Royal Holloway, University of London

Doctoral Thesis

Essays in Environmental Economics and Human Capital

Author: Sefi Roth

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in
Economics

Declaration of Authorship

I confirm that the first and the third chapters of this thesis have been co-authored with Victor Lavy (University of Warwick, Hebrew University of Jerusalem and NBER) and Avraham Ebenstein (Hebrew University of Jerusalem).

Chapter 2 is entirely my own contribution.

Signed (Candidate):

Date:

ROYAL HOLLOWAY, UNIVERSITY OF LONDON

Abstract

Department of Economics

Doctor of Philosophy

Essays in Environmental Economics and Human Capital

by Sefi Roth

This PhD thesis examines the link between air pollution and human capital. The first two chapters evaluate the effect of short-term exposures to ambient and indoor pollution on test scores. I exploit the panel structure of the data to estimate models with individual fixed effects. I find that exposure to elevated levels of ambient and indoor pollution has a statistically and economically significant effect on high school exit exams in Israel and on university examinations in the UK. In the third chapter I study whether random disturbances during high-stakes examinations, induced by pollution, have long-term consequences for schooling and labor force outcomes. To do this I examine the same Israeli students 10 years after graduation and find that exposure to ambient pollution during their high school exit exams is negatively associated with post-secondary educational attainment and earnings. Overall, this thesis shows that a narrow focus on traditional health outcomes, such as hospitalization, may understate the true cost of pollution and highlights how reliance on noisy signals of student quality can lead to allocative inefficiency.

Acknowledgements

I would like to express my sincere gratitude to my Ph.D. supervisor, Professor Victor Lavy for the continuous academic and holistic support, patience, and immense knowledge.

I would also like to thank my thesis advisors, Dr. Arnaud Chevalier and Professor Jonathan Wadsworth for their outstanding encouragement, support and feedback.

Finally, I would like thank my colleagues at Royal Holloway, University College London and The London School of Economics for the inspiring, motivating and stimulating discussions.

Contents

List of Figures

List of Tables

Introduction

Air pollution has significant consequences for human health and life expectancy (Pope et al. 2009, Chay and Greenstone 2003). Researchers have documented that short-term exposure to air pollution may affect respiratory and cardiovascular conditions, such as asthma and heart attacks (Pope at al., 1995; Dockery, 2009; Donaldson et al., 2000; Weinmayr et al., 2010). Human intake of pollution can also lead to milder health effects such as irritation of the airways, coughing or difficulty breathing[1]. Despite the empirical evidence on links between air pollution and human health, the literature on the effect of pollution on cognitive performance is exceptionally scarce. A potential link between pollution and cognitive performance would imply that a narrow focus on traditional health outcomes may understate the true cost of pollution as mental acuity is essential to productivity in most professions.

The first chapter of this thesis evaluates this important relationship by estimating the effect of ambient pollution exposure on standardized test scores among Israeli high school high-stakes tests (2000-2002). Since students take multiple exams on multiple days in the same location after each grade, we can adopt a fixed effects strategy estimating models with city, school, and student fixed effects. we focus on fine particulate matter ($PM_{2.5}$) and carbon monoxide (CO), which are considered to be two of the most dangerous forms of air pollution. We find that while $PM_{2.5}$ and CO levels are only weakly correlated with each other, both exhibit a robust negative relationship with test scores. We also find that $PM_{2.5}$, which is thought to be particularly damaging for asthmatics, has a larger negative impact on groups with higher rates of asthma. For CO, which affects neurological functioning, the effect is more homogenous across

[1] For further details on such effects see http://www.epa.gov/pm/health.html

demographic groups. Furthermore, we find that exposure to either pollutant is associated with a significant decline in the probability of receiving a Bagrut certificate, which is required for college entrance in Israel.

In the second chapter I examine whether mental acuity is associated with indoor air quality using a sample of university final examination results from a British higher education institution. To the best of my knowledge this study is the first to estimate the effect of indoor air pollution on cognitive performance with indoor pollution measures. This is of particular importance as most jobs are indoors and mental acuity is essential to productivity in most professions. To account for potential confounders, I exploit the panel structure of the data to estimate models with subject and student fixed effects. I find that exposure to elevated levels of particulate matter (PM_{10}) has a statistically and economically significant effect on test scores and long-term academic indicators that are potentially correlated with future career outcomes. Furthermore, I find that the effect is larger among male, high ability and STEM subgroups and is present at levels considerably lower than current EPA standards. The results suggest that a narrow focus on traditional health outcomes, such as hospitalization, may understate the true cost of pollution as indoor air quality also affects productivity.

The third chapter evaluates whether random disturbances to cognitive performance during high-stakes exams can have permanent consequences for long-term schooling attainment and labor market outcomes. We examine this hypothesis among Israeli high school students who took a series of high stakes matriculation exams between 2000 and 2002. As a source of random (transitory) shocks to high-stakes matriculation test scores, we use exposure to ambient air pollution during the day of the exam. First, we document a significant and negative relationship between average $PM_{2.5}$ exposure during exams and student composite scores, post-secondary educational attainment, and earnings during adulthood. Second, using $PM_{2.5}$ as an instrument, we estimate a large economic return to each point on the exam and each additional year of post-secondary education. Third, we examine the return to exam scores and schooling across sub-populations, and find the largest effects among boys, better students, and children from higher socio-economic backgrounds. The results suggest that random disturbances during high-stakes examinations can have long-term consequences for schooling and labor market outcomes, while also highlighting the drawbacks of using high-stakes examinations in university admissions.

Chapter 1

The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation

1.1 Introduction

Ambient air pollution has significant consequences for human health and life expectancy (Pope et al. 2009, Chay and Greenstone 2003). Researchers have documented that short-term acute exposure to particulate matter decreases circulatory performance and leads to increased illness and hospitalization rates (Pope et al. 1995). Exposure to fine particulate matter is particularly dangerous since these small particles penetrate deep in to the lungs and may also affect other aspects of human life, such as cognitive performance, due to their impact on blood flow and circulation (Pope and Dockery 2006). Recent work has also demonstrated a link between carbon monoxide and higher incidents of respiratory and heart related emergency room visits (Schlenker and Walker 2011). Medical research has also identified symptoms that point to a diagnosis of carbon monoxide poisoning, including headaches, dizziness, and confusion (Piantadosi 2002). A potential link between cognition and harmful forms of ambient air pollution would suggest that the benefit of pollution reduction could be underestimated by focusing only on health outcomes (Chay and Greenstone 2005). However, evidence documenting a link between cognition and ambient air pollution is extremely limited. A potential link between cognitive performance and pollution exposure would imply high costs of pollution in terms of lost labor productivity, as mental acuity is critical to productivity for many occupations.

There are several challenges posed in trying to estimate the relationship between cognitive performance and air pollution. First, ambient pollution is often correlated with other factors correlated with wellbeing, such as wealth, generating a potential source of omitted variables bias that is similar to the challenges faced in measuring the health impact of air pollution. However, measuring air pollution's impact on cognition poses unique challenges as well. First, unlike with health problems, poor cognitive outcomes are generally not measured precisely. Whereas short-term dysfunction can result in a hospital admission, short-term cognitive decline is unlikely to be recorded. Even if short-term cognitive dysfunction results in injury, such as from a car accident, it is unlikely that this will be recorded in a systematic manner. In our study, since we observe students engaged in a difficult mental task with precise measurement of performance, it is more likely we can observe an effect (if there is one). A second issue is that cognitive tests (e.g. IQ) are only administered to self-selected groups, such as military recruits, making samples less representative than in samples of individuals exposed to air pollution with observed health outcomes. As we will describe, since the Israeli examination we analyze in our study is taken by nearly all high school students, and our dataset includes the entire universe of test takers, our results presumably have more external validity than results generated from a self-selected group.

In this paper, we examine a unique data set of merged high school high-stakes exit exams (*Bagrut* tests) and pollution data for the universe of Israeli test takers during 2000-2002 where we observe pollution and outcomes for over 400,000 subject examinations. Since we observe the same student at multiple test administrations following each year of high school, we can control for both time invariant features of both a school and of a particular student. The rigorous nature of the *Bagrut* tests and the precise scoring of the exams provide a context to analyze a potential link between cognition and air pollution, even if there are only modest declines in cognitive performance due to pollution. Furthermore, Israel's small size and well-developed monitoring system implies that most of its testing locations are near a station where we observe precise levels of pollution concentration. Lastly, Israel's ethnic heterogeneity provides a context to examine the responsiveness of different groups to pollution, and potentially distinguish between different mechanisms by which pollution may affect cognitive performance.

In this study, we examine the impact of fine particulate matter and carbon monoxide exposure on exam outcomes. These two pollutants are particularly harmful to human health, and are available in the data provided by the Israeli monitoring system. We find that a 10 unit increase in the ambient concentration of fine particulate matter ($PM_{2.5}$) as measured by the Air Quality Index (AQI) reduces *Bagrut* test scores by .46 points, or roughly 1.9% of a standard deviation of the *Bagrut* (σ =23.7). Alternatively, relative to a day with average air quality, a 1 standard deviation increase in the $PM_{2.5}$ AQI value (σ =22.81) is associated with a .65 point decrease in score, or 2.8% of a standard deviation. We also find that a 10 unit increase in the ambient concentration of carbon monoxide (CO), as measured by the Air Quality Index (AQI), reduces *Bagrut* test scores by .85 points, or roughly 3.5% of a standard deviation. This implies that relative to a day with average air quality, a 1 standard deviation increase in the CO AQI value is associated with a .54 point decrease in score, or 2.4% of a standard deviation. We also examine whether pollution has a non-linear impact on test takers using specifications where we include dummy variables for clean, moderately polluted, or very polluted days. We find that our results are largely driven by poor performance of test takers on very polluted days, with an AQI reading above 101 for $PM_{2.5}$ associated with a decline in test score of 1.95 points, or 8.2% of a standard deviation. For CO, test administrations in the top 5% of most polluted days are 10.16

points lower, a decline of 42.8% of a standard deviation. These results suggest that modest pollution levels have only a marginal impact, but very polluted days can have much larger impacts, suggesting a non-linearity in pollution's relationship with cognitive performance. In several placebo exercises, we find that the correlation between *Bagrut* test scores and pollution readings other than the test pollution level is insignificant in most specifications, further supporting our claim of a causal interpretation to our results. Our results also indicate that test outcomes for afternoon examinations are more affected by carbon monoxide than morning examinations. This is consistent with a prior that carbon monoxide, which is generated primarily by automobile emissions, will worsen over the course of the day. Our results for fine particulate matter, which are primarily the byproduct of sandstorms and coal-burning power plants, are more similar for morning and afternoon examinations.

We examine mechanisms for our findings by estimating treatment effects for different groups in Israel, in combination with a prior on how each pollutant should affect test takers. In particular, we find that demographic groups with higher rates of asthma have larger treatment effects of $PM_{2.5}$, suggesting that exacerbation of respiratory health problems could be a mechanism for pollution to affect test outcomes. Our results for $PM_{2.5}$ seem to be consistent with the patterns of relative risk for asthma found by Laor et al. (1993) from military records in Israel, which reflect much higher incidence among boys and Ashkenazi Jews, and among lower socio-economic groups in other countries (Basagana et al. 2004, Eriksson et al. 2006). Carbon monoxide exposure, which is thought to decrease neurological functioning, has a more homogenous impact on Israel's demographic groups. This may be due to a more similar responsiveness to carbon monoxide poisoning, which may affect all individuals, even those without prior respiratory conditions.

We also find that exposure to $PM_{2.5}$ or CO on examination days has a significant impact on a particular student's long-term academic outcome, and potentially has implications for the welfare consequences of using the *Bagrut* for ranking students. We find that a one standard deviation increase in the fraction of exam days that are heavily polluted is associated with a 2.19 and 2.70 percentage point decline in the probability of receiving a *Bagrut* matriculation certificate for $PM_{2.5}$ and CO respectively. Note that this certificate is a prerequisite for college entrance, preventing some students from accessing higher education. In addition, since access to college majors is also determined by *Bagrut* performance, air pollution may have long-term consequences for students who pass the *Bagrut* but are forced to choose a less desirable college major. An implication of this finding is that by temporarily lowering the productivity of human capital, high pollution levels lead to allocative inefficiency as students with lower human capital are assigned a higher rank than their more qualified peers. This may lead to inefficient allocation of workers across occupations, and possibly a less productive workforce. The results highlight the danger in assigning too much weight to a student's performance on a high-stakes exam, rather than their overall academic record.

Our results provide novel and compelling evidence that cognition is affected by air pollution exposure. Epidemiologists have examined the relationship between air pollution and cognition, but the evidence is generally cross-sectional in nature, with little attention paid to a potential correlation between omitted variables and pollution. For example, Suglia et al. (2008) found that in a sample of 202 children, those living near higher levels of black carbon (which is a solid fraction of $PM_{2.5}$) performed worse on cognitive function assessments. Wang et al. (2009) found that children in higher-traffic areas (with higher levels of carbon monoxide) performed worse on neurobehavioral examinations. Both of these studies, however, were cross-sectional in nature and did not account for a potential correlation between unobservable determinants of test outcomes and the measures pollutants. Our examination of Israeli *Bagrut* exams is, to our knowledge, the first attempt to measure fine particulate matter and carbon monoxide's impact on cognitive performance using rich data and a panel approach.[2] Our results underscore the need for tighter pollution regulations relative to policy made taking only human health effects into account. The results may also highlight a mechanism by which individuals in highly polluted areas, such as those living in cheaper industrial areas of cities, could have economic disadvantage exacerbated by pollution (Brown 1995).

The rest of the paper is laid out as follows. In the second section, we present background on the Israeli context, and summarize in greater detail the relevant existing work on acute air pollution and human welfare. Section III presents our data and Section IV presents our empirical strategy. In Section V, we present our empirical results and in Section VI we conclude.

1.2 Background and Data

1.2.1 Air pollution and Cognitive Performance

We consider two air pollution measures. Our first air pollution measure is particulate matter ($PM_{2.5}$), which is a complex mixture of solid and liquid microscopic droplets found in the air that consists of various components including acids, metals, dust particles, organic chemicals and allergens. In Israel, the main sources of particulate matter are sand storms, coal-burning power plants, and certain industrial processes. Our second air pollution measure is carbon

[2] These results contribute to a limited but growing literature in economics documenting that a narrow focus on hospitalization rates or excess mortality rates may understate the impact of air pollution on human wellbeing, though these studies focus primarily on consequences of illness rather than a direct impact on cognition. Currie et al. (2009) find that carbon monoxide exposure increases absenteeism among elementary and middle school children students. Oliva and Hanna (2015) present evidence that labor supply is reduced in Mexico City on days with high pollution levels. Ham et al. (2011) examine the relationship between pollution and test scores using data from California elementary schools. They find significant but modest effects for ozone, fine particulate matter, and coarse particulate matter. However, they are unable to observe the same student over multiple examinations, and are therefore forced to rely on grade-school fixed effects.

monoxide, which is generated by automobile emissions, fossil-fuel furnaces, and fires (Piantadosi 2002). Human intake of particulate matter or carbon monoxide inhibits proper blood flow, leading to elevated risk of heart disease, stroke, and lung cancer (Dockery and Pope 1996; Schlenker and Walker 2011). It is less clear whether either of these air pollutants affect cognition. Since the brain consumes a large fraction of the oxygen needs of the body, any deterioration in oxygen quality can in theory affect cognition (Clark and Sokoloff, 1999). Long-term exposure to ambient pollution can lead to the growth of white-matter lesions, potentially inhibiting cognition (Calderón-Garcidueñasetal et al. 2008). Air pollution can also impact the nervous system, leading to symptoms such as memory disturbance, fatigue and blurred vision (Kampa and Castanas, 2008), and may also impact labor productivity (Graff Zivin and Neidell 2012). Fine particle matter can also travel through small passageways, suggesting that high levels of pollution may affect test takers even indoors (Branis et al. 2005). These papers provide compelling evidence that cognition may be affected by pollution. They also suggest that while particulate matter may affect the respiratory system, carbon monoxide will primarily affect the release of oxygen to human tissues, including the brain. This implies that particulate matter may have a larger impact on sensitive or unhealthy groups – such as asthmatic groups – while carbon monoxide will affect healthy and unhealthy groups more similarly. However, as stated by Suglia et al. (2008), “the possible neurodegenerative effect of air pollution remains largely unexplored.” The Israeli *Bagrut* examination provides a unique context to assess the relationship empirically, which is discussed in the next section.

