poorly ventilated rooms, for at least three hours, displayed lower performance on cognitive functions (including basic activity level, applied activity level, focused activity level, crisis response, information usage, breadth of approach, strategy) (Allen et al., 2016). Experimental evidence shows that children in poorly ventilated classrooms struggle to pay attention and display lower performance in attention, concentration and memory tests (Bakó-Biró et al., 2012; Fisk, 2017). However, there is a lack of long-term field studies that estimate the impact of poor classroom ventilation on educational outcomes.

This paper provides the first evidence on the implications of variation in ventilation on student performance based on a large sample of primary schools, including more than 5,500 primary school students (ages 5 to 13). Primary school students spend most of their school days 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

[1]Similar to the US, nearly 30% of schools in the Netherlands rely on outdated HVAC systems. In the Netherlands, where this study is taking place, the central government recently provided a one-off subsidy of nearly €400 million to primary and secondary schools, exclusively for improving the ventilation infrastructure (Rijksoverheid, 2020).

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

sensors, continuously monitoring the indoor environmental conditions in 216 classrooms across 27 primary schools in the Netherlands over a period of five school terms.[3] Each sensor collects high-frequency measurements on a range of indoor environmental variables: CO_2, coarse and fine particles, temperature, humidity, noise levels, and light intensity. To estimate the impact of a classroom's ventilation quality on the cognitive development of students, we relate daily measures of indoor air quality in classrooms to student scores on nationally standardized tests. During the sample period (2017-2020), each student took an average of seven tests across a range of subjects, including mathematics, spelling, reading, and vocabulary, resulting in more than 37,000 unique test outcomes. All tests in our sample were designed by a national examination center (i.e., not by the teacher) and are part of a national tracking system to monitor the development of students throughout their primary school education.

Our primary measure of ventilation in classrooms is based on the concentration of carbon dioxide (CO_2) in the room. Humans produce and exhale CO_2, which is removed from the room by either mechanical systems (i.e., HVAC) or natural ventilation (i.e., opening windows) that exchange indoor with outdoor air. Engineers and scientists widely use CO_2 as an indicator of how much fresh (outdoor) air is brought into a room, and public officials use it to set guidelines and evaluate the performance of ventilation systems in buildings. As the ventilation rate (i.e., the replacement rate of indoor air with fresh, outdoor air) 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 test domain and student fixed effects. Our estimates identify the impact of indoor air quality (i.e., CO_2) during the prior school term on end-of-term test results, by leveraging within-student variation in air quality over multiple school terms.

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 the CO_2 level during the school term leads to a 0.11-standard-deviation reduction in test scores. The effects are strongest for mathematics, where a one-standard-deviation increase in the CO_2 level during the school term is associated with an 0.21 standard deviation decrease in test scores. Evidence from a heterogeneity analysis suggests poor ventilation impairs learning outcomes of students most strongly for ages 8-12 years old. Our identification strategy relies on the premise that variation in air quality over successive terms for a given student is uncorrelated with unobserved determi-

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

nants of learning.[4] The results from a specification-curve analysis that tests 160 specifications of our main regression analysis show the magnitude and significance of the coefficient associated with CO_2 levels remains largely unchanged when we include several combinations of indoor environmental controls. Similarly, we provide the results of a set of falsification tests in which we match students' learning outcomes with data from sensors deployed in other classrooms. The lack of significant effects suggests our results are not driven by spurious correlation associated with general school characteristics.

A battery of robustness tests further sheds light on the role of three main sources of variation in the classroom over time in CO_2 concentrations in classrooms in our sample: changes in (1) student activity patterns, (2) teacher behavior, and (3) ventilation infrastructure. First, we include controls for changes in activity patterns of students in the classroom. The noise sensor in our study is able to capture minute-by-minute changes in background noise that correlate with a variety of student behavior related to the production of CO_2, such as screaming or moving in the class during teaching hours. The noise sensor also picks up background noise associated with the opening of windows. In an analysis that tests multiple specifications for noise patters in the classroom, we show the magnitude and significance of the coefficient associated with CO_2 levels remains largely unchanged after the inclusion of multiple noise indicators, suggesting the activity patterns in the classrooms are not a main driver of our effects.

Second, we test the correlation between teacher quality, as measured by the test scores of the three cohorts of their students, and the average daily peak CO_2 concentration levels in their classrooms over all five semesters in the sample. The results show no significant correlations, suggesting a lack of correspondence between teacher quality and classroom air quality. Third, we implement an instrumental variable strategy relying on the plausibly exogenous variation in CO_2 concentration associated with failures in school ventilation systems, in the subsample of schools that are mechanically ventilated (85% sample). The failure of an HVAC system on a given day results in sustained and abnormally high levels of CO_2 in the classroom (20%-40% higher than on days when the HVAC system is functioning properly). The results of a two-stage least-squares identification strategy, using ventilation breakdowns as the instrument, show a reduction in test scores of 0.25 standard deviations – an effect that is larger and more robust than results reported in the main specification.

This study is the first to show exposure to poor air quality *inside* the classroom can hinder

[4] We include controls for classroom infrastructure attributes by including classroom fixed effects as well as time-varying factors that could be contemporaneous and correlated with CO_2, such as air particles, temperature, noise, and relative humidity.

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 certain elements of school infrastructure were either in extremely poor condition or non-existent, which suggests new school construction projects are generally 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, using high-frequency measures of indoor environmental quality, rather than relying on monetary indicators of changes in school infrastructure. Our outcome-based approach to school quality can facilitate more precise estimates than a purely input-based approach.[5]

In addition, this paper contributes to the nascent literature exploring the role of environmental factors (i.e., air pollution and extreme heat) in determining cognitive performance and human capital development. Over the last decade, a number of studies have provided 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 respiratory problems in early life (e.g., asthma), 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).[6]

Beyond the health damage, increasing evidence shows 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 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;

[5]See Hanushek (2003) for a discussion of misallocation of resources in school investments driven by input-based approaches.

[6]Additionally, numerous studies show 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 7 million premature deaths due to poor air quality (WHO, 2014).

Graff Zivin et al., 2020).[7] Finally, accumulated exposure to traffic or industry-induced pollution during an academic year has been associated 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 satellites, air quality, or weather stations located miles away from the schools where the pupils are learning and taking their tests.[8] We collect data on air quality and other environmental metrics inside the classrooms in which our subjects are learning and taking exams, overcoming the challenge of measurement error that could result from mis-assigning environmental conditions to individuals (Moretti and Neidell, 2011; Roth, 2018). The data retrieved from our large sensor network allow us to control for a rich set of factors often neglected in the literature (e.g., noise, humidity). Second, we provide evidence on the impact of environmental conditions on children in primary schools, a cohort in which the implications of exposure to poor air quality or extreme temperatures are still largely unexplored. Whereas the current evidence mostly relies on samples of high school or university students, the children in our study are 5 - 13 years old, a critical age range for cognitive and human capital development (Howard-Jones et al., 2012; Heckman, 2006).[9] Our estimates are based on a high-quality panel that contains individual standardized tests in all core learning dimensions that each child takes twice a year, throughout all primary school years.

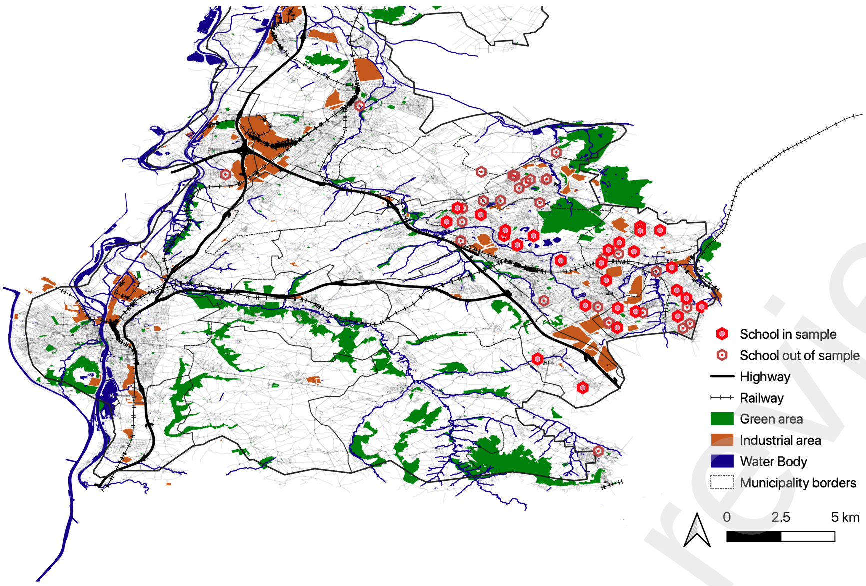
The remainder of the paper is organized as follows. In section 2, we describe the study design and descriptive statistics of the main variables in the study. Section 3 describes the empirical strategy used to link indoor environmental conditions to student academic performance. Section 4 presents the estimation results, section 5 investigates their robustness, and section 6 concludes.