1.2.2 The Israeli High-School Matriculation Exam System

Israeli post-primary education consists of middle school (grades 7–9) and high school (grades 10–12). High-school students are enrolled either in an academic track leading to a matriculation certificate (*Bagrut* in Hebrew[3]) or in a vocational track leading to a high-school diploma. The matriculation certificate is a prerequisite for university admission and is one of the most economically important education milestones. Students complete the matriculation process by passing a series of national exams in core and elective subjects following tenth grade and eleventh grade, and then a larger set following twelfth grade. Students choose to be tested at various levels of proficiency, with each test awarding the student between one and five credit units per subject, varying by the difficulty of the exam. The exam focuses on seven mandatory subjects and one elective subject, allowing us to observe students completing exams with separate grades for each subject.[4] The most basic level of study is three credits and a minimum of twenty credits is required to qualify for a matriculation certificate. About 52 percent of high-school graduates in 2002 and 46 percent of the overall cohort received matriculation certificates.[5]

The examinations are given bi-annually during the two exam "seasons", a winter examination given in January and a summer examination in May/June, and are graded by two independent and anonymous examiners. The *Bagrut* final score in each subject is a simple average of the *Bagrut* exam score and a school score, or *Magen* score, on this subject. The *Magen* score is based on a school exam (the *Matchonet* examination)[6] that precedes the *Bagrut* exam by week to three weeks and has the same format as the nationally-administered *Bagrut* exam, except that it is graded by the student's secondary school subject teacher and on the

[3] Many countries and some American states have similar high-school matriculation exams, e.g., the French Baccalaureate, the German Certificate of Maturity (*Reifezeugnis*), the Italian Diploma di Maturità, the New York State Regents examinations, and the recently instituted Massachusetts Comprehensive Assessment System.

[4] The seven core subjects are Math, English, Hebrew, History, Literature, Religious Studies and Civics. It is possible to be awarded a *Bagrut* certificate despite a failing mark on one of the exams if one of following conditions is satisfied: (1) the mark is not below 45 (2) the mark is below 45 but the candidate has two more exams with 3 credit units or more that their scores combined sums to at least 150 (3) the failing mark is not in the Hebrew subject exam.

[5] See the Israel Ministry of Education web site (www.education.gov.il) and Lavy (2002).

[6] This exam is called *matkonet*, derived from the word *matkon*, recipe, meaning that the school exam follows the "recipe" of the state exam. The school exam follows the exact format of the *Bagrut* exam and it also draws its questions from the same bank of questions used for the *Bagrut* exam.

student's overall performance in this subject during the academic year. We only observe the overall *Magen* score and not its two components. The weights of these two factors can vary and the overall *Magen* score is therefore a natural measure for ranking the students in terms of quality which we use in our analysis to stratify the sample.

Students are admitted to post-secondary programs on the basis of their average matriculation scores and based on an SAT-style examination from a psychometric examination administered by the National Testing Center. Each higher education institution ranks applicants according to the same formula, thus producing an index based on a weighted average of the student's average score on all her matriculation exams and the SAT-style examination. Therefore, pollution levels can affect students' post-secondary schooling by affecting their probability of passing *Bagrut* exams, and also by affecting the average score in these exams. The first channel will affect the eligibility for post-secondary admission while the second will affect which programs (or majors) will be available to the student.

1.2.3 Data

Our data set is generated by combining Israeli test score data with air pollution and meteorological data for 2000-2002. The *Bagrut* exam information and demographic information for each test taker were provided by the Israeli Ministry of Education. These files also contain each student's Israeli identification number, allowing us to observe rich demographic information on the student and the student's family, such as parental education level, country of origin, and ethnicity. For each exam, we also know the exact date of the test and the precise location of the testing site, allowing us to assign pollution measures to each test administration. Our pollution data are taken from files published by the Israeli Ministry of Environmental Protection, which reports daily mean readings of particular matter less than 2.5 microns in width, or $PM_{2.5}$ ($\mu g/m^3$) and carbon monoxide (CO) at 139 monitoring stations throughout Israel for the sample period (see Figure 1.1).[7] Readings are taken at 5 minute intervals and averaged over the course of the day.

Each school is assigned the average pollution reading for all monitoring stations within the city limits in which it is located, or within 2.5 kilometers of the city limits. Since Israeli cities are not very large, we generally are taking readings from stations very close to the schools. While we ideally would have a measure of pollution inside the classroom, the air quality inside a school is presumed to be highly correlated with the ambient reading outdoors (Branis et al. 2005). Schools that had no monitoring station within the city limits or 2.5 kilometers of the city limits were dropped from the sample.[8] These stations also record temperature and relative humidity, which are used as control variables. We assign pollution and weather to each test by averaging all non-missing values among stations within 2.5 kilometres of the test site. Our analysis is executed using the Air Quality Index (AQI) measurement associated with our $PM_{2.5}$ and CO readings. The AQI measure converts the pollutant measures in micrograms ($\mu g/m^3$) into an index score that ranges from 0-500 using a formula specified by the US EPA.[9] The US defines values above 101 as "unhealthy for sensitive groups" and values above 150 as "unhealthy". In our empirical analysis, we classify air quality on a particular day as being beyond the threshold if the $PM_{2.5}$ (AQI) reading is greater or equal to 101 (AQI). Since Israel's CO measures are

[7] The Israeli monitoring system also records readings for a set of other pollutants. In this paper we focus on $PM_{2.5}$ and CO since these are considered among the most harmful, and are monitored most extensively by the monitoring system. We also examined the relationship between PM_{10} and SO_2 and test score data, finding zero effects for PM_{10} and modest effects of very high levels of SO_2. The results are available from the authors upon request.

[8] Since Israel's population is densely concentrated in several metropolitan areas, this led to the dropping of less than 5% of schools.

[9] We used the EPA's breakpoints table (see Table A3.1) and the following formula to generate the $PM_{2.5}$ (AQI) measurement: $PM_{2.5}\ (AQI)=\{(I_{HI}-I_{LO}) / (BP_{HI}-BP_{LO})\}\ (C_P-BP_{LO}) + I_{LO.}$ Where C_P is the rounded concentration of the pollutant, BP_{HI} is the breakpoint that is greater than or equal to C_P, BP_{LO} is the breakpoint that is less than or equal to C_P, I_{HI} is the AQI value corresponding to BP_{HI} and I_{LO} is the AQI value corresponding to BP_{LO}. A similar formula is used for CO.

relatively lower, and have few days where the AQI score exceeds 101, we chose a lower threshold where we generate a dummy for a test occurring on a day in the top 5 percent of the most polluted days.

It is also worth noting that the correlation between our two measures of pollution is very close to zero (R=.0028).[10] The two pollutants are associated with different causes: particulate matter is generated by sand and dust storms and coal-powered electric plants, whereas carbon monoxide is associated with high traffic density or other combustion processes. As such, this provides an opportunity to exploit two different and largely independent measures of pollution to assess the link between air pollution and cognitive performance.

The summary statistics for our sample are presented in Table 1.1. Our sample includes 489,419 examinations taken by 71,383 students at 712 schools throughout Israel. Our key variables are the measures of $PM_{2.5}$ ($\mu g/m^3$ or AQI), CO, and our standardized test outcomes in *Bagrut* exams. We also use the *Magen* score as a proxy for student quality that can be used to stratify the sample. This score is determined primarily by a test which is a similar to the *Bagrut* test a few weeks prior to the *Bagrut*, and by the student's course grade. In columns (2)-(4), we stratify the sample by sex and *Magen* score. The table indicates that girls perform somewhat better than boys on the exam. Also, as anticipated, the students who have higher *Magen* scores, and are placed in our top quality group, have on average higher *Bagrut* scores and come from more educated families. The parents of the higher *Magen* group have, on average, some postsecondary schooling, whereas parents of the lower *Magen* group have, on average, less than a high school diploma. The lower *Magen* group also come from larger families (more siblings)

[10] In a robustness check, we estimate whether there is a conditional correlation between the two pollutant measures by estimating models with the two pollutants simultaneously, shown in Table A1.2. Our results indicate that the results are largely unchanged by including both measures in the same model for our dichotomous measure of pollution, but the result for our continuous measure of $PM_{2.5}$ is significantly reduced for models with student fixed effects. This is discussed in the empirical results section.

and are more likely to be of African/Asian ethnic origin (*Sephardi*). These means are not shown in the table but are available from the authors, and provide additional indication that the students in the low *Magen* group are from more disadvantaged socio-economic backgrounds. This characterisation of the low *Magen* group will feature in our later discussion of potential mechanisms for our results, since the incidence of asthma is much higher among disadvantaged populations. Also note that Table 1.1 indicates that air pollution, temperature, and humidity do not vary by gender or *Magen* score. For example, the mean $PM_{2.5}$ AQI index for boys and girls is similar: 59.5 and 59.9, less than a tenth of a standard deviation. Similarly, the average $PM_{2.5}$ AQI index value for high and low *Magen* students are 60.1 and 59.5 respectively, and the mean temperature of the days of exams for both groups is 23.8. The balancing of our data on observables when stratified by these groupings by gender and by our measure of student quality is important in light of our findings that the effect of pollution is very different for these sub-populations. We discuss this further in the empirical section, but the similarity on observables is suggestive evidence that selection on unobservables is unlikely to be driving our results.

1.3 Empirical Strategy

Our estimation strategy is relatively straightforward. We estimate OLS models where we examine the partial correlation between our air pollution measures and test scores outcomes. For identification, we crucially rely on the panel structure of the data and the repeated nature of the *Bagrut* exam. Since we observe the exact location of the test, we can include city or school fixed effects. Since we observe the students taking exams following each grade, we can include student fixed effects. Formally, the models we estimate are of the following form:

$$(1)\ R_{ist} = \beta_0 X_{it} + \beta_1 POL_{st} + \beta_2 Temp_{st} + \beta_3 RH_{st} + M_t + L_l + I_i + \varepsilon_{ist}$$

where R_{ist} is the test score of student i at school s at time t; X_{it} is a vector of individual characteristics possibly related to test outcomes, such as parental education[11]; POL_{st} is our measure of air pollution ($PM_{2.5}$ or CO) at school s at time t; $Temp_{st}$ is the mean temperature[12] at school s at time t; RH_{st} is the relative humidity measure at school s at time t; M_t and L_l are month and exam proficiency level fixed effects respectively; I_i is our fixed effect for the individual; and e_{ist} is an idiosyncratic error term. Note that in different specifications we will use city or school fixed effects in place of our individual fixed effects.

The key identifying assumption for inferring a causal relationship between pollution and test scores estimated by equation (1) is that unobserved determinants of student's test scores are uncorrelated with ambient pollution. Without any fixed effects to absorb unobserved variation in schools or individuals, this assumption is likely violated since it is likely that pollution is correlated with time invariant features of a testing location or a particular student. For example, if poorer schools are located in more polluted parts of cities, OLS will likely overstate the causal link between pollution and test scores. Conversely, if schools in denser (and wealthier) cities have more pollution exposure, OLS might understate the true cost of pollution, as it is mitigated by other compensating factors (e.g. tutoring). More generally, endogenous sorting across schools, heterogeneity in avoidance behavior, or measurement error in assigning pollution exposure to individuals will all bias results that do not properly account for unobserved factors correlated with both our outcome of interest and ambient pollution (Moretti and Neidell 2011). In our setup, since we account for time invariant features of schools and students with fixed

[11] Our results with individual fixed effects exclude individual controls.
[12] In the empirical analysis, we include linear and quadratic terms in both temperature and humidity, and linear and quadratic interaction terms of the two variables.

effects, the challenge relevant to our estimation is to account for omitted variables that are varying over time but are potentially correlated with pollution and *Bagrut* outcomes. For example, if weather or traffic the day of the exam is correlated with pollution, our fixed effects models will fail to identify the true effect. In our empirical analysis, we include controls for time-varying factors that could be contemporaneous with pollution, such as daily temperature and relative humidity, but of course it is untestable whether there are factors that are unobserved that are both correlated with pollution and *Bagrut* exam scores. As such, we conduct a rich set of robustness checks and placebo tests. These are discussed further in the next section.

1.4 Empirical Results

1.4.1 Main Results

In Table 1.2, we report our baseline results of the relationship between the Air Quality Indicator values for $PM_{2.5}$, CO, and *Bagrut* test scores. In columns (1) and (2) of Panel A, we report the correlation between *Bagrut* scores and a continuous measure of $PM_{2.5}$ (AQI) using OLS without city, school or student fixed effects. In column (1), we estimate that a 1 unit increase of $PM_{2.5}$ is associated with a 0.055 points decrease in a student's test score, significant at the 1% level. The results also indicate that a relatively small part of the variation in test scores (R-squared = 0.003) is explained by air pollution. This result indicates, as one would expect, that variables other than air pollution are responsible for the vast majority of the variation in test scores. In column (2) we report the results with the addition of controls for parental education, sex, temperature, relative humidity and dummies for the month of the exam and difficulty of the exam. The results are similar and slightly larger in magnitude, with our coefficient estimate indicating that a 1 unit increase in pollution is associated with a 0.065 decrease in a student's score. Note that the sample with controls is roughly 20% smaller, as we have incomplete demographic information for these individuals. The similarity of the results with and without controls, and with the smaller sample size, is suggestive that there is no strong correlation between observables and pollution. We also used the smaller sample to estimate the OLS regression without any controls and obtained estimates almost identical to those reported in column 1, which suggest the sample of students with some missing characteristics is not on average selectively different from the rest of the sample.

In columns (3)-(5) of Table 1.2, we take advantage of the panel structure of our data and include city, school, and student fixed effects, respectively. These account for variation in time-invariant unobserved heterogeneity that could be correlated with ambient pollution. The estimates from a regression with city or school fixed effects in columns (3) and (4), are somewhat larger, with estimated coefficients of -0.082 and -0.069 respectively. Adding student fixed effects weakens the results slightly, with our preferred estimate indicating that a 1 unit increase in $PM_{2.5}$ is associated with a 0.046 (σ =0.007) decline in the *Bagrut* score. This estimate implies that a test score in an exam on a day with average pollution (AQI=59.74) will be lowered relative to an exam taken on a day with the minimum pollution level (AQI=10.1) by 0.10 (.046*(59.7-10.1)/22.8) standard deviations. Our results for CO in columns (6-10) largely mirror our results in columns (1-5). Our results in column 10 indicate that a 1 unit increase in CO is associated with a 0.085 (σ =0.017) decline in the *Bagrut* score, significant at the 1 % level. Note however that since the Israeli monitoring system failed to collect CO readings at all stations during our sample period, our $PM_{2.5}$ analysis is based on a much larger sample.[13]

[13] We investigated whether there was something systematic about which stations did not collect CO measures. We found no noticeable pattern in our data, though coverage for CO was much poorer in northern Israel and in the areas surrounding Haifa.

In Panel B, we perform a similar analysis but replace our continuous measure of pollution with a dichotomous indicator for whether the test occurred in a day classified as having "poor" air quality. The results are qualitatively similar to the results using the continuous measure for $PM_{2.5}$ but much larger for CO. Specifically, in our specification in column 5 where we include student fixed effects, the data indicate that having "poor" $PM_{2.5}$ air quality the day of the exam is associated with a 1.95 point decline in the student's *Bagrut* score, equivalent to 8.2% of a standard deviation. Our specification in column 10 indicates that having "poor" CO air quality the day of the exam is associated with a 10.16 point decline in the student's *Bagrut* score, equivalent to 42.8% of a standard deviation. [14]

The effect of $PM_{2.5}$ on *Bagrut* scores for the 99th percentile of exposure in our sample (AQI=137) is very large and implies a decline of roughly a sixth (.149) of a standard deviation relative to an average day. This effect is similar to the estimated effect of reducing class size from 31 to 25 students (Angrist and Lavy, 1999) and larger than the test scores gains associated with paying teachers large financial bonuses based on their students' test scores (Lavy, 2009). Unfortunately, days with elevated levels of particulate matter are not unusual in Israel and in neighboring countries in the Middle East, as they are often the result of sandstorms that originate in the Sahara desert and are relatively common in the spring and summer months, with serious health effects (Bell et al. 2007). For CO, our results similarly suggest a large response of students to very poor days. The 99th percentile of CO, AQI=56, would imply a similar decline of .158 standard deviations relative to a day with average levels of CO. Since Israel's CO level is actually quite similar to the levels found in other large cities, such as Los Angeles, CA,[15] and

[14] It is also worth noting that the CO results for our threshold measure of pollution may be affected by several *extremely* polluted exam administrations. In the highest CO reading, students were subjected to AQI=270, roughly twenty times the average reading.
[15] http://www.usa.com/los-angeles-ca-air-quality.htm

may indicate that these results may affect student performance in polluted areas of these cities as well.

In light of the fact that $PM_{2.5}$ and CO are only weakly correlated, these results suggest a robust relationship between different air pollution measures and test scores, as two largely independent pollution measures are associated with appreciable declines in test scores. To explore the role of each pollutant further, in Table A1.2 we estimate models where both pollutants are included simultaneously. The results indicate that our dichotomous measure of each pollutant's impact is extremely robust to simultaneous estimation, and the continuous measure for CO is almost unchanged by the inclusion of $PM_{2.5}$. However, our continuous measure of $PM_{2.5}$ is weakened by inclusion of CO in models with student fixed effects. This may be because our sample for $PM_{2.5}$ is more than twice as large as our sample for CO, and partly due to a weak residual correlation with CO.

In Table 1.3, we report results where we examine whether pollution has a non-linear impact on test takers using specifications where we include dummy variables for clean, moderately polluted, or very polluted days.[16] For $PM_{2.5,}$ we define moderately polluted days as days where the AQI score ranges from 51-100 (which the EPA defines as moderate pollution) and AQI scores above 101 (which the EPA defines as unhealthy for sensitive groups) as poor or very polluted days (see Table A1.1). Since our CO scores are consistently lower than our $PM_{2.5}$ scores (a mean score of 13 versus a mean score of 59 for $PM_{2.5}$), we define moderately polluted days as days above the median pollution level and below the top 5% of the most polluted days, and very polluted days as the top 5% of the sample's CO readings. Column 5 indicates that having poor air quality from $PM_{2.5}$ exposure the day of the exam is associated with a 2.89 point

[16] It is worth noting that students cannot reschedule their examination, and so avoidance behavior in response to high pollution on the day of the *Bagrut* is unlikely to be common.

decline in the student's *Bagrut* score, which is more than double the size of the coefficient for moderately polluted days. Similarly, Column 10 indicates that having "poor" CO air quality the day of the exam is associated with a 10.89 point decline in the student's *Bagrut* score, which is more than ten times the size of the coefficient for moderately polluted days. These results indicate that our results are largely driven by poor performance of test takers on very polluted days, suggesting that pollution's impact on cognitive performance is mostly relevant on days with very poor air quality.

1.4.2 Placebo Tests

In this section, we perform a set of placebo tests where we examine the relationship between air pollution on days *other than* the actual exam and exam scores. In Table 1.4, we examine whether there is a correlation between pollution from the day of the previous *Bagrut* and the score on the exam. Note that since students take the *Bagrut* exams over a short period of time, this will generally be a pollution reading taken from several days prior. As shown in Panel A of Table 1.4, the correlation between *Bagrut* outcomes is weak relative to the correlation with the actual exam. While some of the specifications are statistically significant, our preferred specification with student fixed effects are either statistically insignificant, or with the wrong sign. For example, in our estimates using our threshold measure with student fixed effects, the impact of $PM_{2.5}$ during the previous exam is a .78 point *increase* in the student's score, and the result is not statistically significant. This can be compared to our main result using the $PM_{2.5}$ reading from the day of the *Bagrut*, where poor air quality reduces scores by 1.9 points (significant at the 1% level). For our dichotomous measure of CO, the results are also reassuring:

after including school or student fixed effects, no significant relationship between the placebo pollution reading and the exam score is observed.

In Panel B, we perform a similar exercise but using the air pollution on the date exactly one year before the exam. For the continuous measure of pollution, column 5 indicates a negative and statistically insignificant relationship between $PM_{2.5}$ and test scores, while column 10 indicates a *positive* and statistically significant relationship between CO and test scores. For our dichotomous measure of pollution, we observe a correlation between exams and $PM_{2.5}$ in the previous year when we include no fixed effects: having a day classified as polluted in the previous year is associated with a 2.8 point decline in scores in models with controls, even though there should be no relationship. This underscores the importance of including fixed effects to absorb a time-invariant correlation between pollution and student quality, and suggests that more polluted areas have lower exam scores in general. Once we include student fixed effects in our models, the correlation between $PM_{2.5}$ from the previous year and the *Bagrut* score declines to 1.15 points, and it is only marginally significant. For the dichotomous measure of CO, the results for the previous year's reading are counter-intuitive: we find a *positive* correlation between pollution levels from the previous year and exam scores. While this result is surprising, it suggests that our CO results may be less stable than our $PM_{2.5}$ results due to a smaller sample size and more extreme values for pollution. The results for CO, therefore, should be interpreted with greater caution.

In Figure 1.2, we examine the impact of $PM_{2.5}$ and CO from three days prior to the exam, the day of the exam, and three days following the exam on test scores. As shown in the figure, the main effects of $PM_{2.5}$ are concentrated on the day of the exam, and no significant relationship between pollution readings and the exam score is observed for days before and after the exam.

The figure indicates that the coefficient on pollution *the day of* the exam is much larger and more negative than the other days: an additional 100 units of AQI is associated with a 0.2 point decrease in student scores, and the coefficient estimates are small and positive on the days before and after the exam. In contrast, the results for CO are less conclusive, with somewhat larger negative coefficients for the day of the exam relative to the days before and after. As such, our results for CO should be interpreted with greater caution.

In Table A1.3, we exploit the fact that we know the exact time of day that the examination was administered, and consider whether our pollutants have different effects at different times of day.[17] While the majority of our sample is given a 9AM examination time, roughly 40% of examinations are given after 12PM. **W**e posit that fine particulate matter, which is generated from sandstorms and coal-burning plants, will affect students throughout the day in a similar manner at all hours of the day (or night). Carbon monoxide is produced primarily by automobile emissions, and is likely to be more relevant for exams later in the day. As shown in the table, our coefficient estimates for $PM_{2.5}$ are relatively similar for both afternoon and morning examinations. In our preferred student fixed effect specification, we find that having poor air quality from our $PM_{2.5}$ exposure measure for an afternoon exam is associated with a .045 point and .054 point decline per unit of AQI respectively. Likewise, our results using the dichotomous measure are similar; we observe a 3.16 point decline in the student's *Bagrut* score for days with very high AQI in afternoon exams, which is about 20% larger than the coefficient for morning exams. For CO, our estimates are much larger for afternoon exams using both the continuous and dichotomous measures.[18] For example, using the dichotomous measure, having

[17] As an additional robustness check, we also estimate our main models with fixed effects for the day of the week on which the exam is given. The results are largely unchanged, and available upon request.
[18] Note that since we have fewer observed tests for each student, our results using student fixed effects will be less stable.

poor CO air quality for afternoon exams is associated with a 10.45 point decline in the student's *Bagrut* score, which is almost ten times the size of the coefficient for morning exams. The results are consistent with a prior that carbon monoxide exposure should be more problematic later in the day, and the results for particulate matter will be similar at different times.

1.4.3 Heterogeneity and Implied Mechanisms

In this section, we examine heterogeneity in the treatment effects reported in Table 1.2. Our interest is twofold. First, we wish to identify whether there are sub-populations that may be particularly responsive to poor air quality. Second, this may help to identify mechanisms for the observed reduced form relationship between air pollution and cognition. In particular, our prior is that $PM_{2.5}$ which affects the respiratory system will have a larger impact on weaker groups who are more sensitive to poor air quality. In contrast, we expect that CO, which affects the tissues and neurological system, to have a more similar impact across different groups.

We build on a set of stylized facts regarding who would be most sensitive to poor air quality from the medical literature. First, Israeli boys are more likely to be asthmatic than Israeli girls. As shown by Laor et al. (1993) the rate of asthma incidence in Israel is 25 percent higher among boys. Second, children of lower economic status are known to have higher rates of asthma and respiratory illnesses (Eriksson et al. 2006, Basagana et al. 2004). Third, Laor et al. (1993) also found that *Ashkenazic* Jews (ethnic origin from America and Europe) have 63% higher incidence of these illnesses than *Sephardim* (ethnic origin from Africa and Asia). This gives a rich set of potential comparisons for gauging whether asthma is a mechanism for the observed reduced form relationship between pollution and exam outcomes.

In Table 1.5, we examine our results separately by gender. The results highlight that men are significantly more likely to have their test outcomes affected by $PM_{2.5}$ than women. Our results indicate that treatment effects among men are between 2 and 4 times larger than among women. For example, in models with student fixed effects, we estimate that an additional 10 units of $PM_{2.5}$ (AQI) is associated with a .078 point decline among men and a .021 decline among women. We posit that the difference could be generated by the different asthma rates in these cohorts. Another possibility is that male students are more likely to be affected by small cognitive decline and distraction, consistent with higher rates of Attention Deficit Disorder in males (Biederman et al. 2002). In contrast, the results for CO are largely similar for men and women, with the results for men being moderately larger. For instance, in our model with student fixed effects, we estimate that an additional 10 units of CO (AQI) is associated with a .099 point decline among men and a .075 decline among women.

In Table 1.6, we break down our sample of test takers by our ex-ante expectation of their performance. This is proxied by their *Magen* score, which is a reasonable measure of student quality as it reflects their achievement in the full-year class and on a test similar to the *Bagrut*, and is correlated with family income and other measures of wellbeing because it is highly correlated with parental schooling, family size and ethnic origin. It may be that poorer families are more affected by air pollution as well, due to lower ability to engage in compensating behavior (Neidell 2004). Poorer children also have higher incidence of asthma (Basagana et al. 2004, Eriksson et al. 2006). When we stratify the students by whether their *Magen* score is above or below the median, our estimated treatment effects for $PM_{2.5}$ are more than two times larger among those classified as low quality. For a low quality student, we estimate that a 10 unit increase in $PM_{2.5}$ (AQI) is associated with a .061 point decline versus only a .028 point impact among higher quality students. However, we see no large difference between the responsiveness of higher and lower quality students to CO when using the continuous measure for AQI. This is consistent with our earlier results that CO's effect may be less heterogeneous. However, when using our dichotomous measure of pollution, both $PM_{2.5}$ and CO have larger effects on weaker students.

The results by student quality are investigated further in Table A1.4, which reveals that when the sample is stratified into quartiles, there is a monotonic relationship between treatment effects and our student quality measure for $PM_{2.5}$. Specifically, using our continuous measure of $PM_{2.5}$, we find that poor air quality lowers scores by 0.08 and 0.04 points in the lowest and the second-lowest quartile respectively. For the two quartiles above the median, the treatment effect is -0.03 and -0.02 respectively, neither of which is statistically significant. This suggests that student vulnerability is rising sharply with respect to student quality and may reflect the correlation between the incidence of asthma and socio-economic status. In contrast, the relationship between CO and test scores among the stratified sample is more mixed and the monotonic relationship is not evident for the continuous measure. Again, for the dichotomous measure of CO, the result is monotonic, leaving the results mixed regarding distinguishing between $PM_{2.5}$ and CO on this dimension. The results do, however, consistently point to large effects of both pollutants on student outcomes.

In Table A1.5, we exploit the unique ethnic heterogeneity of Israel to estimate models for sub-populations. Israel's population is composed primarily of Jews and Arabs, and the Jewish population is composed of immigrants from ethnically distinct source countries. The primary distinction is between *Sephardic* Jews of Middle Eastern and North African origin, and *Ashkenazic* Jews who are from Eastern Europe and Russia. The former group has lower rates of asthma and respiratory conditions (Laor et al. 1993). We find that the impact of air pollution is larger among *Ashkenazic* Jews relative to *Sephardic* Jews using both our measures of $PM_{2.5}$ and CO. For example, *Ashkenazic* Jews are a third more responsive to $PM_{2.5}$ (.046/.035) and almost twice as responsive using our dichotomous measure of $PM_{2.5}$ (1.73/1.01). For CO, however, the results are similar across groups, with *Ashkenazic* Jews being slightly *less* responsive than *Sephardic* Jews for both our continuous (.056/.61) and our dichotomous measure (8.28/10.56).

1.4.4 Impact of Particulate Matter on Academic Outcomes with Long-run Implications

While our analysis focuses on the impact of short-term exposure to particulate matter on cognition, in our context this can have a large effect on academic success in the long-run. Success on the *Bagrut* exam facilitates entry in to university, and higher scores allow a student to choose more lucrative college majors, such as medicine or computer science. To assess directly the potential harmful long-term effect of pollution on human capital formation in our context, we examine in Tables 1.7, 1.8, and 1.9 the relationship between exposure to air pollution and academic outcomes related to *Bagrut* exams.