[7]Air pollution also affects labor market outcomes. In particular, the literature provides evidence of air pollution affecting the productivity of agricultural workers (Graff Zivin and Neidell, 2012), the productivity of factory workers (Chang et al., 2016), and the performance of 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).

[8]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.

[9]Persico and Venator (2019) is a notable exception, investigating the impact of proximity to industrial sites or busy highways on the performance of primary school students, studying the learning performance of children from grade 5 onwards (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).

2 Study Design and Data

This study exploits data from a large-scale network of sensors that we deployed in 216 classrooms across 27 schools in the Netherlands. In each of the five school semesters in the sample period (2018-2020), the schools have an aggregate 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 the province of Limburg, the most southern province in the Netherlands.[10] The bright red hexagons in Figure 1 show the location of each school.[11] All schools are located in the same metropolitan area, exposed to similar levels of outdoor temperature and outside 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 to €25.6k, compared to the national median household income of €25.8k.

All schools in the sample follow the same teaching curriculum, and children are evaluated

[10]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.

[11]The dark hexagons display the schools that are part of the school board, but were not selected for the study.

Table 1: Description of Schools and Groups in Our Sample

Panel A: Schools	Mean	St. Dev.	Min.	Max.
Classrooms per school	8	2	5	13
Age of school building (in years)	29	21	2	88
School with mechanical ventilation (in %)	85			
Age of ventilation system (in years)	8	6	0	21
Panel B: Student Groups				
Student Age	9	2	5	13
Group Size	28	10	15	63
Years of Education	3	1	0	8
Panel C: Average test per student				
All tests domains & all groups	9	4.3	1	22
Mathematics	2.4	1.0	0	5
Language skills	6.6	3.4	0	17
Age: 5−7 year old	4.1	2.0	1	10
Age: 8−9 year old	5.0	2.0	1	11
Age: 10−13 year old	8.1	3.9	1	22

Notes: The variable years of education captures the number of years that a student has been enrolled in a primary school. Group size describes the number of students in groups. Age of building describes the number of year since the construction of the school to the beginning of our sample period (January 2018). Row "School with mechanical ventilation (in %)" describes the percentage of schools in our sample with a mechanical ventilation system. Panel C describes the average number of tests that students in our sample take overall, as well as the number of tests in mathematics or language skills (spelling, reading and vocabulary). Finally, it reports average tests taken by different age groups in our sample.

following the same set of nationally standardized exams (Section 2.1.1 describes the national testing system in depth). Panel A of Table 1 describes the characteristics of the 27 schools in the sample. The average school has 8 classrooms, and is 29 years old. The majority of schools in our sample (85 %) have a mechanical ventilation system, that was installed eight years ago, on average.[12] All schools in our sample have heating systems, and none of the schools have an air-conditioning (cooling) system, with relatively mild temperatures in the Netherlands.

Panel B of Table 1 describes the distribution of students' and groups' characteristics in our sample. The average student in our sample is 9 years old. The youngest students are five years old, and the oldest students are thirteen years old. In our study, we focus on students aged almost six and older, the age at which the standardized test system starts for all core subjects in the curriculum.

2.1.1 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 do-

[12] In schools without mechanical ventilation, the ventilation is "natural" (i.e. by opening and closing windows).

mains, including mathematics, reading, spelling, and vocabulary, and apply to students from Kindergarten through 6th grade.[13]

We collect the scores in all tests over the entire primary school education for each student in our sample. The source of this data is the *OnderwijsMonitor Limburg* – a collaboration between 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 (Group 1 and Group 2), given the limited amount of 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 called *CITO*, administered at each individual school, and graded by the teachers using a standardized grading scheme. The raw scores are transformed into percentiles by the examination center *CITO*. The rules for transformation are constant over time, such that results for the same student can be compared between periods.The main outcome variable is the standardized score for each student, in each test period and domain.[14]

Panel C of Table 1 describes the coverage of tests in our sample. On average, each student in our sample takes 9 tests, 2.4 in math and 6.6 in language skills (including spelling, vocabulary and reading proficiency). [15]. The dataset covers students throughout the primary school education. Students in our sample between 5 and 7 years old take an average of 4.1 tests, 8-9 year-old students take 5 tests on average and 10-13 take 8.1 on average.[16]

2.2 School Ventilation and Environmental Conditions in Classrooms

For each school in the sample, the environmental conditions in each classroom with students in group three and above are continuously monitored throughout the sample period (i.e., 2018-2020), 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

[13]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.

[14]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.

[15]The remaining tests are studying skills and listening comprehension

[16]The sample split based on age is based on the tertiles in the distribution of number of students in the sample.

Table 2: Summary Statistics for Environmental Conditions in Classrooms

Panel A: Classroom Ventilation	Mean	St. Dev.	Min.	Max.
Daily Peak of CO_2 (in ppm)	1495	624	737	4665
Daily Average CO_2 (in ppm)	988	336	485	2783
Percentage of days with CO_2 in classroom > 1000 ppm	77	29	0	100
Percentage of days with CO_2 in classroom > 2000 ppm	15	26	0	100
Panel B: Indoor Environmental Quality	**Mean**	**St. Dev.**	**Min.**	**Max.**
Temperature (in °C)	21	1	18	25
Humidity (in rH)	43	6	28	57
Noise (in dBA)	56	2	47	64
PM_{10} (in Count/L)	1104	576	72	3465

Notes: The summary statistics of the environmental parameters presented in Panel B are based on the distribution of daily averages over school periods.

to a cloud-based server whereis aggregated at the minute frequency.[17]

We assess the degree of ventilation in each classroom based on the levels of CO_2 concentration, 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 mechanical or natural ventilation removes it and exchanges it with outdoor air. The recommendation from the main US institution setting standards in building ventilation, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE, Standard 62-2001) highlights that building ventilation rates should keep indoor CO_2 concentrations at a maximum of 1,000-1,200 parts per million in schools (ppm) (i.e., 700 ppm above outdoor concentrations) The levels of outdoor CO_2 are almost always considerably lower than those in occupied rooms – the global average atmospheric carbon dioxide in 2020 was 412.5 ppm (NOAA, 2021).[18] The variation in CO_2 levels inside classrooms that we observe in our sample is almost fully caused by indoor sources, i.e. human breathing, with negligible variation from outdoor levels.

Table 2 provides the summary statistics of the indoor environmental conditions retrieved from the sensors, in days when the students in our sample are inside the classrooms. We restrict the sample period to the official school days and exclude those in which a classroom is emptied, which we infer by observing relatively low levels of CO_2 and noise in the room.

[17] The deployment of sensors took place between January 2018 and December 2018. 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.). Supplementary Figure D.2 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 network is fully deployed in January 2019, and the network has been fully operational since then.

[18] 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).

Panel A in Table 2 describes the distribution of CO_2 levels in our sample. Consistent with Fisk (2017), who provides a recent review of the literature of CO_2 in schools, all students in our sample are exposed to CO_2 levels above the recommended threshold (700 parts per million (ppm) above outdoor concentrations (i.e., 1,100-1,200ppm) recommended by ASHRAE). The distribution of daily peaks of CO_2 indicates that the average student in our sample is exposed to 1,495 ppm on a given school day, with a range spanning from 737 ppm to 4,665 ppm (i.e. almost four times the limit recommended by ASHRAE). The levels of CO_2 exceed 1,000 (2,000) ppm 77% (15%) of teaching days.

Panel B in Table 2 provides the summary statistics of the 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 level is 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).[19] The relative humidity levels are between the recommended thresholds by the Environmental Protection Agency (i.e., 30-50%). [20]. Finally,the levels of noise in the classrooms are below the levels affecting human health (Hammer et al., 2014). [21]

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 on average, relative to their own score following a school term in a classroom with "good" air quality (or vice versa).

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

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

[20]For more details about humidity control in schools, see `https://www.epa.gov/iaq-schools/moisture-control-part-indoor-air-quality-design-tools-schools`

[21]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).

general trends in test scores in our sample for each testing domain. More formally, we estimate the following specification:

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

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 or June 2019). We define 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.[22] We standardize 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 (e.g. changes in levels of outdoor pollution or ambient temperatures). The classroom fixed effects (α_c) control for all time-invariant characteristics of a classroom, such as views or angle to the sun, and the teaching infrastructure in the room (e.g. digital boards, furniture, etc.).[23]

Finally, X_{ct} is a vector of indoor environmental quality, 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 series of dummmies describing the years of schooling by the time they take the test in their school time.