In Table 1.7 Panel A, we examine the relationship between air pollution exposure and the probability of failing a particular *Bagrut* exam. In Panels B and C, we carry out the analysis at the student level. For these results, our new measure of pollution is the average pollution reading across *all* exams the students has taken. Our continuous measure of pollution is the average over all the exam days, and our threshold measure is the average over all days of whether the exam was administered on a day with pollution in the top 5% of most polluted days. As such, the coefficients will represent the impact of raising pollution on all days for the continuous measure, or increasing the fraction of exams taken during very polluted days from 0% to 100% for the threshold measure. As we will show, the results indicate that having poor $PM_{2.5}$ or CO on the days of the *Bagrut* exams is associated with a lower *Bagrut* composite score and lower probability of receiving the matriculation certificate. These outcomes can have a permanent impact on an individual's probability of attending college, and the majors that are available upon matriculation.

As shown in Panel A, in our preferred specification with student fixed effects, having elevated levels of $PM_{2.5}$ or CO using the continuous measure have a statistically insignificant effect. However, for the threshold measure, both indicate a large decline in a student's probability of passing the exam on very polluted days: a student is 2.4 and 12.3 percentage points less likely to pass an exam on very polluted days relative to a normal day. In Panel B, the estimated effect of $PM_{2.5}$ is negative and significant, and in our preferred specification, which includes school fixed effects, we estimate that an additional 10 units of AQI on average for each test would lead to a decline in the student's average score of 1.66 points, roughly 9.8% of a standard deviation. Similarly, increasing the fraction of days with high $PM_{2.5}$ readings by 10% reduces the average score by .96 points. A student's probability of passing the *Bagrut* is also sensitive to these measures. A 10 point increase in $PM_{2.5}$ AQI reduces a student's probability of receiving the *Bagrut* certificate by 3.3 percentage points, and increasing the fraction of days with very pollution readings by 10% reduces certificate achievement by 1.5 percentage points. Our estimates for CO are somewhat more modest: a 10 unit increase in the AQI average reading during the student's tests reduces scores by .86 points, and a 10% increase in the share of days with high pollution readings reduces scores by .75 points. Similarly, a 10 point increase in CO AQI reduces a student's probability of receiving the *Bagrut* certificate by 0.5 percentage points, and the result is not statistically significant. Finally, increasing the fraction of days with very pollution readings by 10% reduces certificate achievement by 1.4 percentage points. This suggests that CO only affects long-run outcomes among students who are exposed to extremely elevated levels of CO, and that more modest levels may have an extremely small impact.

In Table 1.8, we examine these results broken down by two sub-populations that may be more sensitive to air pollution: boys and students of lower quality. The results indicate that boys are more sensitive to $PM_{2.5}$ than girls, and lower quality students are more likely to be detrimentally affected than stronger students. In particular, raising the fraction of days with very polluted air by 10 percentage points is associated with a .57 percentage point increase for boys in the chance of failing a particular *Bagrut* in models with student fixed effects. Girls appear largely unaffected, with the increased chance of not passing being statistically indistinguishable from zero. The gap is even more striking for student with low *Magen* scores: a 10 percentage point increase in the fraction of days with very polluted air is associated with a .59 percentage point increase in failure probability. The second outcome we examine is the student's probability of failing the composite *Bagrut*. Boys are nearly a third more sensitive to air pollution by this measure, where a 10 percentage point increase in polluted days is associated with a 1.74 percentage point increased chance of not receiving their matriculation certificate, whereas girls only experience a 1.19 percentage point increase. The results are even more striking for low scoring *Magen* students, who are 1.07 percentage points more likely to not receive a *Bagrut* certificate for a 10 percentage point increase in the share of days with poor air quality.

In Table 1.9, we present results parallel to those shown in Table 1.8 but for CO rather than $PM_{2.5}$. While the results for our continuous measure are statistically insignificant, the results for our threshold measure are negative and statistically significant. Interestingly, we find very similar results for boys and girls in their probability of failing the *Bagrut* exam or not receiving a matriculation certificate. For instance, a 10 percentage point increase in days above the CO threshold is associated with a 1.42 percentage point increased chance of not receiving their matriculation certificate for boys, and girls experience a similar 1.44 percentage point increase. The results are also similar for low scoring *Magen* students, who are 1.02 percentage points more likely to not receive a matriculation certificate for a 10 percentage point increase in the share of days with poor air quality, versus a 1.24 increase for high scoring *Magen* students. This suggests that the long-run effects of CO are similar across different groups.

1.5 Conclusion

This paper has examined the relationship between cognitive performance and ambient pollution exposure. Using a large sample of Israeli high-school *Bagrut* examinations (2000-2002), we have presented evidence that there is a robust negative relationship between outcomes and ambient pollution concentrations. We also find that among Israeli sub-populations with higher rates of asthma and respiratory illnesses, our estimated treatment effects for $PM_{2.5}$ are larger, suggesting that physiological impairment is a potential mechanism for our findings. In contrast, our results for CO are largely consistent among Israeli sub-populations, suggesting that neurological impairment may be a mechanism for our findings. The measured impact of our pollutants may have a permanent effect on a student's human capital formation, because it affects whether the student earns the *Bagrut* in a timely fashion and can matriculate in college following the army, or must complete additional coursework prior to starting college, delaying matriculation. In the overall economy, the mis-ranking of students due to variability in pollution exposure may result in bad assignment of workers to different occupations, resulting in reduced labor productivity.

While our results are robust to a variety of specification checks, it is worth noting several important caveats. First, our result is in a completely reduced form and we cannot trace out the pathways. While we posit that asthmatics and other sensitive groups are driving our results for $PM_{2.5}$, this is difficult to determine definitively in the absence of health measures for the test takers. Second, we cannot fully examine whether the effect is due to pollution only on the day of the exam, versus through a build-up effect from the days prior to the exam. We report the relationship between the exam outcome and ambient pollution, but we are unable with our data to fully disentangle the exact timing of the effect. Third, it may be that increased pollution is contemporaneous with other factors affecting test outcomes. For example, it is possible that traffic on the way to the exam is correlated with pollution and with reduced test performance. In spite of these limitations, our results present new evidence of a connection between reduced cognitive performance and fine particulate matter or carbon monoxide exposure.

The results presented here suggest that the gain from improving air quality may be underestimated by a narrow focus on health impacts. Insofar as air pollution may lead to reduced cognitive performance, the consequences of pollution may be relevant for a variety of everyday activities that require mental acuity. Traffic accidents, injuries in the workplace, and reduced worker productivity may all be the byproduct of reduced cognitive performance. As such, the results presented here highlight a channel by which the consequences of pollution are vastly understated by a narrow focus on the immediate and acute health consequences, and suggest that improvements in air quality may yield tremendous benefits in welfare.

Table 1.1

Descriptive Statistics

Variable	All (1)	By Sex: Boys (2)	By Sex: Girls (3)	By *Magen* Score (Course Grade[1]): Low Scores (4)	By *Magen* Score (Course Grade[1]): High Scores (5)
$PM_{2.5}$	21.05	20.89	21.18	21.15	20.96
(μg/m^3)	(10.86)	(10.57)	(11.10)	(10.88)	(10.87)
$PM_{2.5}$	59.74	59.47	59.98	60.01	59.51
(AQI Index)	(22.81)	(22.50)	(23.08)	(22.89)	(22.75)
$PM_{2.5}$	0.05	0.05	0.05	0.05	0.05
(AQI ≥101)	(0.21)	(0.21)	(0.22)	(0.22)	(0.21)
CO	1.21	1.22	1.21	1.25	1.17
(μg/m^3)	(1.05)	(1.08)	(1.02)	(1.15)	(0.92)
CO	13.77	13.81	13.73	14.19	13.29
(AQI Index)	(11.58)	(11.93)	(11.27)	(12.80)	(10.18)
CO	0.04	0.04	0.04	0.04	0.03
(>95th percentile)	(0.20)	(0.20)	(0.20)	(0.21)	(0.18)
Bagrut Exam Score	70.76	68.91	72.33	53.22	77.10
(1-100 points)	(23.74)	(24.86)	(22.64)	(30.69)	(22.18)
Magen Score	75.45	73.27	77.30	64.09	86.93
(1-100 points)	(21.37)	(22.50)	(20.19)	(23.25)	(10.47)
Bagrut Composite	83.03	81.37	84.49	73.18	95.05
Score	(16.84)	(17.48)	(16.11)	(14.59)	(10.33)
Matriculation	0.68	0.64	0.71	0.48	0.91
Certificate (1=yes)	(0.47)	(0.48)	(0.45)	(0.50)	(0.28)
Failed a *Bagrut* Exam	0.19	0.21	0.17	0.33	0.04
(1=yes)	(0.39)	(0.41)	(0.37)	(0.47)	(0.19)
Mother's Education	11.44	11.60	11.30	10.79	12.08
(years)	(5.04)	(5.09)	(5.00)	(4.87)	(5.13)
Father's Education	11.62	11.83	11.44	10.85	12.39
(years)	(5.03)	(5.02)	(5.03)	(4.84)	(5.10)
Temperature	23.81	23.81	23.82	23.84	23.83
(celsius)	(2.61)	(2.61)	(2.62)	(2.66)	(2.50)
Relative Humidity	50.90	50.86	50.94	50.98	50.95
(percent saturation)	(14.71)	(14.52)	(14.87)	(15.08)	(14.35)
Observations	415,219	190,410	224,809	206,571	204,527

Notes: Standard deviations are in parentheses. The measures of pollution are particulate matter smaller than 2.5 microns, or $PM_{2.5}$, and carbon monoxide, CO. We also report the AQI value for each reading, which is calculated from a formula that converts micrograms (μg/m3) into a 1-500 index value. We also report dummies for days with $PM_{2.5}$ (AQI) >100 or CO readings in the top 5% of days in our sample. Relative humidity is the amount of moisture in the air as a share of what the air can hold at that temperature. Receiving a Matriculation Certificate is determined by a combination of the average Bagrut score across exams, and the Magen score, which is composed of the student's course grade and an exam similar in content to the Bagrut. [1]The low and high subsamples were based on being above or below the median of the Magen score.

Table 1.2

Pooled OLS and Fixed Effect Models of Air Pollution's Impact on *Bagrut* Test Scores

	RHS Pollutant Measure: Particulate $Matter_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No Controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No Controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Panel A: Air Quality Index (continuous measure)										
Pollutant	-0.055	-0.065	-0.082	-0.069	-0.046	-0.047	-0.054	-0.133	-0.083	-0.085
	(0.015)	(0.011)	(0.008)	(0.007)	(0.007)	(0.017)	(0.020)	(0.018)	(0.017)	(0.017)
Female (1=yes)		3.22	3.30	2.72			3.82	3.88	3.15	
		(0.34)	(0.34)	(0.22)			(0.50)	(0.50)	(0.38)	
Mother's Education		0.165	0.141	0.112			0.182	0.191	0.113	
		(0.063)	(0.062)	(0.034)			(0.097)	(0.093)	(0.057)	
Father's Education		0.410	0.396	0.241			0.451	0.463	0.251	
		(0.061)	(0.058)	(0.033)			(0.095)	(0.090)	(0.050)	
R-squared	0.003	0.042	0.046	0.145	0.493	0.001	0.054	0.060	0.174	0.531
Observations	415,219	380,435	380,435	380,435	380,435	158,647	153,528	153,528	153,528	153,528
Panel B: Air Quality Index above Threshold Value										
Dummy for AQI>100[1]	-3.00	-2.63	-2.75	-2.68	-1.95	-6.04	-6.68	-9.16	-9.56	-10.16
	(1.54)	(1.03)	(0.84)	(0.70)	(0.74)	(1.15)	(1.31)	(1.28)	(0.96)	(1.02)
Female (1=yes)		3.19	3.25	2.68			3.84	3.91	3.19	
		(0.340)	(0.337)	(0.219)			(0.498)	(0.496)	(0.377)	
Mother's Education		0.158	0.143	0.111			0.185	0.192	0.117	
		(0.064)	(0.063)	(0.035)			(0.096)	(0.092)	(0.055)	
Father's Education		0.409	0.396	0.241			0.452	0.465	0.252	
		(0.061)	(0.058)	(0.033)			(0.094)	(0.090)	(0.048)	
R-squared	0.001	0.040	0.043	0.143	0.492	0.002	0.056	0.062	0.177	0.534
Observations	415,219	380,435	380,435	380,435	380,435	158,647	153,528	153,528	153,528	153,528

Notes : Standard errors are clustered by school. All regressions include suppressed controls for temperature and humidity on the exam date, which are included as linear and quadratic terms in each, and linear and quadratic interaction terms of the two variables. [1]For carbon monoxide, we generate a dummy for a test occurring on a day in the top 5% of most polluted days.

Table 1.3

Air Pollution's Impact on *Bagrut* Test Scores on Polluted and Extremely Polluted Days

	RHS Pollutant Measure: Particulate Matter$_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No Controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No Controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Dummy for AQI >50 & < 101[1]	-2.32	-2.29	-3.02	-2.41	-1.43	0.56	-1.92	-1.26	-1.36	-0.72
	(0.50)	(0.42)	(0.31)	(0.29)	(0.33)	(0.71)	(0.75)	(0.67)	(0.60)	(0.66)
Dummy for AQI ≥ 101	-4.42	-4.07	-4.92	-4.34	-2.89	-5.76	-8.56	-10.39	-10.88	-10.87
	(1.61)	(1.10)	(0.87)	(0.73)	(0.78)	(1.27)	(1.55)	(1.42)	(1.14)	(1.19)
Female (1=yes)		3.20	3.27	2.70			3.86	3.92	3.20	
		(0.339)	(0.335)	(0.217)			(0.498)	(0.497)	(0.378)	
Mother's Education		0.166	0.142	0.112			0.180	0.190	0.114	
		(0.064)	(0.063)	(0.035)			(0.096)	(0.092)	(0.055)	
Father's Education		0.411	0.395	0.241			0.455	0.466	0.252	
		(0.061)	(0.058)	(0.034)			(0.095)	(0.090)	(0.049)	
R-squared	0.003	0.041	0.046	0.145	0.493	0.003	0.056	0.063	0.178	0.534
Observations	415,219	380,435	380,435	380,435	380,435	158,647	153,528	153,528	153,528	153,528

Notes : See Table 2. [1]For carbon monoxide we generate a dummy for a test occurring on a day above the median pollution level and below the top 5% of the most polluted days as the intermediate pollution category.

Table 1.4

Placebo Tests Measuring the Relationship between the *Bagrut* and Pollutants on Irrelevant Days

	RHS Pollutant Measure: Particulate Matter$_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No Controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No Controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Panel A: Pollutant Level from Previous Exam										
Pollutant (AQI)	-0.024	-0.035	-0.049	-0.034	-0.005	-0.128	-0.130	-0.307	-0.080	0.097
	(0.013)	(0.010)	(0.007)	(0.006)	(0.006)	(0.065)	(0.074)	(0.164)	(0.079)	(0.055)
Pollutant (Threshold)	-1.10	-0.48	-0.66	-0.29	0.78	-3.61	-1.48	-5.68	1.48	-2.53
	(1.37)	(0.87)	(0.78)	(0.68)	(0.71)	(2.29)	(2.85)	(3.09)	(2.82)	(3.06)
Observations	358,584	328,974	328,974	328,974	328,974	131,579	127,341	127,341	127,341	127,341
Panel B: Pollutant Level from Previous Year										
Pollutant (AQI)	-0.008	-0.033	-0.027	-0.014	-0.006	-0.032	-0.060	0.061	0.063	0.147
	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)	(0.017)	(0.038)	(0.029)	(0.023)	(0.048)
Pollutant (Threshold)	-2.78	-2.89	-1.03	-0.75	-1.15	0.98	2.38	3.87	4.90	5.55
	(0.81)	(0.68)	(0.76)	(0.73)	(0.69)	(1.06)	(0.76)	(0.74)	(0.70)	(0.81)
Observations	261,091	248,759	248,759	248,759	248,759	291,555	193,764	193,764	193,764	193,764

Notes : See Table 2.

Table 1.5

Air Pollution's Impact on *Bagrut* Test Scores, Separately for Boys and Girls

	RHS Pollutant Measure: Particulate Matter$_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No Controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No Controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Panel A: Boys Only										
Pollutant (AQI)	-0.087	-0.100	-0.118	-0.104	-0.078	-0.035	-0.055	-0.142	-0.080	-0.099
	(0.018)	(0.013)	(0.009)	(0.009)	(0.009)	(0.021)	(0.026)	(0.024)	(0.020)	(0.019)
Pollutant (Threshold)	-4.83	-5.62	-5.59	-5.33	-4.10	-6.12	-7.49	-10.73	-10.28	-11.28
	(1.95)	(1.26)	(0.96)	(0.82)	(0.87)	(1.41)	(1.65)	(1.53)	(1.22)	(1.23)
Observations	190,410	174,250	174,250	174,250	174,250	73,054	70,311	70,311	70,311	70,311
Panel B: Girls Only										
Pollutant (AQI)	-0.031	-0.036	-0.054	-0.041	-0.021	-0.058	-0.052	-0.125	-0.091	-0.075
	(0.014)	(0.012)	(0.009)	(0.007)	(0.008)	(0.017)	(0.017)	(0.021)	(0.018)	(0.023)
Pollutant (Threshold)	-1.67	-0.30	-0.55	-0.66	-0.38	-6.07	-5.79	-7.62	-8.79	-9.29
	(1.35)	(1.03)	(0.90)	(0.80)	(0.83)	(1.13)	(1.26)	(1.34)	(1.10)	(1.16)
Observations	224,809	206,185	206,185	206,185	206,185	85,593	83,217	83,217	83,217	83,217

Notes : See Table 2.

Table 1.6

Air Pollution's Impact on Test Scores, Separately for Students with Low and High *Magen* Scores

	RHS Pollutant Measure: Particulate Matter$_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No Controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No Controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Panel A: Low *Magen* Scores										
Pollutant (AQI)	-0.074	-0.078	-0.081	-0.075	-0.061	-0.073	-0.097	-0.064	-0.065	-0.048
	(0.015)	(0.013)	(0.011)	(0.010)	(0.011)	(0.032)	(0.031)	(0.021)	(0.020)	(0.029)
Pollutant (Threshold)	-4.64	-4.79	-3.77	-3.86	-3.49	-6.22	-9.66	-11.79	-11.56	-12.14
	(1.58)	(1.24)	(1.12)	(1.04)	(1.10)	(1.31)	(1.49)	(1.64)	(1.28)	(1.45)
Observations	206,571	185,030	185,030	185,030	185,030	134,126	128,078	128,078	128,078	128,078
Panel B: High *Magen* Scores										
Pollutant (AQI)	-0.027	-0.024	-0.037	-0.030	-0.028	-0.023	-0.027	-0.068	-0.052	-0.055
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.015)	(0.015)	(0.016)	(0.016)	(0.015)
Pollutant (Threshold)	-0.94	-0.93	-1.30	-0.93	-0.76	-2.97	-4.09	-4.61	-4.88	-4.57
	(0.42)	(0.71)	(0.66)	(0.57)	(0.68)	(0.59)	(0.69)	(0.78)	(0.81)	(0.85)
Observations	204,527	191,790	191,790	191,790	191,790	128,758	126,284	126,284	126,284	126,284

Notes : See Table 2. The sample is stratified by whether the student did below (Panel A) or above (Panel B) the median on the *Magen* score. The *Magen* score is based on the student's class performance and on an exam similar to the Bagrut.

Table 1.7

Air Pollution's Impact on Long-term Academic Outcomes Related to the *Bagrut* Examination

	RHS Pollutant Measure: Particulate $Matter_{2.5}$				RHS Pollutant Measure: Carbon Monoxide			
	Pooled OLS	Fixed Effects			Pooled OLS	Fixed Effects		
	Controls (1)	City (2)	School (3)	Student (4)	Controls (5)	City (6)	School (7)	Student (8)
Panel A: Failing a Particular *Bagrut* Exam (1=yes)								
Pollutant (AQI, 100 units)	0.025	0.035	0.018	-0.017	0.048	0.071	0.026	0.019
	(0.015)	(0.010)	(0.009)	(0.010)	(0.024)	(0.034)	(0.036)	(0.049)
Pollutant (Threshold)	0.036	0.039	0.036	0.024	0.082	0.103	0.111	0.123
	(0.016)	(0.012)	(0.010)	(0.010)	(0.018)	(0.021)	(0.017)	(0.019)
Observations	380,435	380,435	380,435	380,435	153,528	153,528	153,528	153,528
Panel B: *Bagrut* Exam Composite Score								
Pollutant (AQI, 100 units)	-6.77	-26.79	-16.60		-2.88	-22.72	-8.56	
	(4.29)	(3.40)	(1.87)		(3.68)	(4.53)	(3.23)	
Pollutant (Threshold)	-3.77	-10.93	-9.55		-10.98	-8.43	-7.54	
	(4.96)	(3.66)	(2.70)		(2.24)	(1.55)	(1.15)	
Observations	50,899	50,899	50,899		25,730	25,730	25,730	
Panel C: Received a *Bagrut* Matriculation Certificate (1=yes)								
Pollutant (AQI, 100 units)	-0.236	-0.537	-0.328		-0.063	-0.301	-0.054	
	(0.099)	(0.082)	(0.048)		(0.075)	(0.102)	(0.100)	
Pollutant (Threshold)	-0.184	-0.255	-0.146		-0.214	-0.188	-0.142	
	(0.125)	(0.093)	(0.050)		(0.051)	(0.035)	(0.027)	
Observations	50,899	50,899	50,899		25,730	25,730	25,730	

Notes : See Table 2. In Panel A, each observation is an examination. In Panel B and Panel C, each observation is a student. For the models estimated in Panel B and Panel C, pollution is averaged over all of the *Bagrut* tests taken following grades 10-12 for each student.

Table 1.8

Particulate Matter's Impact on Failing a *Bagrut* Exam and
Receiving a Matriculation Certificate by Sex and *Magen* Score

	LHS: Failed *Bagrut* Exam (1=yes)				LHS: Received Matriculation Certificate (1=yes)		
	Pooled OLS	Fixed Effects			Pooled OLS	Fixed Effects	
	Controls (1)	City (2)	School (3)	Student (4)	Controls (5)	City (6)	School (7)
Panel A: Boys Only							
$PM_{2.5}$ (AQI, 100 units)	0.070	0.082	0.064	0.021	-0.243	-0.621	-0.345
	(0.020)	(0.014)	(0.013)	(0.014)	(0.109)	(0.089)	(0.057)
$PM_{2.5}$ (Threshold)	0.085	0.086	0.077	0.057	-0.283	-0.403	-0.174
	(0.022)	(0.017)	(0.015)	(0.015)	(0.133)	(0.094)	(0.065)
Observations	174,250	174,250	174,250	174,250	23,830	23,830	23,830
Panel B: Girls Only							
$PM_{2.5}$ (AQI, 100 units)	-0.012	-0.002	-0.017	-0.046	-0.231	-0.468	-0.328
	(0.015)	(0.011)	(0.010)	(0.011)	(0.107)	(0.097)	(0.056)
$PM_{2.5}$ (Threshold)	0.000	0.003	0.004	-0.001	-0.061	-0.073	-0.119
	(0.014)	(0.012)	(0.010)	(0.010)	(0.142)	(0.116)	(0.062)
Observations	206,185	206,185	206,185	206,185	27,069	27,069	27,069
Panel C: Low *Magen* Scores							
$PM_{2.5}$ (AQI, 100 units)	0.038	0.038	0.026	0.001	-0.086	-0.299	-0.204
	(0.020)	(0.017)	(0.015)	(0.017)	(0.096)	(0.079)	(0.049)
$PM_{2.5}$ (Threshold)	0.077	0.071	0.066	0.059	-0.124	-0.178	-0.107
	(0.020)	(0.017)	(0.016)	(0.016)	(0.103)	(0.082)	(0.049)
Observations	185,030	185,030	185,030	185,030	24,892	24,892	24,892
Panel D: High *Magen* Scores							
$PM_{2.5}$ (AQI, 100 units)	-0.026	-0.031	-0.035	-0.041	-0.219	-0.197	-0.118
	(0.006)	(0.006)	(0.005)	(0.007)	(0.074)	(0.077)	(0.050)
$PM_{2.5}$ (Threshold)	0.004	0.006	0.000	-0.002	-0.284	-0.252	-0.042
	(0.007)	(0.006)	(0.006)	(0.007)	(0.146)	(0.131)	(0.044)
Observations	191,790	191,790	191,790	191,790	26,007	26,007	26,007

Notes: See Table 2. Each cell in the table represents a separate regression. Each observation in columns (1)-(4) is an examination, and in columns (5)-(7) each observation is a student.

Table 1.9

Carbon Monoxide's Impact on Failing a *Bagrut* Exam and Receiving a Matriculation Certificate by Sex and *Magen* Score

	LHS: Failed *Bagrut* Exam (1=yes)				LHS: Received Matriculation Certificate (1=yes)		
	Pooled OLS	Fixed Effects			Pooled OLS	Fixed Effects	
	Controls (1)	City (2)	School (3)	Student (4)	Controls (5)	City (6)	School (7)
Panel A: Boys Only							
CO (AQI, 100 units)	0.049	0.089	0.027	0.037	-0.022	-0.240	0.009
	(0.033)	(0.030)	(0.029)	(0.036)	(0.079)	(0.164)	(0.128)
CO (Threshold)	0.090	0.123	0.118	0.132	-0.236	-0.196	-0.142
	(0.023)	(0.025)	(0.020)	(0.023)	(0.061)	(0.049)	(0.040)
Observations	70,311	70,311	70,311	70,311	11,990	11,990	11,990
Panel B: Girls Only							
CO (AQI, 100 units)	0.045	0.057	0.031	0.007	-0.101	-0.389	-0.243
	(0.024)	(0.048)	(0.050)	(0.067)	(0.083)	(0.111)	(0.125)
CO (Threshold)	0.071	0.084	0.103	0.115	-0.191	-0.176	-0.144
	(0.018)	(0.022)	(0.018)	(0.020)	(0.049)	(0.036)	(0.028)
Observations	83,217	83,217	83,217	83,217	13,740	13,740	13,740
Panel C: Low *Magen* Scores							
CO (AQI, 100 units)	0.031	0.073	0.029	0.053	-0.107	-0.180	0.033
	(0.034)	(0.060)	(0.061)	(0.073)	(0.076)	(0.138)	(0.108)
CO (Threshold)	0.136	0.197	0.192	0.220	-0.183	-0.163	-0.102
	(0.025)	(0.027)	(0.023)	(0.029)	(0.053)	(0.045)	(0.035)
Observations	71,192	71,192	71,192	71,192	11,962	11,962	11,962
Panel D: High *Magen* Scores							
CO (AQI, 100 units)	0.028	-0.01	-0.014	-0.015	-0.053	-0.143	-0.052
	(0.010)	(0.016)	(0.017)	(0.025)	(0.043)	(0.093)	(0.087)
CO (Threshold)	0.026	0.019	0.023	0.026	-0.097	-0.11	-0.124
	(0.009)	(0.010)	(0.010)	(0.012)	(0.035)	(0.035)	(0.024)
Observations	80,728	80,728	80,728	80,728	13,768	13,768	13,768

Notes: See Table 2. Each cell in the table represents a separate regression. Each observation in columns (1)-(4) is an examination, and in columns (5)-(7) each observation is a student.

Figure 1.1
Locations of Air Quality Monitoring Stations in Israel

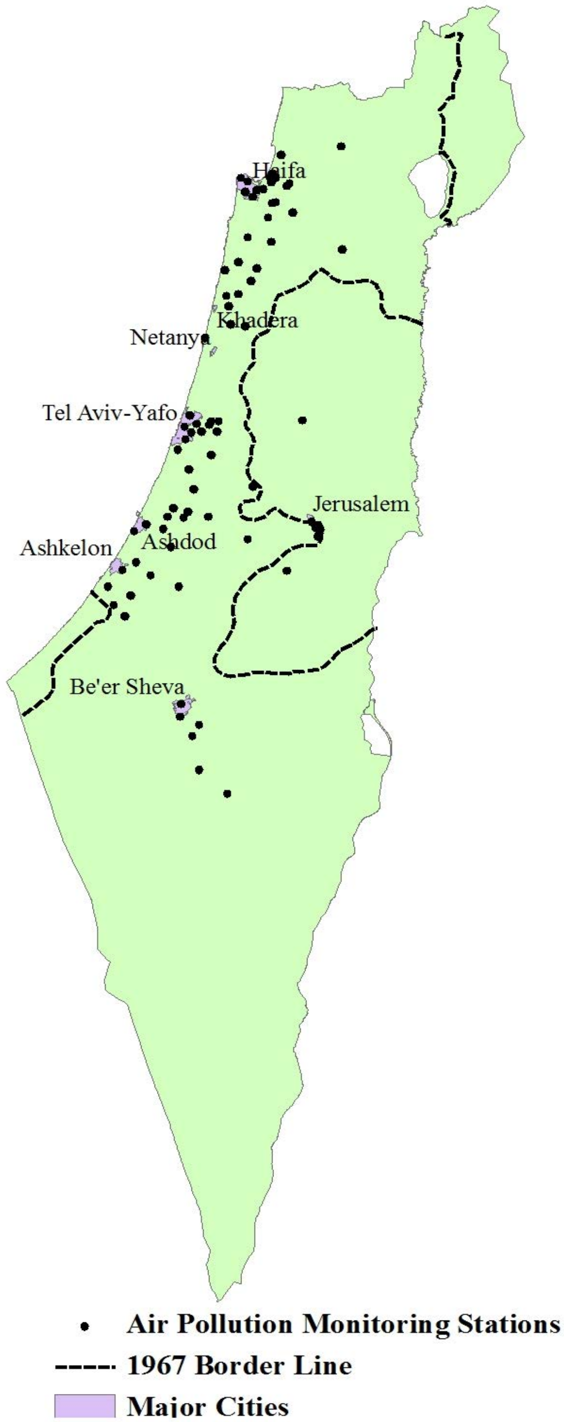
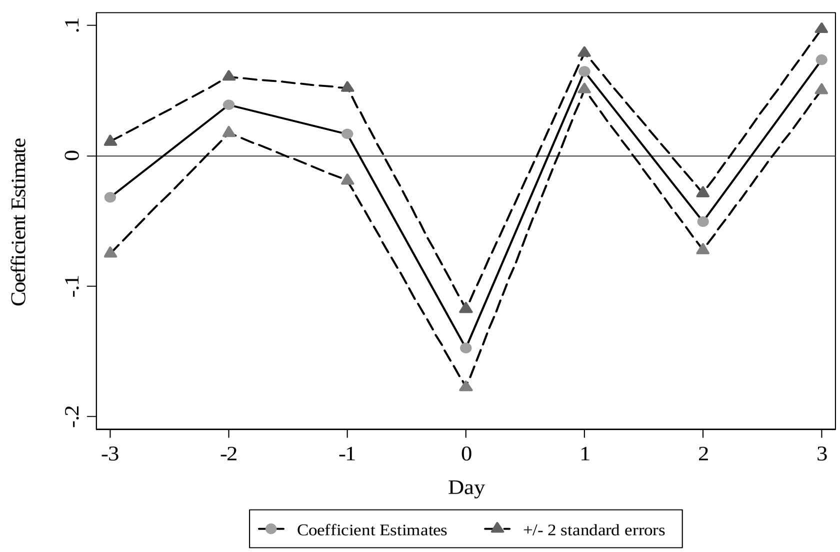
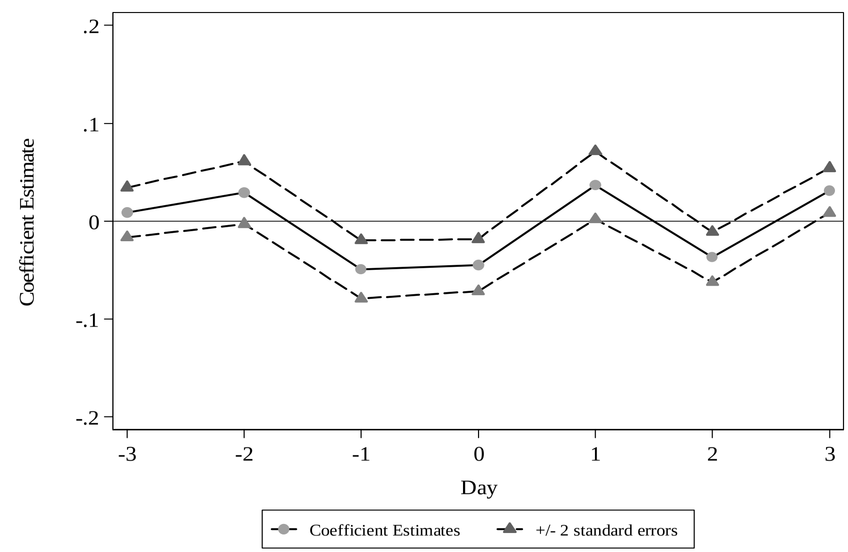