Standard errors (ε_{ict}) are clustered at the classroom-by-period level to control for correlation among test results for students learning in the same classroom at the same time, and following

[22] Section 5.1 includes an specification curve, where we test a wide variety of specifications of our treatment variable to test the sensitivity of our results to the specification of the exposure to CO_2.

[23] 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. 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 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. 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.

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 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, conditional on the set of fixed effects, is independent of other variables that might be correlated with CO_2 levels. 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 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.[24] 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, a challenge to our identification is that teachers might sort into rooms based on indoor environmental quality, with high-quality teachers selecting classrooms with better ventilation. In our setting, it is the school principals who determine classroom allocation to teachers and student groups, rather than individual teachers. Furthermore, the classroom fixed effects of the fixed-effect model partly control for possible 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. In a robustness check, we collect data on the allocation of teachers to classrooms in each term and for each school in our sample to test how the average value added of each teacher in children scores correlates the average levels of CO_2 over the entire sample period of the study. The lack of correlation between average CO_2 levels and teacher quality supports the robustness of our results.

Finally, 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. The vast majority of schools in the sample (85%) are mechanically ventilated. The failure of a mechanical ventilation system on a given day results in sustained and abnormally high levels

[24] 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.

of CO_2 in the classroom. We construct 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. Using this variation, we build a two stage least square strartegy to instrument the average exposure over the term with the number of days that the ventilation system in the classroom is not working properly.

4 Results

4.1 Indoor Air Quality and Student Performance

This section presents our main estimation results linking the variation of classroom CO_2 concentrations in the school term to student performance in subsequent standardized tests, followed by a heterogeneity analysis, and a series of robustness checks.

4.1.1 Main Results

In Table 3, we report our baseline results of the relationship between classroom ventilation as described by the CO_2 concentrations and test scores. The table displays the (standardized) estimated coefficients in Eq. 1, introducing sequentially the set of time-varying controls and fixed-effects to test the sensitivity of the parameters to multiple specifications. The results indicate that high concentrations of CO_2 in the classroom during the school term lead to significantly lower performance of students in the test.

Column (1) in Table 3 describes the estimates in a regression only including student, classroom and period fixed effects. The results from this simple regression show a negative relationship between average concentrations of CO_2 in the term and the test scores. The estimated damage of CO_2 on test scores after controlling for basic environmental conditions in the classroom (Column (2)) shows that a one standard deviation higher exposure to average daily peak CO_2 levels during the school-term lowers student performance by 0.094 standard deviations. 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, 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, our proxy for particulate matter in our sample. The changes over time in those conditions are controlled for

Table 3: Average Daily Peak CO2 Concentration During Learning Period on Standardized Test Scores

	Dependent Variable: Standardized Test Scores				
	(1)	(2)	(3)	(4)	(5)
CO_2	-0.066^{*}	-0.094^{**}	-0.115^{***}	-0.108^{**}	-0.110^{***}
	(0.037)	(0.041)	(0.043)	(0.045)	(0.038)
PM_{10}		0.016	0.010	0.044	0.047
		(0.062)	(0.058)	(0.057)	(0.071)
Temperature		−0.014	−0.011	−0.006	−0.016
		(0.029)	(0.029)	(0.025)	(0.031)
Humidity		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				−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
Years of schooling	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

Notes: All models relate average daily peak of CO_2 concentration in the classroom over the school term with subsequently obtained standardized test scores. We standardize all coefficients in the table to facilitate the comparisons. Column (1) provides results for a model with fixed effects for classroom, students by each subject domain (i.e., math, spelling, vocabulary, etc.) and period. 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) displays results for a model that adds the students' age which is included based on the tirtile of the distribution, setting the youngest group in our sample Age 5 to 7 as reference group. Column (4) includes controls for class size (i.e., number of students enrolled) and average noise levels (used as a proxy for student behavioral patterns in the classroom). Finally, the last column (Column (5)) adds school by period fixed effects, relying only in the variation within each specific school in a given school term (i.e., variation in conditions across classrooms) and rule out unobserved heterogeneity at the school level. Clustered standard errors at the classroom by period level are shown in brackets and significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

by the period fixed effects and school-period fixed effects.

The decline in test scores following terms with poor air quality relative to their own scores following terms with good air quality in the classroom are not driven by other channels potentially correlated with CO_2 in the school term leading up to the test, as seen in the next three last columns of Table 3 (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)

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

	by Domain			by Age			Advice to University prep.(1=Yes)
	Spelling (1)	Math (2)	Reading (3)	[5-7] (4)	[8-9] (5)	[10-13] (6)	(7)
CO_2	−0.022	−0.215***	−0.116**	0.109	−0.173***	−0.185***	−0.127***
	(0.061)	(0.057)	(0.059)	(0.072)	(0.052)	(0.064)	(0.003)
Obs.	10,627	10,437	8,434	9,553	13,164	14,734	193
Adj. R^2	0.757	0.784	0.750	0.720	0.775	0.762	0.031

Notes: All models relate average daily peak of CO_2 concentration in the classroom with standardized test scores in spelling (Column (1)), Maths (Column (2)) and Reading (Column (3)). In addition, Columns (4)-(6) provide the results decomposed by different student age. All specifications in Columns (1)-(6) include the complete set of environmental controls (particulate matter, temperature, humidity and noise), class size, and classroom, student and period fixed effects. In addition to this list, Columns (4)-(6) includes student by test domain fixed-effects. Column (7) relates average daily peak of CO_2 concentration in the classroom with the probability that the teachers advise that the student attend a university preparing high-school. Specification in Column (7) includes environmental controls and class size as controls. Clustered standard errors at the classroom by period level are shown in brackets. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

has nearly no effect on the point estimate (Column (3) in Table 3). Controlling for changes in class characteristics leaves our estimates nearly unchanged. This suggests that our estimates are measuring direct impacts of poor ventilation in the classroom, and not changes in class size or changes in activity patterns in the classroom (Column (4) in Table 3).

Finally, Column (5) presents the results of a regression including school-period fixed effects, controlling flexibly for any school-level changes across periods in our sample, such as outdoor environmental conditions (e.g. air pollution or outdoor temperature), investments in equipment, or changes in the leadership of the school. Again, the results remain largely unaltered, with a 0.11 standard deviation lower average test result for a one standard deviation increase in CO_2 exposure. Overall, the results reported in the table support our hypothesis that our findings reflect the direct effects of ventilation quality in the classroom during the school term, rather than confounded characteristics of schools or classes.

4.2 Heterogeneity

In this section, we examine heterogeneity in the treatment effects reported in Table 3. In particular, we test whether there are differences in treatment effects across tests in our sample.

Testing domain. Table 4 presents the estimates of our main regression for the different sub samples, including the full set of fixed effects and controls. Column (1) to (3) in Table 4 show the estimated impact of CO_2 separately for the three major subject domains in our sample of tests - i.e, spelling, mathematics, and reading. The Table shows that the results in the pooled estimation, including all tests in the sample, are mainly driven by mathematics tests and reading comprehension tests, while that spelling does not seem to be significantly affected by bad ventilation quality in the classroom.

Student age and secondary-school advice. Columns (4) to (6) in Table 4 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 groups. This is important, since the Netherlands has high school education at different levels, and the choice for a specific level is made when pupils are in the final primary school year.

The final year of primary school education includes a set of exams that shape the long term education of students. In their final year, students take a final set of standardized tests conducted with the sole purpose of informing the teacher's advice about what type of secondary school students should attend (university vs. professional education oriented secondary schools).[25]

In a cross-sectional analysis, we connect the scores in the test informing the teacher's advice on the type of secondary school that the student should attend with the CO_2 levels in the term preceding the test. We relate the school term CO_2 exposure in the classroom preceding the tests informing secondary-school advice with its result, and with the probability that the test results point to advising the student to go to a university-oriented high school. Column (5) in Table 4 shows results for this analysis, including controls for the other environmental parameters, age, and the class size. This result suggests that a one standard deviation increase in CO_2 concentration is associated with a 13% lower chance to get a test result recommending a student to go to the two most advanced high school levels.

Treatment decomposition by day of the week. Finally, we explore differences in impacts associated with high concentrations of CO_2 across school week days. We construct CO_2 concentration measures specific for each day of the week, and re-estimate our main specification 1. Supplementary Table A.1 displays the coefficients associated with the average concentration

[25] The secondary school system in the Netherlands includes three different types of high-school that prepare students for three different types of career paths after completing high-school (academic university, polytechnic schools or straight to the labor market after high-school). Students in the Netherlands can attend different types of high schools, according to their performance in primary school. The first type of high school, "VMBO", requires the lowest performance in primary school education, and it is orientated towards acquiring practical sets of skills. This all but precludes a subsequent college education. The second level, "HAVO", requires an intermediate academic level and focuses on a mix of applied and conceptual skills. This type of high school prepares students for higher education at the polytechnic or community college level, after which it is still possible to enter university. Finally, for students with the highest level of academic achievement, "VWO" schools teach more conceptual subjects with most of these students entering research universities after high school completion (Supplementary Figure D.3 displays the distribution of test scores in the final test separately for each recommended high school track). Teachers calculate a comprehensive score combining all these test results and compare it with their own advice. If the advice shown by the tests points at a more advanced high school level than advised by the teacher, teachers are obliged to review their decision, if in turn the opposite occurs, there are no consequences and the original advice remains.

of CO_2 in different school days. The results indicate that poor environmental conditions within the classroom are equally important across school days, except for Wednesdays, when we do not find a significant effect of CO_2 exposure. In the Netherlands, the current educational calendar indicates that children should only be at school for half a day on Wednesdays, and therefore being exposed to conditions in the classroom for half of the time than a regular day. This result provides supportive evidence on the role of exposure to CO_2 driving our estimates of the drop in test scores associated with poor air quality, since our estimates are weakest for the days that students are exposed to CO_2 in the classroom for the shortest period of time (half of time in normal days).

5 Robustness

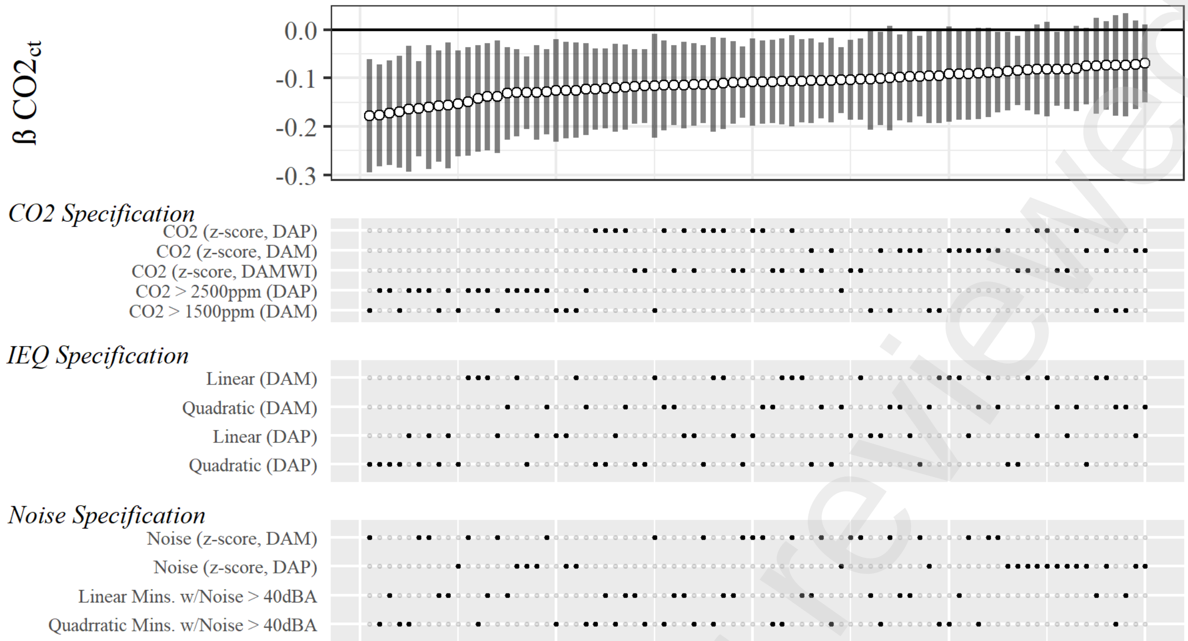
5.1 Specification Tests

In our main results, we include all indoor climate regressors as the average of the daily peak exposure 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 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 dozens of combinations of different forms of the CO_2 variable 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)) [26], a dummy variable indicating that the school term average of daily peaks in CO_2 concentration was above 2500 ppm ($CO_2 > 2,500$ ppm (DAP)), and a dummy variable indicating that the school term average of daily averages of CO_2 was above 1500 ppm ($CO_2 > 1,500$ ppm (DAM)). Similarly,

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

Figure 2: Specification Sensitivity Curve



Notes: Figure plots coefficients from separate (standardized) estimates of the main effect of CO_2 on test scores (i.e, coefficient β in Equation 1). Dots describe point estimates. Vertical lines indicate 95% confidence intervals. Alternative models include different specifications for CO_2 exposure levels during the school term. This includes the average daily mean of CO_2 ($CO2(z-score, DAM)$),average daily mean of CO_2 including only the minutes that the classroom is occupied ($CO2(z-score, DAMWI)$) and two dummy variables indicating number of days in the term that the daily peak of CO_2 is above 2,500ppm ($CO2 > 2,500ppm, DAP$)) or the daily mean of CO_2 is above 1,500ppm ($CO2 > 1,500ppm, DAM$)). Next set varies redefines the specification of envrionmental controls. Including a linear ($Linear(DAM)$) and quadratic ($Quadratic(DAM)$) form of daily averages of each environmental factor, and a linear ($Linear(DAP)$) and quadratic ($Quadratic(DAP)$) form of daily peaks of each environmental factor. Finally, the last includes multiple transformations of noise readings in the classroom. Background noise is included as average daily average over the term ($Noise(z-score, DAM)$), average daily peak over the term ($Noise(z-score, DAP)$), number of minutes in a day with noise levels above 40dBA ($LinearMins.w/Noise > 40dBA$) and its squared term ($QuadraticMins.w/Noise > 40dBA$).

we test different specifications of the environmental controls based on the school term daily averages (Linear (DAM)) and the school term daily peaks (Linear (DAP)), 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 and combinations of the treatment and environmental control variables. The coefficient shows high stability in terms of its magnitude, sign, and statistical significance. The estimates in our main specification are within the confidence intervals of the CO_2 coefficient in each specification of the curve.[27] The occupancy patterns in the classroom are the key source of variation in the levels of CO_2 and 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

[27]The few specifications that are not statistically significant are associated with the school term average of daily mean, without the adjustment for the minutes that children are in the classroom (CO_2 (z–score) DAM).

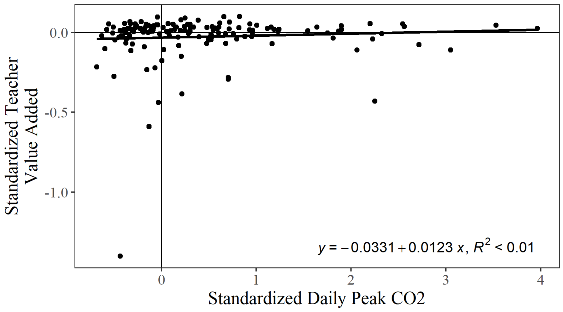
specifications supports the robustness of the findings, and indicates that our results are not driven by a specific definition of exposure measurement.

Noise measures of activity patterns in the classroom. Alterations in children's activity patterns might generate changes in CO_2 levels. As 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,[28] 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 a lack of influence of classroom activity patterns on our estimates, and therefore indicates that alterations of activity patters in the classroom do not play a meaningful role in the relationship between ventilation quality and learning outcomes we document in our study.

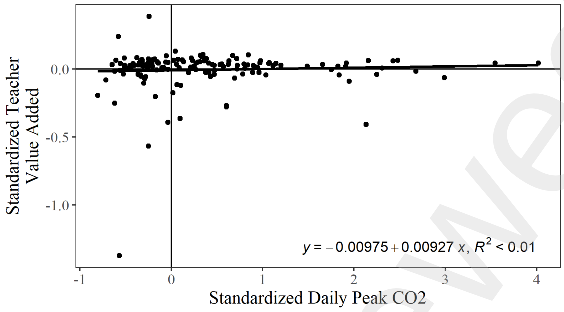
Falsification test: miss-assignment of sensor to student. Finally, we perform a set of falsification tests where we misspecify the exposure to CO_2 during the school term for each student. We randomly choose a sensor from another classroom than the one to which the student was allocated and relate the CO_2 concentration in the *wrongly* assigned classroom with the student's subsequent test scores. We reproduce this test 1000 times, each time randomly drawing a sensor different from the one allocated to the classroom where the student received lessons during the school term, and create a distribution for the coefficients that we obtain from fitting Equation (1) using the miss-assigned CO_2 concentration. In Supplementary Figure B.1, we show the distribution of these estimated coefficients associated with CO_2 concentrations from classrooms within the same school but different to the one where the student learned (left plot) and from classrooms located in any other school in our sample. These estimates are mostly insignificant and distribute fairly normal around zero, with the estimate from our main specification falling within values located in the left tail in each case.