Figure 1.2
Impact of $PM^{2.5}$ and CO on Test Scores in the Days
Pre and Post Examination

Particulate Matter$_{2.5}$



Carbon Monoxide



Notes : The figure plots the coefficients from a regression of *Bagrut* test scores on $PM_{2.5}$ and CO AQI readings in the days prior to, the day of (Day=0), and the days following the examination. Standard errors are clustered by school.

Table A1.1

Breakpoints for $PM_{2.5}$ ($\mu g/m^3$) and AQI Index Categories

$PM_{2.5}$ (μg/m3)	AQI Index Value	Category
0.0 - 15.4	0 - 50	Good
15.5 -40.4	51 - 100	Moderate
40.5 - 65.4	101 - 150	Unhealthy for Sensitive Groups
65.5 - 150.4	151 - 200	Unhealthy
150.5 - 250.4	201 - 300	Very unhealthy
250.5 - 350.4	301 - 400	Hazardous
350.5 - 500.4	401 - 500	Hazardous

Source : United States Environmental Protection Agency

Table A1.2

Pooled OLS and Fixed Effect Models of Pollutant Matter on Test Scores with Both Measures Included Simultaneously

	RHS Pollutant Measure: Particulate Matter$_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Pollutant (AQI)	-0.049	-0.030	-0.043	-0.033	-0.009	-0.046	-0.052	-0.126	-0.078	-0.083
	(0.018)	(0.018)	(0.016)	(0.014)	(0.019)	(0.017)	(0.020)	(0.018)	(0.017)	(0.017)
Pollutant (Threshold)	-3.77	-1.93	-2.09	-2.60	-1.73	-6.19	-6.89	-9.33	-9.77	-10.26
	(1.03)	(1.01)	(0.94)	(0.74)	(0.82)	(1.15)	(1.30)	(1.28)	(0.96)	(1.01)
Observations	158,647	153,528	153,528	153,528	153,528	158,647	153,528	153,528	153,528	153,528

Notes : The table reports the coefficients from estimating the models with *both* measures of pollution as independent variables. The results in the first row and columns (1) and (6) are from the same regression, and the results from (2) and (7) are from the same regression, and so on. The results in the second row and columns (1) and (6) are from the same regression, and the results from (2) and (7) are from the same regression, and so on.

Table A1.3

Pooled OLS and Fixed Effect Models of Pollutant Matter on Afternoon and Morning Test Scores

	RHS Pollutant Measure: Particulate Matter$_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No Controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No Controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Panel A: Afternoon Examinations										
Pollutant (AQI)	-0.019	-0.079	-0.116	-0.082	-0.045	-0.099	-0.079	-0.183	-0.152	-0.135
	(0.017)	(0.015)	(0.013)	(0.010)	(0.013)	(0.024)	(0.014)	(0.022)	(0.015)	(0.020)
Pollutant (Threshold)	-2.94	-3.11	-3.04	-2.56	-3.16	-8.82	-7.77	-9.24	-9.89	-10.45
	(2.11)	(2.01)	(1.95)	(1.35)	(1.42)	(1.23)	(1.22)	(1.29)	(1.10)	(1.31)
Observations	162,912	148,026	148,026	148,026	148,026	68,161	65,984	65,984	65,984	65,984
Panel B: Morning Examinations										
Pollutant (AQI)	-0.074	-0.067	-0.074	-0.066	-0.054	0.017	0.007	0.086	0.130	0.239
	(0.016)	(0.013)	(0.009)	(0.008)	(0.010)	(0.026)	(0.046)	(0.069)	(0.034)	(0.083)
Pollutant (Threshold)	-2.93	-3.13	-2.97	-3.13	-2.46	-1.50	-2.45	-7.50	-0.08	-1.12
	(1.38)	(1.11)	(0.87)	(0.73)	(0.97)	(2.79)	(3.35)	(8.30)	(3.24)	(4.38)
Observations	252,307	232,409	232,409	232,409	232,409	90,486	87,544	87,544	87,544	87,544

Notes : See Table 2. The examinations that are given at 12PM and later are classified as afternoon exams.

Table A1.4

Air Pollution's Impact on Test Scores, Separately by *Magen* Score Quartile

	RHS Pollutant Measure: Particulate $Matter_{2.5}$				RHS Pollutant Measure: Carbon Monoxide			
	Low *Magen* Score Percentile		High *Magen* Score Percentile		Low *Magen* Score Percentile		High *Magen* Score Percentile	
	(0-0.25)	(0.25-0.50)	(0.50-0.75)	(0.75-1.00)	(0-0.25)	(0.25-0.50)	(0.50-0.75)	(0.75-1.00)
Pollutant (AQI)	-0.080	-0.044	-0.033	-0.022	-0.134	-0.064	-0.082	-0.088
	(0.017)	(0.010)	(0.008)	(0.006)	(0.042)	(0.020)	(0.015)	(0.028)
Pollutant (Threshold)	-4.99	-2.29	-1.24	-0.35	-18.22	-10.93	-7.52	-3.56
	(1.38)	(1.19)	(0.86)	(0.63)	(1.86)	(1.16)	(0.87)	(1.18)
Observations	90,354	94,676	94,288	97,502	36,901	38,446	38,022	38,551

Notes : See Table 2. All models include student fixed effects. The columns include the students within the listed percentile range on the *Magen* score, which is based on the student's class performance and on an exam similar to the *Bagrut* .

Table A1.5

Pooled OLS and Fixed Effect Models of Pollutant Matter on Test Scores, Separately for *Ashkenazi* and *Sephardi* Students

	RHS Pollutant Measure: Particulate Matter$_{2.5}$					RHS Pollutant Measure: Carbon Monoxide				
	Pooled OLS		Fixed Effects			Pooled OLS		Fixed Effects		
	No Controls (1)	Controls (2)	City (3)	School (4)	Student (5)	No Controls (6)	Controls (7)	City (8)	School (9)	Student (10)
Panel A: *Ashkenazi* (Europe, America & Australia)										
Pollutant (AQI)	-0.048	-0.056	-0.081	-0.062	-0.046	-0.038	-0.035	-0.122	-0.077	-0.056
	(0.018)	(0.014)	(0.011)	(0.009)	(0.012)	(0.024)	(0.028)	(0.035)	(0.033)	(0.025)
Pollutant (Threshold)	-2.91	-2.85	-3.03	-2.19	-1.73	-5.87	-5.98	-7.41	-9.10	-8.28
	(1.94)	(1.35)	(1.08)	(1.07)	(1.13)	(1.30)	(1.30)	(1.29)	(1.19)	(1.19)
Observations	88,635	80,156	80,156	80,156	80,156	31,437	30,156	30,156	30,156	30,156
Panel B: *Sephardi* (Asia, Middle East & Africa)										
Pollutant (AQI)	-0.077	-0.058	-0.068	-0.058	-0.035	-0.037	-0.059	-0.109	-0.073	-0.061
	(0.018)	(0.014)	(0.011)	(0.010)	(0.009)	(0.020)	(0.020)	(0.019)	(0.018)	(0.020)
Pollutant (Threshold)	-4.11	-1.17	-1.15	-1.32	-1.01	-4.96	-6.85	-9.06	-9.72	-10.56
	(1.67)	(1.33)	(1.20)	(0.98)	(1.00)	(1.29)	(1.49)	(1.61)	(1.50)	(1.56)
Observations	61,889	54,822	54,822	54,822	54,822	22,702	22,116	22,116	22,116	22,116

Notes : See Table 2.

Chapter 2

The Contemporaneous Effect of Indoor Air Pollution on Cognitive Performance: Evidence from the UK

2.1 Introduction

Recent decades have seen a dramatic increase in the level of public concern surrounding the adverse effect of ambient pollution. However, the importance of indoor air quality has been often overlooked. This is of particular interest given that the US population spends 89% of their time indoors of which 21% in non-residential environments, such as offices and schools (Klepeis et al., 2001; Wu et al., 2001). Studies have shown that indoor pollution can cause immediate health effects including irritation of the eyes, nose, and throat, headaches, dizziness, and fatigue (Young, 2001; Brenstein, 2008)[19]. Exposure to particulate matter can also affect cognitive acuity as any deterioration in oxygen quality may in theory impair brain functioning (Clark and Sokoloff, 1999). Nevertheless, evidence on the effect of indoor pollution on cognitive performance is remarkably scarce. A potential link between pollution and cognitive performance would imply that a narrow focus on traditional health outcomes, such as hospitalization and increased mortality, may understate the true cost of pollution as mental acuity is essential to productivity in most professions.

There are many challenges in identifying the link between air pollution and human health such as heterogeneity in avoidance behavior, measurement error and the presence of unobserved correlated factors. However, identifying the causal relationship between indoor air pollution and

[19] There is also strong evidence on the long-term effect of indoor air pollution on human health. These effects include respiratory disease, heart disease and even cancer. See http://www.who.int/mediacentre/factsheets/fs292/en/.

cognitive performance possess an additional challenge. Whilst the impact of air pollution on health outcomes is likely to be recorded, data on the adverse effect of air pollution on cognitive performance may be unobserved by researchers as impaired cognitive performance is unlikely to lead to health encounters and may not even be noticed by the affected individual (Chang et al., 2014). As such, this paper provides a unique opportunity to asses such potential link by using university final examinations in the UK as a measure of cognitive performance.

I perform my analysis using a unique data set which combines readings of indoor air pollution (PM_{10}) with administrative data on 2,458 students taking 11,522 exams at a leading public research university within the Greater London Urban Area. To account for potential confounders I crucially rely on the panel structure of the data to estimate models with subject, venue and student fixed effects. Importantly, students are exogenously allocated to exams venues by the university a few weeks prior to the examination date and avoidance behavior is therefore unlikely. By collecting air pollution data from within the examination sites I overcome the challenge of measurement error which could result from assigning pollution to individuals. This is of particular importance as most studies in the literature use data from ambient air pollution monitors which are usually located a few miles away from the location of the individual. As such, they are likely to be subject to considerable measurement error due to significant spatial variation even within finely defined areas (Moretti and Neidell, 2011; Lin et al., 2001). I also include controls for time-varying factors that could be contemporaneous and correlated with pollution, such as daily temperature and relative humidity. Nevertheless, it is still possible that other unobserved factors that are correlated with both pollution and test scores remain present. In order to ease such concern, I conduct a rich set of placebo and robustness tests. More specifically, I examine the correlation between test scores and indoor air pollution from the previous exam and also the correlation between ex-ante test scores and elevated levels of PM_{10}. The correlations in both placebos are not statistically different from zero, lending further supports to the causal interpretation of the analysis.

My results demonstrate that elevated levels of Particulate Matter (PM_{10}) have a statistically and economically significant effect on test scores. I find that a one unit increase in PM_{10} (µg/m3) or being above the World Health Organization (WHO) guideline reduces student's test scores by 0.060 and 2.868 points respectively. The effect for the dichotomous indicator is equivalent to 0.15 of a standard deviation which is very large and similar to the estimated effects found in studies that have measured the impact of paying teachers and students large financial incentives (Jackson, 2010) or reducing class size from 31 to 25 students (Angrist and Lavy, 1999). Furthermore, I explore whether indoor air pollution has heterogeneous effects across subpopulations and academic disciplines. My interest is twofold: first, to test whether some subgroups are more sensitive to indoor pollution; and second, to examine whether the effect of indoor pollution varies across subjects. I find that the effect is larger among male, high ability and STEM subgroups[20].

I also examine the possible non-linear impact of indoor air pollution on test scores by including dummy variables for different levels of pollution exposure simultaneously. Specifically, I define dummies for PM_{10} (µg/m3) being less than 25, between 25 and 50, between 50 and 75, and above 75. The analysis reveals a nonlinear and monotonic relationship between pollution and test scores with a possible threshold at 50 (µg/m3) which is the WHO guideline. Importantly, this threshold is well below current US Environmental Protection Agency (EPA)

[20] The acronym STEM is widely used in the US and refers to academic disciplines of Science, Technology, Engineering and Mathematics.

standards which suggest that it may be economically beneficial to lower existing guidelines[21]. The results imply that taking an exam with pollution above 75 (μg/m3) reduces student's scores by 4.13 points, or approximately 23% of a standard deviation. Finally, I show that transitory decline in cognitive performance has a robust negative relationship with long-term academic indicators that are potentially correlated with future career outcomes. More specifically, I find that exposure to coarse particulate matter reduces student's composite scores and therefore the probability of receiving an upper second classification or above. This is of particular interest since an upper second classification is a threshold requirement for most prestigious jobs and academic graduate programs in the UK.

Overall my results provide compelling evidence that short-term exposures to elevated levels of indoor PM_{10} affect cognitive performance. Epidemiologists have already examined this potential link but such studies are predominantly cross sectional in nature and do not account convincingly for confounding factors (Mendell et al., 2005). To the best of my knowledge, this paper is the first to estimate the causal effect of indoor air pollution on cognitive performance with indoor pollution measures[22]. My findings imply that a narrow focus on health outcomes understate the true cost of pollution as indoor air quality also affects productivity.