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

Figure 3: Correlation Between Teacher Quality and Classroom CO_2 levels



(a) Only Teacher FEs



(b) Teacher, Domain, Period, and Years of schooling FEs

Notes: The figure describes the correlation between the estimated teacher fixed effects in Equation 2 and the estimated teacher fixed effects in Equation 3. Panel (a) includes no controls in the regression models. Panel (b) controls for test domain, period and student years of schooling at the time of the test. Thus defined, the Pearson correlation coefficients between teacher quality and ventilation quality are 0.18 and -0.12, respectively, both statistically insignificant. The regression coefficients associated with the average CO2 peak levels in the classroom (denoted by x in the regression equations displayed in the figures) show a small and not statistically significant relationship between CO2 levels and value added measures of teaching quality.

5.2 Teacher Quality and Classroom Ventilation

A key challenge in the identification of classroom infrastructure at schools has to do with its correlation with teacher quality. It is possible that high-quality teachers understand the importance of good (ventilation) infrastructure better, or that good teachers are better at choosing classrooms with such infrastructure. This correlation between teacher quality and ventilation quality would therefore challenge the interpretation of our results as air quality impacts.

In this subsection, we estimate the correlation between teacher quality and the levels of CO_2 in the classroom where they teach. We estimate teacher quality using teacher value-added measures (Chetty et al., 2014; Rivkin et al., 2005), based on differences in test-scores by students at the end and the beginning of the school-term taught by the teacher. We link the value-added in test-scores of teachers with the average levels of CO_2 in the classrooms where the teacher taught. [29]

Figure 3 displays the link between the two sets of fixed effects, to test for a possible link

[29] We calculate the teacher-specific differences by extracting teacher fixed-effects in the following regression models:

$$CO2_{ctp} = Teacher_p^{CO_2} + \varepsilon_{itp} \tag{2}$$

and

$$\Delta_{t-1}^{t} Score_{cdp} = Teacher_p^{tests} + \varepsilon_{itdp} \tag{3}$$

where $Teacher_p^{CO_2}$ in Equation 2 describes the teacher fixed effects, containing dummy variables capturing differences across teachers in the CO_2 levels observed at the classrooms (c) allocated in the school term (t) to the specific teacher (p). $\Delta_{t-1}^{t} Score_{cdp}$ describes the difference between a student's test scores at the end of school-term t and $t1$ (just before and after being instructed by teacher p). Similarly, in Equation 3, $Teacher_p^{tests}$ describes the teacher fixed effects, containing dummy variables capturing differences across teachers in the test scores of their students.

between teacher quality and classroom CO_2. The lack of a statistically significant relationship between the two sets of fixed effects indicates that teachers are not material for the impacts of CO_2 on the performance of their students.

5.3 Instrumenting CO_2 Concentrations in the Classroom with Failures in Mechanical Ventilation Systems

In this subsection, we explicitly test for the role of the school infrastructure (i.e. HVAC system) as a driver of our main effects. We develop a data-driven algorithm to detect infrastructure failures based on the presence of jumps (i.e. sudden, transitory and abnormally high levels of CO_2) in the time series of daily peaks of CO_2 in classrooms. Then, we implement a two-stage least square strategy (2SLS) where the school term average concentration of CO_2 is instrumented with the number of days that the algorithm detects that the ventilation system supplying fresh air to the classroom is broken. For this analysis, we restrict the sample to 85% of schools that are mechanically ventilated. These schools have an HVAC system, refreshing the air in the room with outdoor air.

The main consequence of a failing HVAC system is the reduction of air exchange rates in a classroom, letting the level of CO_2 accumulate to (abnormally) high levels. High concentrations of CO_2 are among the main indicators used by technicians to detect failures in ventilation systems. The failure of engines or blockage of pipes in the system are common failures that generate abnormally high concentrations of CO_2 in a classroom, which 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 and have no power over.

We use data from our sensors to detect spikes in the daily concentration of CO_2 in the classroom. We design an 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 peak 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 (For a detailed description of the algorithm, see Appendix C). We then create dummy variables $Days\_Broken\_Vent_{ctv}$ corresponding to bins $v = (0, 1, 2, 3, > 3)$ indicating the number of days with a detected broken ventilation over the school term, to build an instrument of the treatment variable in our main specification (eq. 1).[30]

[30] Appendix Table C.2 displays the results using a linear specification of the instrument (number of days with

Table 5: CO2 Concentration on Standardized Test Scores: Instrumental Variables Using Number of Days With Broken Ventilation

	IV				OLS	
	Peak CO_2		Average CO_2		Peak CO_2	Average CO_2
	(1)	(2)	(3)	(4)	(5)	(6)
	2^{nd} Stage	1^{st} Stage	2^{nd} Stage	1^{st} Stage		
CO_2	-0.250^{***}		-0.338^{***}		-0.111^{**}	-0.089^{*}
	(0.047)		(0.103)		(0.048)	(0.049)
1 day broken		0.650^{**}		0.604^{**}		
		(0.253)		(0.253)		
2 days broken		0.440^{***}		0.267^{**}		
		(0.086)		(0.116)		
3 days broken		2.119^{***}		1.129^{***}		
		(0.194)		(0.195)		
> 3 days broken		1.122^{***}		1.405^{***}		
		(0.310)		(0.298)		
Kleibergen-Paap						
F-statistic		43.118		24.843		
p-value		0.0000		0.0000		
Obs.	32,442	32,442	32,442	32,442	32,442	32,442
Adj. R^2	0.745	0.965	0.743	0.957	0.746	0.746
Fixed Effects						
Student by Domain	Y	Y	Y	Y	Y	Y
Period	Y	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y	Y
Years of Schooling	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

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. Column (1) shows results from instrumenting average daily peaks in CO_2 concentration during the school term and Column (2) shows results from the first stage of the instrumentation. Columns (3) and (4) show the same set of results as columns (1) and (2), but from instrumenting average daily mean levels of CO_2 concentrations. Columns (5) and (6) show the corresponding OLS estimates from fitting Equation (1) using the subsample of schools with mechanical ventilation. Clustered standard errors at the classroom by period level are shown in brackets and significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 5 shows the coefficients from the two-stage instrumental variable regression, instrumenting school-term average CO_2 concentrations with the number of days that the mechanical ventilation system was failing. Column (2) includes the first stage coefficients associated with the set of dummy variables describing the number of days that the ventilation system was identified

the ventilation system broken). The results are consistent with the dummy variable specification, showing the robustness of our results to different specifications.

as broken. The high values of the F-statistic indicates strong predictive power of the ventilation breakdown events on the school-term average CO_2 average daily peak in the classroom over the school term, after controlling for classroom fixed effects and the remaining factors listed in our main regression (Eq. 1). Column (1) shows 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. Column (3) and (4) shows the results using daily averages of CO_2 as treatment. Similarly, Supplementary Table C.3 replicates the heterogeneity analysis using the 2SLS strategy, showing that the coefficients are consistent to our main coefficients in sign and statistical significance, and larger in magnitude than the original regression results.

6 Discussion and Conclusion

This paper provides the first 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 primary schools, continuously monitoring the indoor air quality conditions in 216 primary-school classrooms, including 5,500 children at ages 5 - 13 years old. Using within-child variation in exposure to air quality conditions, we document that systematic exposure to poorly ventilated classrooms during the school term (as measured by high levels of CO_2) impairs student performance on nationally standardized tests. An increase in classroom CO_2 level during the school term by one standard deviation reduces subsequent test scores by 0.11 standard deviations. Our findings imply that exposure to poor indoor air quality directly interferes with learning, highlighting the role of physical school infrastructure in determining educational outcomes.

To understand the magnitude of the results, we benchmark our estimates with the literature that evaluates the impact of a variety of other factors on test scores. First, we compare our results to estimates from studies evaluating the impact of environmental stressors on test scores. Recent studies show that child exposure to outdoor air pollution leads to a 0.04 to 0.05 standard deviation lower test score (Persico and Venator, 2021; Heissel et al., 2022), which is about half the baseline effect associated with elevated CO_2 levels, as documented in this study. Similarly, estimating the effect of heat exposure on standardized test scores, Park et al. (2020) document that a one standard deviation increase in average temperature over the school year (SD =

$65°F$) leads to a 0.15 standard deviation decrease in test scores, about 1.4 times the estimated performance decrease associated with a standard deviation increase in CO_2.

Another relevant comparison is to recent findings by Engzell et al. (2021), who study the effects of primary school closings, due to COVID-19, on student performance, using the same national tests exploited in this study. The estimates show that eight weeks of interruption in in-person learning led to a 0.08 standard deviation decrease in standardized test scores,[31] and a 0.11 standard deviation decrease in test scores for the sample of students with a lower socio-economic background (comparable to students in our sample). The magnitude of the COVID-19-induced performance decrease is almost identical to the results documented in this study.

Finally, we can compare the cost-effectiveness of investments required to improve indoor air quality (through mechanical ventilation system upgrades) with two alternative school investments evaluated in the literature: class size reductions and the installation of air-conditioning in schools. A program evaluation of the Tennessee STAR experiment shows that the average cost associated with reductions in class size is \$163 per child, for each percent of a standard deviation increase in test scores (**?**). Similarly, the cost of increasing test scores by one percent of a standard deviation through the installation of air-conditioning is ranging from \$25 and \$125 per student per year (Park et al., 2020). A recent report on ventilation systems in Dutch primary schools estimates the cost to upgrade a non-mechanically ventilated school to mechanical ventilation is \$592,190 (RuimteOK, 2021), or \$463 per student for an average primary school of 230 children[32]. This upgrade aims to reduce the measured concentration of CO_2 below 900ppm, that is the equivalent in our sample of reducing the average levels of CO_2 by one standard deviation. Our baseline estimates show that a decrease in CO_2 by one standard deviation translates into an 0.11 standard deviation increase in test scores. The amortized cost of improving test scores by 1 percent of a standard deviation via mechanical ventilation upgrades is USD42 per student per year, which is much lower than reducing class size, and at the lower bound of the cost of upgrading air conditioning in classrooms, as estimated by Park et al. (2020). [33]

Our results also yield several lessons for ongoing response policies against the spread of COVID-19. The airborne transmission of the SARS-COV-2 virus has elevated the salience of room ventilation as an important factor to prevent the spread of the disease. Schools buildings are among the major targets in many countries, due to the high density of children in classrooms,

[31] In the Netherlands, all classes were immediately migrated online, mitigating any interruptions in instruction time.

[32] €500,997, Conversion based the average exchange rate in 2021.

[33] Amortized cost calculations are based on a 5 percent discount rate, an average economic life of the system of 20 years, and fixed costs of \$592,190 per school per year.

and the general lack or state of disrepair of ventilation and air treatment systems in schools. Many countries are preparing investment programs to improve ventilation, through installation or modernization of HVAC systems, or upgrading the standards of ventilation in buildings. However, the ambition of these investment programs is quite limited. For example, the Dutch government has allocated €6 billion to remediate the results of the COVID-19 pandemic on learning outcomes, as compared to allocating €400 million for repairs and maintenance of school ventilation systems. Our results suggest that improving in indoor air quality in schools has relevance beyond the reduction of the spread of viral diseases, such as COVID-19, and can support children's educational outcomes.

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.

Aihounton, G. B. and A. Henningsen (2021): "Units of measurement and the inverse hyperbolic sine transformation," The Econometrics Journal, 24, 334–351.

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.

ASHRAE (2022): "ASHRAE Position Document on Indoor Carbon Dioxide," https://www.ashrae.org/file%20library/about/position%20documents/pd_indoorcarbondioxide_2022.pdf, [Accessed September 2022].

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.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," American Economic Review, 104, 2633–79.

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

Cornman, S. Q., J. J. Phillips, M. R. Howell, and L. Zhou (2022): "Revenues and Expenditures for Public Elementary and Secondary Education: FY 20 (NCES 2022-301)," U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved September 2022 from https://nces.ed.gov/pubsearch.

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.

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.

Engzell, P., A. Frey, and M. D. Verhagen (2021): "Learning loss due to school closures during the COVID-19 pandemic," Proceedings of the National Academy of Sciences, 118, e2022376118.

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," .

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.

Heissel, J. A., C. Persico, and D. Simon (2022): "Does pollution drive achievement? The effect of traffic pollution on academic performance," Journal of Human Resources, 57, 747–776.

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.

——— (2021): "The effects of local industrial pollution on students and schools," Journal of Human Resources, 56, 406–445.

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

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005): "Teachers, schools, and academic achievement," Econometrica, 73, 417–458.

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," .

RuimteOK (2021): "Verdiepend onderzoek gebouwtypen en ventilatiesystemen (In-depth research of building types and ventilation systems)," .

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 Estimates by different days of the week

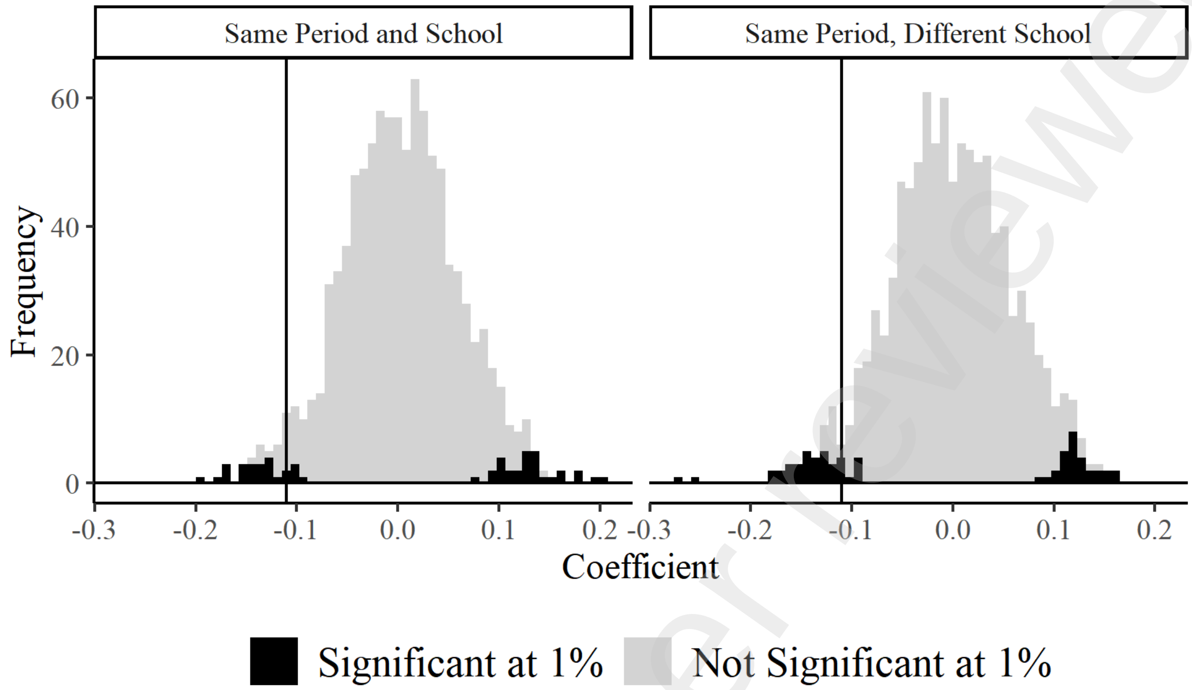
Table A.1: Decomposition of CO_2 impacts by day of the week

	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

Note: 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. School days in Dutch primary schools start at 8am and finish at 4:00pm, except for Wednesdays when school days end at 12:30pm. The last column includes daily average peak CO_2 observed in all days of the week. Clustered standard errors at the classroom by period level are shown in brackets and significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

B Falsification Test

Figure B.1: Falsification Test



Note: Left panel shows the frequency of coefficient point estimates obtained from fitting Equation (1) 1000 times, each time randomly assigning to each classroom CO_2 concentrations observed in another classroom of the same school and in the same period. Right panel shows the frequency of coefficient point estimates obtained from fitting Equation (1) 1000 times, each time randomly assigning to each classroom CO_2 concentrations observed in another classroom in another school and in the same period. The vertical lines indicate the coefficient we estimated as reported in Table 3.

C Broken ventilation: robustness

C.1 Description of algorithm

We use data from our sensors to detect spikes in the daily concentration of CO_2 in the classroom. We design an 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:

$$CO2^{s}_{ch\tau} = \alpha_{cht} + \alpha_{cwt} + \alpha_{\tau} + \varepsilon_{ch\tau} \tag{4}$$

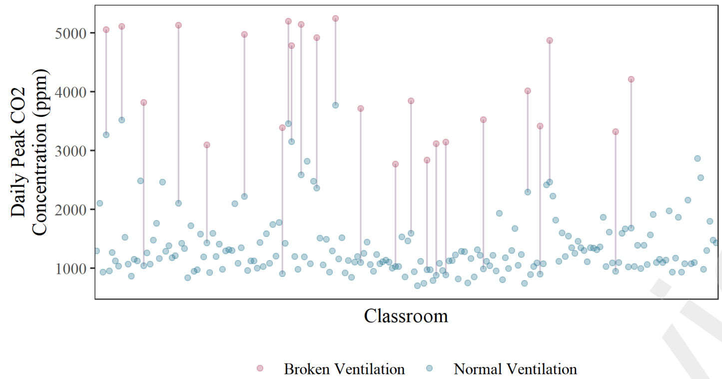
, where $CO2_{ch\tau}$ is the maximum level of CO_2 observed in classroom c, during the lesson that started at time h on date τ, α_{cht} is a classroom c by hour of the school day h during learning period t, α_{cwt} is a classroom by day of the week w during learning period t, and α_{τ} is a date fixed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

To determine whether a specific classroom encounters 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 on date τ if $\hat{\varepsilon}_{ch\tau} > 1,500$ ppm for both, average and maximum levels. The key rationale 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 during the day, rather than just being a spike in CO_2 of a few minutes.