The rest of the paper is laid out as follows. In the second section, I present background information on coarse particulate matter and summarize the existing literature on identifying the impact of air pollutions on various health and academic outcomes. Section III describes the data while Section IV presents my identification strategy. In Section V, I present my empirical results and in VI I conclude.

[21] In order to determine the optimal regulatory action a full-fledged cost benefit analysis must be conducted.

[22] Stafford (2015) examines the effect of indoor air quality (IAQ) on academic outcomes in Texas. She found that IAQ renovations have a significant positive effect on standardized tests. However, she was unable to observe actual level of indoor quality, and is therefore forced to rely on variation in the timing of IAQ renovations across schools.

2.2 Background on Air pollution and Cognitive Performance

Particulate matter (PM) is a mixture of solid particles and liquid droplets suspended in the air that consists of various components including acids, metals, dust particles, organic chemicals and allergens. Particle pollution is classified into two main categories namely "inhalable coarse particles" (PM_{10}) and "fine particles" ($PM_{2.5}$) based on their size. The former corresponds to particles that are larger than 2.5 and smaller than 10 micrometers in diameter and the latter to particulate matter that is 2.5 micrometers in diameter or smaller[23]. The size of particles is associated with their ability to cause health problems. Therefore, in 1987, The EPA replaced the earlier Total Suspended Particulate (TSP) air quality standard with a PM_{10} standard and in 1997 also established an annual and 24-hour National Ambient Air Quality Standard (NAAQS) for $PM_{2.5}$.[24] In 2008, the European (EU) Parliament also set legally binding limits for coarse and fine particulate matter. The 2008 EU ambient air quality directive replaced most previous EU air quality legislation and was made law in England in 2010[25].

The air pollution measure in this study is PM_{10}, which comprises of smoke, dirt, dust, mold, spores and pollen. The emission of ambient PM_{10} comes from various sources such as factories, farming and roads. Nevertheless, indoor concentrations of coarse particles are not simply a byproduct of ambient pollution; they are also the result of emissions from indoor sources. The leading indoor sources of particles in education establishments are human activities, plants and various building materials (Chatzidiakou et al., 2012). Indoor concentrations of coarse particles in classrooms tend to surpass outdoor levels during the daytime, which highlights the significant contribution of indoor sources (Madureira et al., 2012)[26]. This is of particular

[23] For comparison, the average human hair is approximately 70 micrometers in diameter, making it 7 times larger than the largest coarse particle.
[24] Total Suspended Particulate corresponds to particles that are less than 100 micrometers in diameter.
[25] Similar regulations also exist in Scotland, Wales and Northern Ireland.
[26] Madureira et al. (2002) also show that $PM_{2.5}$ and PM_1 have the opposite trend.

importance for this study as it suggests that the level of indoor PM_{10} is likely to vary considerably across venues within close proximity of one another, and also within individual venues across time.

The relationship between particulate matter and adverse health outcomes is well documented in the epidemiological literature. The medical explanation for such link is that elevated levels of particles in the air lead to changes in cardiovascular and pulmonary functioning (Seaton et al., 1995). More specifically, human intake of particles may affect respiratory and cardiovascular conditions, such as asthma and heart attacks (Pope at al., 1995; Dockery, 2009; Donaldson et al., 2010; Weinmayr et al., 2010). Particle pollution can also lead to milder health effects such as irritation of the airways, coughing or difficulty breathing.[27] The former types of conditions are likely to be evident in most data sets commonly used in the literature. The latter, however, are likely to be unobserved by researchers as they do not lead to health encounters or even noticed by the affected individual (Chang et al., 2014)[28]. While empirical evidence suggests that symptoms from exposure to particulate matter can manifest within hours or days, it is unclear whether there is also an instantaneous effect (Son et al., 2013). This paper provides a unique opportunity to test this potential immediate effect using a novel quasi-experimental method.

According to the medical and epidemiological literature exposure to particulate matter can affect cognitive acuity but the exact neurodegenerative effect remains largely unexplored (Suglia et al., 2008). One possible explanation is that any deterioration in oxygen quality may in theory impair brain functioning (Clark and Sokoloff, 1999). Air pollution can also impact the

[27] For further details on such effects see http://www.epa.gov/pm/health.html

[28] Schlenker and Walker (2015) show that using an inpatient discharge data substantially underestimate the morbidity effect of ambient pollution. This is because inpatient discharge data excludes emergency room visits which do not require overnight admission.

nervous system, leading to symptoms such as memory disturbance, fatigue and blurred vision which in turn may affect cognitive performance (Kampa and Castanas, 2008).

Despite the growing evidence of strong links between air quality and various health outcomes, research on the effect of air pollution on cognitive performance is remarkably scarce. Epidemiologists have examined such potential link but these studies are predominantly cross-sectional in nature and do not account convincingly for confounding factors (Suglia et al., 2008; Wang et al., 2009). A study by Lavy et al. (2014) examined the causal relationship between ambient pollution and high school exit exams in Israel. In that study they found that increased daily exposure to ambient pollution significantly decrease test scores but were unable to disentangle whether this was caused by exposure during the exam or a build-up effect. Moreover, the results were driven by days with very high levels of pollution which are less frequent in most developed countries and it remained unclear whether lower level of pollution could also lead to reduced cognitive performance in indoor settings[29].

2.3 Data

My data combines self-collected readings of indoor air pollution with administrative data on test scores and demographics of undergraduate students at a leading public research university within the Greater London Urban Area. For exam and demographic information I use a confidential student file which contains the full academic record of all undergraduate students that took exams during the 2012/2013 academic year. The file also contains a unique student identification number which allows me to observe key demographic information on each student

[29] There is also evidence on a link between ambient pollution and indoor physical productivity. More specifically, Chang et al. (2014) found that an increase in ambient PM2.5 leads to decrease in productivity at pear-packing factory.

such as gender, nationality and UCAS tariff points[30]. I also know the exact date, time and location of each exam and the allocation of students across examination sites, allowing me to assign indoor pollution levels to test takers. The indoor pollution data was self-collected from 15 examination sites during the exam term[31]. I used the 3M™ EVM-7 which is an advanced environmental monitor designed to provide real time measurements with a one per second update rate. The monitor provides readings on mean PM_{10} (µg/m3), temperature (Celsius) and relative humidity (%)[32].

According to the WHO, the air quality guidelines for particulate matter are also applicable to indoor spaces (WHO, 2005). Currently, the EPA and the WHO set daily PM_{10} guidelines of 150 and 50 micrograms per cubic meter (µg/m3) respectively[33]. The EPA also report daily air quality using the Air Quality Index (AQI) for the five pollutants regulated by the US Clean Air Act. More specifically, the AQI is divided into six categories ranging from good to hazardous which are associated with different levels of health risks. AQI values above 101, which is about 75 (µg/m3) of PM_{10}, pose various health risks according the EPA[34]. In the UK, The Department for Environment, Food and Rural Affairs use the Daily Air Quality Index (DAQI) to provide information about levels of air pollution and recommended actions and health advice for the same five pollutants. The index is numbered 1-10 and divided into four bands (low, moderate, high and very high). Index value of above 6, which is about 76 (µg/m3) of PM_{10} is defined as high level of pollution in the UK. In my empirical analysis I mainly use the more

[30] The UCAS tariff is a means of allocating points to pre-university qualifications, allowing a broad comparison to be made across a wide range of international qualifications. The tariff points system assist British universities with their admission decisions and their management information. For further details see https://www.ucas.com/ucas/undergraduate/getting-started/entry-requirements/tariff.

[31] The exam period lasts for approximately 4 weeks during the month of May.

[32] More specifically, this is a 6 minutes average taken from each examination site and the monitor was placed at least one meter from the wall and 1.5 meters height from the floor to ensure reliable readings (WHO, 2011).

[33] According to the EPA, an area meets the 24-hour PM_{10} standard if it does not exceed the above level more than once per year on average over a three-year period.

[34] The WHO guideline of 50 (µg/m3) is equivalent to an AQI of 46 which is in the "Good" category.

conservative WHO guideline to generate a threshold dummy which classifies exposure beyond the 50 (µg/m3) standards as hazardous[35].

Table 2.1 presents descriptive statistics of key variables of interest. My sample includes 11,522 examination results of 2,458 students taking exams in 14 different venues across 18 days. Each student took 5.2 exams on average, and the pass rate was 83%. In columns (2)-(5) I stratify the sample by gender and ability. I use UCAS tariff points, which is a means of allocating points to pre-university qualifications, as a proxy for student ability. The table indicates that there are more females in the sample and that they tend to achieve marginally better scores. As expected, the high ability subgroup achieved significantly higher marks compared to their low ability counterparts. It is important to note that pollution, temperature and relative humidity do not vary much by sub-population. The similarity in these observables across gender and ability is important as it suggests that selection on observables is unlikely to drive my results.

2.4 Empirical Strategy

For identification, I crucially rely on the panel structure of the data to estimate models with subject and student fixed effects. More formally, I estimate the following specification:

$$(1)\, R_{ist} = \beta_0 X_{it} + \beta_1 PM_{st} + \beta_2 f(Temp_{st}, RH_{st}) + \beta_3 NUM_{st} + Day_t + TOD_t + Dur_{st} + Site_s + I_i + \varepsilon_{ist}$$

where R_{ist} is the test score of student *i* at site *s* at time *t*; X_{it} is a vector of individual characteristics possibly related to test outcomes, such as gender ; PM_{st} is PM_{10} level at site *s* at time *t*; $Temp_{st}$ is the temperature[36] at site *s* at time *t*; RH_{st} is the relative humidity measure at site *s* at time *t*; NUM_{st} is the number of students taking exam at site *s* at time *t* Day_t , TOD_{st} , Dur_{st}

[35] I also examine lower and higher thresholds in my analysis (see table 2.3 for further details).
[36] I include linear and quadratic terms for relative humidity, 5^{O} bins for temperature, and linear and quadratic interaction terms of mean temperature and relative humidity.

and $Site_s$ are day-of-week, time-of-day, duration and examination site fixed effects respectively; I_i is fixed effect for the individual; and ε_{ist} is an idiosyncratic error term[37]. In order to accurately account for both spatial and serial correlation I use two-way cluster robust standard errors, clustering on both examination site and date[38].

There are three main econometric challenges in identifying the causal effect of air pollution on test scores. First, the possible correlation between pollution exposure and unobserved determinants of students' test scores. For example, if wealthy individuals are sorting themselves into degree subjects that exposed to lower levels of pollution (e.g. better facilities); naïve OLS estimation may underestimate the true causal effect of pollution as it is potentially mitigated by other factors (e.g. private tuition). In order to absorb these potential unobserved time invariant variations in subjects or individuals, I include individual fixed effects in equation (1). I also include controls for time-varying factors, such as daily temperature and relative humidity that could be contemporaneous with pollution. Nevertheless, it is still possible that other unobserved factors that are correlated with both pollution and test scores are still present. In order to limit such concern, I conduct a rich set of placebo tests which are discussed in detail in the next section of this paper.

The second challenge is measurement error in assigning pollution to individuals. Most studies assign pollution data from ambient air pollution monitors to individuals using various interpolation techniques. This is likely to yield some degree of measurement error due to the significant spatial variation in pollution even within finely defined areas (Moretti and Neidell 2011, Lin et al. 2001). In addition, since exams are taken indoors and normally a few miles away

[37] Note that in a different specification I use subject fixed effects in place of the student fixed effects. Subject fixed effect is defined as department and year of study (for example, a second year economics student).

[38] As a robustness check I also clustered at both the student and the examination site level separately. While the former tends to have smaller standard errors the latter yield very similar standard errors as the two-way clustering used in this paper. HAC robust standard errors also yield smaller standard errors and I therefore decided to use the most conservative clustering strategy.

from an ambient monitor station, measurement error is likely to be exacerbated. These concerns are not present in this study as pollution data is collected from inside the examination site. This feature also allows me to ensure that I estimate the effect of exposure during the examination and not the potential build-up effect from exposure to pollution on the way to the exam[39].

Heterogeneity in avoidance behavior is the third challenge for causal inference. The concern is that optimizing individuals will alter their pollution exposure to protect their health as air pollution information is widely available to the public. For example, if sensitive groups adopt compensatory behavior in response to a media alert, equation (1) is likely to understate the true causal effect of PM_{10}. This concern is unlikely to arise in my setting for three reasons. First, the allocation of students across examination sites is determined centrally by the university a few weeks prior to the examination date[40]. Second, unlike ambient pollution, information on indoor pollution levels is unavailable to students. Third, the level of pollution measured in my study is sufficiently low for students not to visually notice the presence of pollution.

Figure 2.1 and 2.2 present compelling evidence on the exogeneity of indoor PM_{10} in this study. Figure 2.1 plots the variation of PM_{10} within a day across different examination sites. As evident from the figure, there is substantial variation across and within sites in a given day. Figure 2.2 which plots the variation of PM_{10} across days within a single examination site, shows a high frequency variation across days, with no evidence for a systematic pattern. Figures 2.1 and 2.2 exemplify the significant time and spatial variation of the data and further reduce concerns regarding possible omitted variable bias in this setting.

[39] Any build up effect is implicitly captured by controlling for day and time of exams, which among other factors will also capture ambient pollution.
[40] A student not attending is deemed to have failed unless extenuating evidence are provided, as such there is no possible selection into different time or examination venue.

2.5 Empirical Results

2.5.1 Main Results

Table 2.2 reports on the link between indoor coarse particulates and test scores. In the first two columns of panel A, I present cross sectional correlations between the continuous PM_{10} measurement and student achievement. The coefficient estimates without any controls, column (1), suggest that a 1 unit increase of PM_{10} is associated with a 0.08 points decrease in a student's test scores. In column (2) I add controls for age, gender, temperature, relative humidity, class size and dummies for day-of-week, examination venue, duration and nationality. I find that a 1 unit increase in PM_{10} is associated with a 0.075 decline in test scores. Both estimates are statistically and economically significant but are cross sectional in nature and therefore should be treated with caution.

In the last two columns of Table 2.2 I exploit the panel structure of the data to estimate models with subject and student fixed effects. Column (3), which includes a subject fixed effect, also shows a negative and highly significant effect with a more precise estimate. In order to account for potential confounders at the student level, column (4) estimates my preferred specification using within student regression. I find that a 1 unit increase in PM_{10} leads to a 0.06 decline in a student's test score, an estimate significant at the 1 percent level. These results imply that a student sitting an exam at a site with an average pollution level (33.15 µg/m3) will suffer a substantial reduction of 0.08 standard deviations in test score, as against that which the same person would have achieved at a site with the lowest level of pollution (4 µg/m3).

Panel B of Table 2.2 reports on the effect of PM_{10} being above 50 (µg/m3) which the WHO considers to be an unhealthy level threshold. The results present negative and significant effects of coarse particles on students' performances under most specifications. In column (4), where I include student fixed effects, I find that taking an exam at a site with pollution level above the WHO standard is associated with a 2.868 decline in a student's test score, which is equivalent to 0.15 of a standard deviation. This effect is very large and similar to estimates found in paying teachers and students large financial incentives (Jackson, 2010) and reducing class size from 31 to 25 students (Angrist and Lavy, 1999). Finally, it is worth noting that the results obtained using the dichotomous indicators suggest the possibility of non-linear relationship between indoor pollution and cognitive performance.