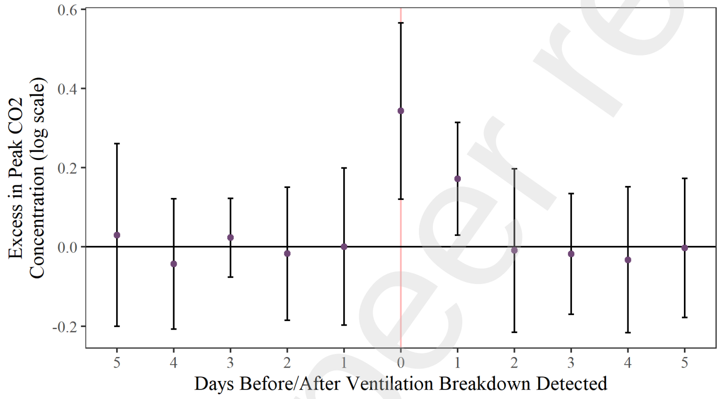
Panel a in Figure C.1 shows average peak CO_2 concentration levels in each classroom during the school term in days with (in red) and without (in blue) broken ventilation. The figure shows how the days detected as having a failing HVAC system show sudden increases in CO_2. Panel (b) in Figure C.1 displays 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. 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 CO_2 levels in the classroom, and remain high (about 20% higher) during the day immediately after the breakdown was detected, suggesting

Figure C.1: 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.

that HVAC systems cannot always be fixed on the same day.

C.2 Two-stage least squares specification

We then create dummy variables $Days\_Broken\_Vent_{ctv}$ corresponding to bins $v = (0, 1, 2, 3, > 3)$ indicating the number of days with a detected broken ventilation, in a regression including the same fixed effects as in our main regression specification (Equation 1):

$$\widehat{CO2_{ct}} = \sum \xi_v Days\_Broken\_Vent_{ctv} + \Gamma X_{ct} + \alpha_{id} + \alpha_c + \alpha_t + \alpha_l + \varepsilon_{idt}, \tag{5}$$

Second stage:

$$Score_{idt} = \beta \widehat{CO2_{ct}} + \Gamma X_{ct} + \alpha_{id} + \alpha_c + \alpha_t + \alpha_l + \varepsilon_{idt}. \tag{6}$$

Table C.1: CO2 Concentration on Standardized Test Scores: Instrumental Variables Using Number of Days With Broken Ventilation

	Dep. Variable: Standardized Test Scores				
	Peak CO2		Average CO2		Reduced Form
	(1)	(2)	(3)	(4)	(5)
	2^{nd} Stage	1^{st} Stage	2^{nd} Stage	1^{st} Stage	
CO_2	-0.250^{***}		-0.338^{***}		
	(0.047)		(0.103)		
1 day broken		0.650^{**}		0.604^{**}	-0.165^{*}
		(0.253)		(0.253)	(0.088)
2 days broken		0.440^{***}		0.267^{**}	-0.053
		(0.086)		(0.116)	(0.087)
3 days broken		2.119^{***}		1.129^{***}	-0.527^{***}
		(0.194)		(0.195)	(0.080)
$>$ 3 days broken		1.122^{***}		1.405^{***}	-0.337^{**}
		(0.310)		(0.298)	(0.137)
Kleibergen-Paap F-statistic		43.118		24.843	
p-value		0.0000		0.0000	
Obs.	32,442	32,442	32,442	32,442	32,442
Adj. R^2	0.745	0.965	0.743	0.957	0.746
Fixed Effects					
Student by Domain	Y	Y	Y	Y	Y
Period	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y
Years of Schooling	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

Note: This table presents results for our main specification when instrumenting CO_2 concentration levels during the school term using dummies indicating the number of days that the ventilation in the school was detected as broken by our algorithm. The table reproduces the coefficients reported in Table 5 (columns (1) and (3)) and their corresponding 1^{st} stage coefficients for the instruments, as well as the Kleibergen-Paap F-statistic and corresponding p-values (columns (2) and (4)). Column (5) presents the reduced form estimates for the non-linear effect for days with broken ventilation on test scores. Clustered standard errors at the classroom by period level are shown in brackets and significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

C.3 Using a log-linear specification for the instrument

Table C.2: CO2 Concentration on Standardized Test Scores: Instrumental Variables Using Number of Days With Broken Ventilation

	Dep. Variable: Standardized Test Scores				
	Peak CO_2		Average CO_2		Reduced Form
	2^{nd} Stage	1^{st} Stage	2^{nd} Stage	1^{st} Stage	
CO_2 (IV)	-0.253^{***}		-0.253^{***}		
	(0.063)		(0.071)		
Number of days broken		0.672^{***}		0.174^{***}	-0.170^{***}
		(0.184)		(0.062)	(0.058)
Obs.	32,442	32,442	32,442	32,442	32,442
Adj. R^2	0.745	0.959	0.744	0.925	0.746
Kleibergen-Paap					
F-statistic		13.38		28.63	
p-val		0.000		0.000	
Fixed Effects					
Student by Domain	Y	Y	Y	Y	Y
Period	Y	Y	Y	Y	Y
Classroom	Y	Y	Y	Y	Y
Years of Schooling	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

Note: This table presents results for our main specification when instrumenting CO_2 concentration levels during the school term using the log number of days that the ventilation in the school was detected as broken by our algorithm. We apply the inverse hyperbolic sine transformation to the number of days with a broken ventilation and take logs on the transformed variable. We fitted each model several times using different scaling parameters as suggested in Aihounton and Henningsen (2021) but observe very minor differences in the estimated coefficients, R^2 values, and log likelihoods achieved, so we present the results without scaling. The comparison set of results are available upon request. The table reproduces the second (columns (1) and (3)) and first stage coefficients, as well as the Kleibergen-Paap F-statistics and corresponding p-values (columns (2) and (4)). Column (5) presents the reduced form estimates for linear effect for days with broken ventilation on test scores. Clustered standard errors at the classroom by period level are shown in brackets and significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

C.4 Heterogeneity in the IV results

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

	by Domain			by Age		
Second Stage	Spelling	Maths	Reading	[5-7]	[8-9]	[10-13]
CO2 (IV)	-0.159^{*}	-0.361^{***}	-0.308^{***}	0.067	-0.233^{***}	-0.216^{*}
	(0.082)	(0.055)	(0.077)	(0.274)	(0.052)	(0.115)
Obs.	9,282	9,028	7,361	8,249	11,674	12,519
Adj. R^2	0.602	0.651	0.632	0.569	0.777	0.707
First Stage						
1 day broken	0.850^{***}	0.592^{***}	0.343^{***}	0.820^{**}	0.120	1.036^{***}
	(0.208)	(0.190)	(0.104)	(0.406)	(0.116)	(0.368)
2 days broken	0.422^{***}	0.354^{***}	0.529^{***}	0.706^{***}	0.511^{***}	0.460^{***}
	(0.157)	(0.085)	(0.132)	(0.254)	(0.088)	(0.084)
> 2 days broken	1.739^{***}	1.868^{***}	1.785^{***}	1.455^{***}	2.040^{***}	1.452^{***}
	(0.215)	(0.260)	(0.137)	(0.190)	(0.103)	(0.477)
Kleibergen-Paap F-stat	47.88	56.43	106.08	25.75	158.18	13.38
p-val	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
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

Note: This table presents results for our main specification when instrumenting CO_2 concentration levels during the school term using the number of days that the ventilation in the school was detected as broken by our algorithm. Each column shows second and first stage results for the subsamples indicated in the corresponding column name, and as described in Table 4. Kleibergen-Paap F-statistics and corresponding p-values are also reported. Clustered standard errors at the classroom by period level are shown in brackets and significance levels are $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

C.5 Temperature and noise during ventilation breakdown

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

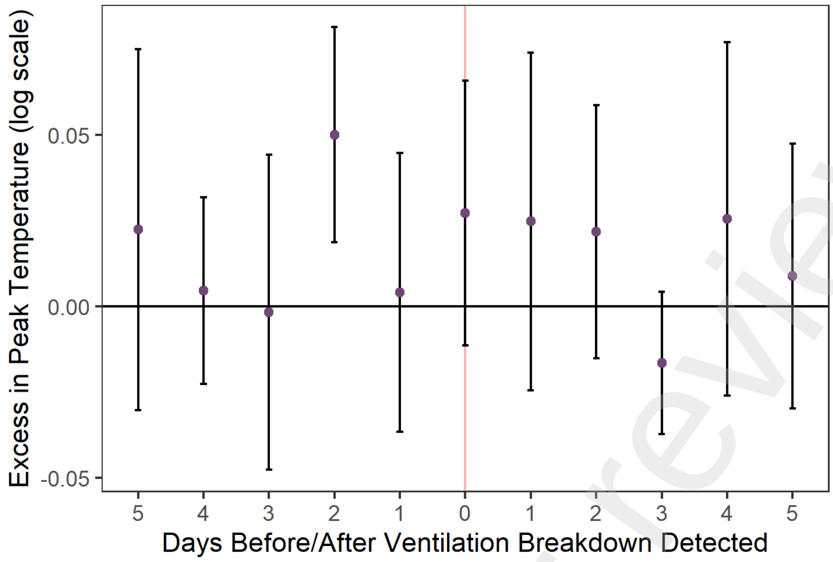
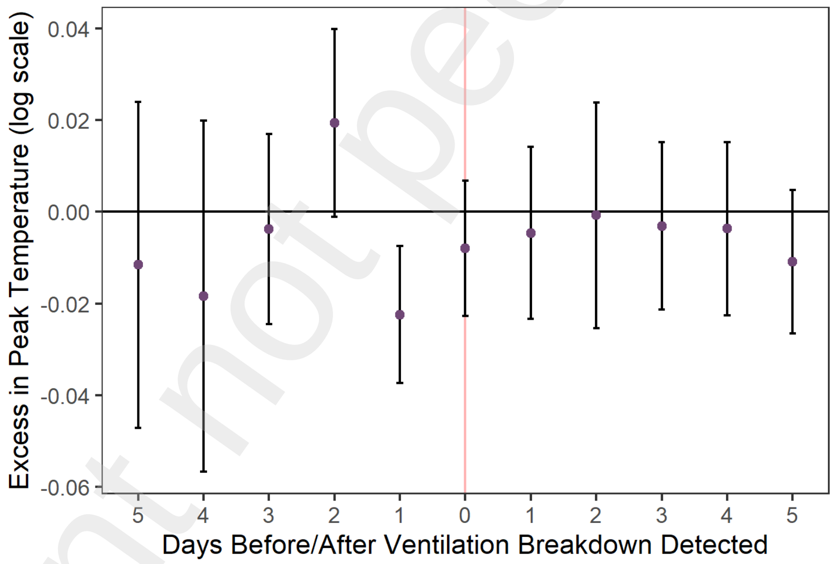


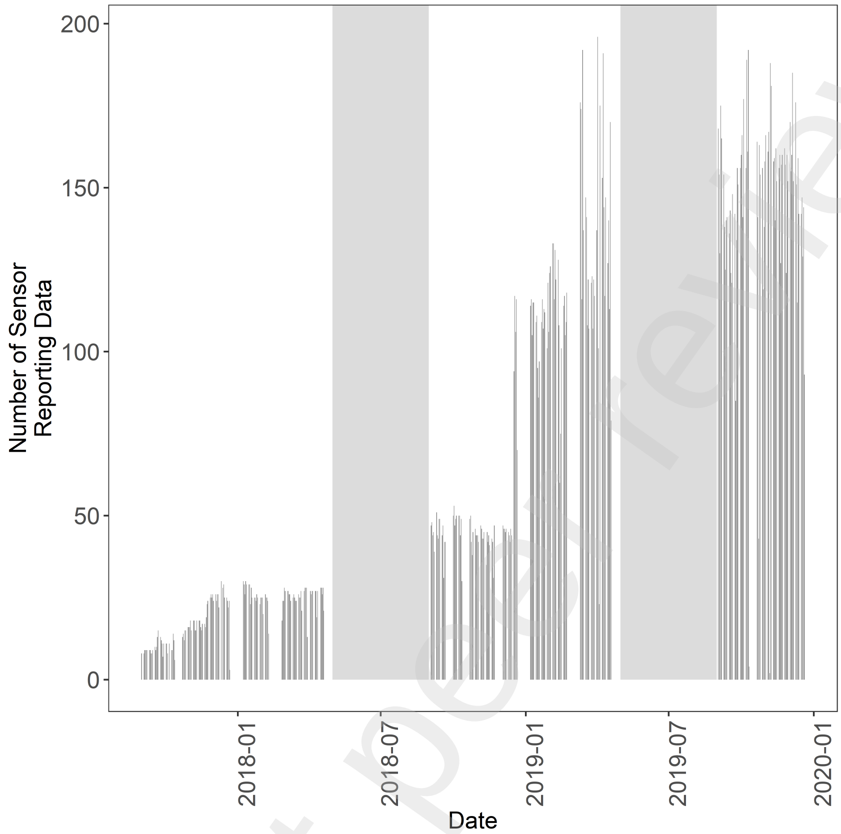
Figure C.3: Average Peak Noise at Days Before and After Ventilation Breakdown Detected



D Data Coverage and Algorithm to Detect Occupancy

D.1 Data Coverage

Figure D.2: Coverage of Sensors by Date

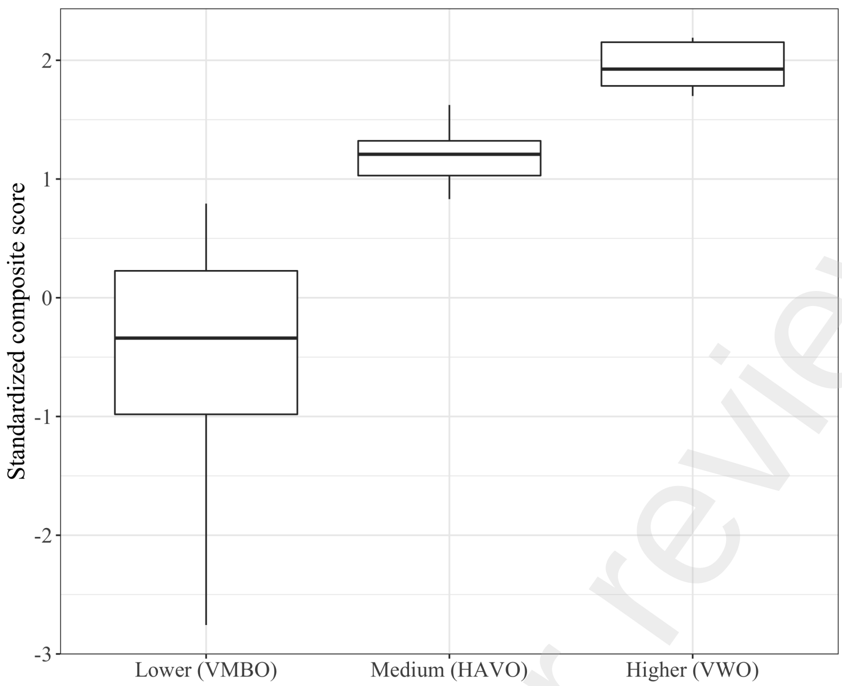


D.2 Secondary School Advice

D.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).

Figure D.3: Composite standard score (all tests) and high school level advice

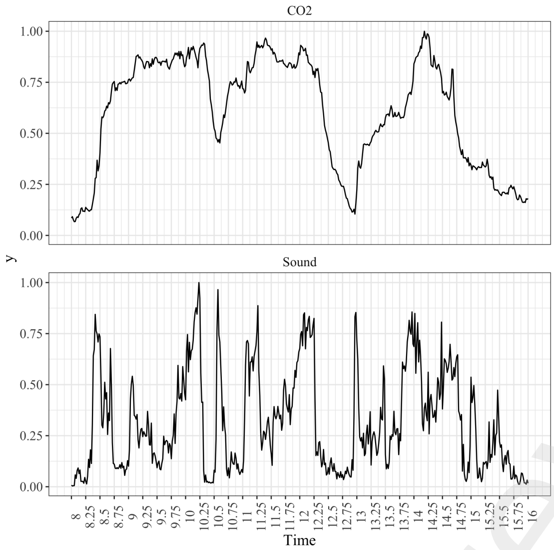


As described in the main text, the algorithm makes use of regularities in the behavior of both CO_2 and sound in the event of children entering or exiting the classroom. Those regularities are easily spotted in Figure D.4. The graph plots how CO_2 concentration and sound decibels move between 8am and 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).

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

Figure D.4: 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.

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 D.5 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.

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.

Figure D.5: Algorithm performance (F1 and R^2)