As such, In Table 2.3 I examine the possible non-linear impact of PM_{10} on test scores by including dummy variables for different levels of pollution exposure simultaneously. Specifically, I define dummies for PM_{10} (µg/m3) being less than 25, between 25 and 50, between 50 and 75, and above 75. I find no significant effect for PM_{10} levels below the WHO standard. Column 4, which show results for my preferred specification using student fixed effect indicates that PM_{10} exposure between 50 and 75 (µg/m3) is significantly associated with a 2.277 decline in the student's score. When PM_{10} reaches 75 (µg/m3) the effect increases to 4.132, which is also significant at the 5% level. Importantly, these results suggest a threshold around 50 (µg/m3) which is well below current EPA standards and therefore it may be economically beneficial to lower existing guidelines. Also note that both the WHO and the EPA guidelines are for 24-hour and there are no existing standards for hourly exposure to PM_{10}. Therefore, my results may suggest a daily threshold below 50 (µg/m3) as air pollution tends to be higher during the day.

2.5.2 Heterogeneity

In this section I explore whether indoor air pollution has a heterogeneous effect across sub-populations and academic disciplines. The reason for this investigation is twofold; first, to test whether some subgroups are more sensitive to indoor pollution than others; and second, to examine whether the effect of indoor pollution varies by subjects. To study the former I stratify by gender and ability and for the latter I break down my sample by subject since the cognitive demands of the exams may differ.

Table 2.4 present estimates on the effects of coarse particulate on test scores stratified by gender, ability and subject, using my preferred specification with student fixed effects. In the first two columns I break down the sample of test takers by gender. Column (1), which reports on the effect for the male subsample only, shows a negative and significant link between indoor levels of PM_{10} and test scores. More specifically, the results suggest that a 1 unit increase in PM_{10} (µg/m3) reduces students' test scores by 0.086 and being above the WHO threshold reduces students' test scores by 3.167. These estimates are considerably higher than the results obtained in the analysis for the full sample which suggests that male students are more sensitive to coarse particulate than their female counterpart. Indeed, Column (2), which reports results for the female subgroup, demonstrates such pattern precisely. The continuous coefficient drops to 0.035 and the threshold dummy declines to 1.406 and is only significant at the 10% level. The results are not statistically different from each other but are suggestive of heterogeneous effect.

A potential explanation for such difference is the higher prevalence of Attention Deficit Hyperactivity Disorder (ADHD) among male students which possibly makes them more vulnerable to distractions induced by elevated levels of indoor pollution (Biederman et al. 2002).

Columns (3) and (4) of Table 2.4 report on the effects of coarse particulate matter on students' test scores by my ex-ante ability measure. As a proxy for ability I use UCAS tariff points, which are a means of allocating points to pre-university qualifications, and I break down the sample above or below the ability median. The results suggest that the effect of indoor air pollution on cognitive performance is larger among high ability students. Specifically, an additional unit of PM_{10} is associated with a 0.070 decline in students' test scores compared to 0.054 among low ability students. When I use the dichotomous measure I find that exposure to indoor PM_{10} reduce test scores for low and high ability types by 2.824 and 3.098 respectively[41]. One possible explanation for this finding is the reasonable assumption of decreasing marginal returns to effort. Hence, high achievers may be more sensitive to random disturbances, such as indoor pollution, since any additional mark requires higher effort.

In the last two columns of Table 2.4 I examine the effect of indoor PM_{10} on different academic disciplines. I follow the guideline of the National Science Foundation (NCF) and stratify my sample into two groups; Science, Technology, engineering, and mathematics (STEM) and all other disciplines (non-STEM)[42]. The motivation for this analysis is to explore if some types of mental tasks are more sensitive to indoor air pollution. The results show that the effect is very large for STEM disciplines (-0.090) compared to the estimate for non-STEM subjects of -0.038. The results suggest that tasks which require higher degree of numerical functioning are more affected by pollution.

2.5.3 The Effect of Indoor Air Pollution on Other Academic Outcomes

In this section I study whether transitory impaired cognitive performance also leads to long-term adverse effects by looking at key academic indicators that are potentially correlated with future career outcomes. In Table 2.5, I estimate the effect of PM_{10} on the probability of failing an exam. The results are highly significant for both the continuous and threshold measures, and suggest that being above the WHO standard increases the probability of failing an exam by 5.3 percentage points. Failing an exam can have a substantial adverse effect on

[41] Note that the estimates are not statically different from each other and they are only suggestive.
[42] Note that the NSF uses a broader definition of STEM which also includes social sciences. In my empirical analysis I classified only one social science (economics) as a STEM subject.

student's future career path due to three main reasons. First, failing can delay graduation and may lead to a change in degree title[43]. Second, since most graduate schemes in the UK require submission of full transcript during the application process, failing an exam can send a bad signal to potential employers. Finally, in case of a retake a student can receive no more than 40 points (a pass) regardless of his or her actual examination score. Hence, failing an exam has a substantial effect on final degree classification which may affect a student's career options.

In Table 2.6, I carry the analysis at the student level. Therefore, the treatment is the average pollution exposure across all examinations. Note that the dichotomous indicator is the fraction of exams above threshold (50 µg/m3) over all exams. In Panel A, I estimate the effect of exposure to coarse particulate matter on students' composite score. The results indicate that an additional 10 units of PM_{10} and a 10% increase in the number of above threshold exposure are associated with a 1.922 and 0.933 decline in a student's composite score respectively. In Panel B, I examine the effect of indoor pollution on the probability of receiving a classification of upper second or above. This is of particular interest as an upper second classification is a threshold requirement to most prestigious graduate jobs and academic graduate programs in the UK[44]. The results show that an additional 10 units of PM_{10} and a 10% increase in the number of above-threshold exposures reduces the probability of a student achieving a second class classification by 4.5 and 19.8 percentage points respectively.

[43] For example, a student that study for a BSc in Management with Economics and fail the core microeconomic module can still graduate with a BSc in Management which may limit future career options.
[44] According to the Association of Graduate Recruiters, 78% of UK employers require an upper second classification (http://www.bbc.co.uk/news/10506798).

2.5.4 Robustness Checks

In this section I conduct two placebo exercises and robustness checks to ease concerns that my estimates may capture unobserved time varying factors which are correlated with both indoor air pollution and test scores. The first placebo exercise uses the level of air pollution from the previous exam as the coefficient of interest. Hence, if equation (1) is correctly specified the coefficients of the lag variables should not be statistically different from zero. The results in Panel A of Table 2.7 suggests that my preferred specification with student fixed effect is indeed not statistically significant. However, the OLS estimates with and without controls are highly significant which exemplifies the importance of controlling for individual unobserved characteristics. Note that the estimate with the subject fixed effect is also insignificant which is reassuring as the last section of the analysis was conducted at the student level and therefore cannot include student fixed effects.

In Panel B, I perform an additional placebo exercise using my ex-ante measure of ability as the dependent variable. Since UCAS tariff points are based on pre-university achievements, they should not be correlated with exposure to indoor air pollution during university after accounting for unobservables. Column (3) which includes subject fixed effect shows that the relationship between indoor levels of PM_{10} and pre-university qualifications is not statistically significant. Again, OLS estimates with and without controls are significant at the 5% and 1% levels respectively. Overall, these results are of great importance for two main reasons. First, they lends further supports to the casual interpretation of my results as it reduces concerns over time varying characteristics that my main specification may fail to capture. Second, it demonstrates that OLS estimates, even with a rich set of controls, may still suffer from omitted variable bias.

Finally, in Panel C I examine whether my estimates capture only the transitory pollution exposure and verify that it is not related to prior exposure. More specifically, I estimate the correlation between the last exam score and the average pollution level from all previous exams. The results are not statistically different from zero in all specifications.

2.6 Conclusion

In this paper I analyze the relationship between short-term exposure to indoor coarse particles and cognitive performance. I perform my analysis using a unique merged data set of indoor PM_{10} levels and administrative student data. I find that a one unit increase in PM_{10} (μg/m3) and being above the WHO guideline reduces student's test scores by 0.060 and 2.868 respectively. I also explore whether indoor air pollution has a heterogeneous effect across sub-populations and academic discipline and find the effect is larger among male, high ability and STEM subgroups.

While my results are robust to a wide range of different specifications it is important to note a few caveats that may limit my analysis. First, since I do not observe the exact composition of my PM_{10} readings, I can not identify whether specific components of coarse particulates are driving my results. Second, despite my rigorous identification strategy, which includes student fixed effects and a rich set of controls it is still possible that other time variant unobserved correlated factors are still present. For example, traffic on the way to the exam can be correlated with both pollution levels and test scores as heavy traffic can increase pollution and stress. Finally, since data on individual health conditions is unavailable I'm unable to identify the exact pathophysiological pathways that drive my results which may be a rewarding area for future research. Despite the above limitations this paper provides compelling evidence on the causal link between indoor air pollution and cognitive performance.

This analysis suggests that a narrow focus on traditional health outcomes, such as hospitalization and increased mortality, may significantly understate the true cost of pollution. This is since mental acuity is essential to most professions and therefore a reduction in indoor air quality can reduce productivity. My analysis also shows that the effect of indoor air pollution on cognitive performance is present at levels considerably lower than current EPA mandates. This is of particular importance as the EPA is currently reviewing whether revisions to the current PM_{10} standards are warranted[45].

[45]For more details see http://www3.epa.gov/airtrends/aqtrnd95/pm10.html.

Table 2.1

Descriptive Statistics

Variable	All (1)	By Gender: Males (2)	By Gender: Females (3)	By Ability: Low (4)	By Ability: High (5)
PM_{10}	33.35	33.66	33.10	34.50	32.33
($\mu g/m^3$)	(21.51)	(21.78)	(21.29)	(21.93)	(21.08)
PM_{10}	0.214	0.215	0.212	0.227	0.202
(PM_{10} >50)	(0.41)	(0.41)	(0.41)	(0.23)	(0.40)
Exam Score	54.59	53.67	55.34	50.59	58.63
(1-100 points)	(18.05)	(19.41)	(16.82)	(19.09)	(15.86)
Temperature	16.37	16.51	16.26	16.40	16.35
(celsius)	(2.31)	(2.16)	(2.41)	(2.29)	(2.32)
Relative Humidity	54.02	54.43	53.67	53.62	54.38
(percent saturation)	(12.15)	(12.08)	(12.20)	(12.26)	(12.07)
Age	21.36	21.46	21.27	21.96	20.71
	(2.77)	(2.95)	(2.61)	(3.42)	(1.18)
Number of Exams	5.178	5.262	5.109	5.066	5.161
	(1.41)	(1.47)	(1.34)	(1.45)	(1.32)
Number of Students	124.6	122.6	126.2	123.7	125.2
	(75.68)	(76.24)	(75.18)	(75.89)	(75.38)
Failed an Exam	0.172	0.201	0.149	0.240	0.104
(yes=1)	(0.38)	(0.40)	(0.36)	(0.43)	(0.31)
Observations	11,522	5,189	6,333	5,580	5,806

Notes: Standard deviations are in parentheses. Relative humidity is the amount of moisture in the air as a share of what the air can hold at that temperature. The ability level is based on UCAS tariff points which is a means of allocating points to pre-university qualifications. The sample is split by whether the student is above or below the median.

Table 2.2

Pooled OLS and Fixed Effect Models of Indoor Air Pollution's Impact on Test Scores

	Polled OLS		Fixed Effects	
	No Controls	Controls	Subject	Student
	(1)	(2)	(3)	(4)
PM_{10} (µg/m3)	-0.083***	-0.075**	-0.075***	-0.060***
	(0.027)	(0.030)	(0.020)	(0.020)
Dummy for PM_{10}>50	-2.814*	-2.697**	-3.445***	-2.868***
	(1.461)	(1.341)	(0.860)	(0.818)
Observations	11,730	11,522	11,522	11,522

Notes: Each cell in the table represents a separate regression.Standard errors are heteroskedastic-consistent and clustered by examination venue and date of pollution assignment. All regressions include suppressed controls for temperature and humidity. *** Signifcant at the 1 percent level, ** Signifcant at the 5 percent level, * Signifcant at the 10 percent level.

Table 2.3

Indoor Air Pollution's Impact on Test Scores

	Polled OLS		Fixed Effects	
	No Controls	Controls	Subject	Student
	(1)	(2)	(3)	(4)
Dummy for PM_{10} >25 & <= 50	-3.087**	-1.980	-1.418	-0.778
	(1.480)	(1.205)	(0.937)	(0.977)
Dummy for PM_{10} >50 & <= 75	-2.920*	-2.374	-2.970**	-2.277**
	(1.540)	(1.702)	(1.138)	(1.082)
Dummy for PM_{10} >75	-5.488**	-5.530**	-5.453***	-4.132**
	(2.473)	(2.377)	(1.636)	(1.797)
Observations	11,730	11,522	11,522	11,522

Notes: See Table 2.2. Each column in the table represents a separate regression.

Table 2.4

Heterogeneity in the Impact of Indoor Air Pollution on Test Scores

	Gender		Ability		Degree Subject	
	Males	Females	Low	High	STEM	non-STEM
	(1)	(2)	(3)	(4)	(5)	(6)
PM_{10} (μg/m3)	-0.086***	-0.035	-0.054**	-0.070***	-0.090***	-0.038**
	(0.021)	(0.022)	(0.023)	(0.021)	(0.032)	(0.019)
Dummy for PM_{10}>50	-3.167***	-1.406*	-2.824***	-3.098***	-3.553**	-1.434*
	(0.781)	(0.791)	(0.999)	(0.804)	(1.634)	(0.738)
Observations	5,189	6,333	5,596	5,822	7,187	4,270

Notes: See Table 2.2. All specifications include student fixed effects.

Table 2.5

Indoor Air Pollution's Impact on Failing an Exam

	Pooled OLS		Fixed Effects	
	No Controls	Controls	Subject	Student
	(1)	(2)	(3)	(4)
PM_{10} (μg/m3)	0.001	0.001*	0.001***	0.001**
	(0.001)	(0.001)	(0.000)	(0.000)
Dummy for PM_{10}>50	0.036	0.057*	0.069***	0.053***
	(0.034)	(0.033)	(0.020)	(0.018)
Observations	11,730	11,522	11,522	11,522

Notes: See Table 2.2. All specifications include student fixed effects.

Table 2.6

Indoor Air Pollution's Impact on Other Academic Outcomes

	Pooled OLS		Fixed Effects
	No Controls	Controls	Subject
	(1)	(2)	(3)
Panel A: Composite Score			
PM_{10} (μg/m3, 10 units)	-1.510**	-1.750***	-1.922***
	(0.630)	(0.625)	(0.454)
Dummy for PM_{10}>50	-3.527	-5.278	-9.333***
	(5.427)	(4.023)	(2.722)
Observations	2,462	2,458	2,458
Panel B: Upper Second Class (yes=1)			
PM_{10} (μg/m3, 10 units)	-0.055***	-0.060***	-0.045***
	(0.019)	(0.015)	(0.011)
Dummy for PM_{10}>50	-0.109	-0.146	-0.198***
	(0.139)	(0.105)	(0.072)
Observations	2,462	2,458	2,458

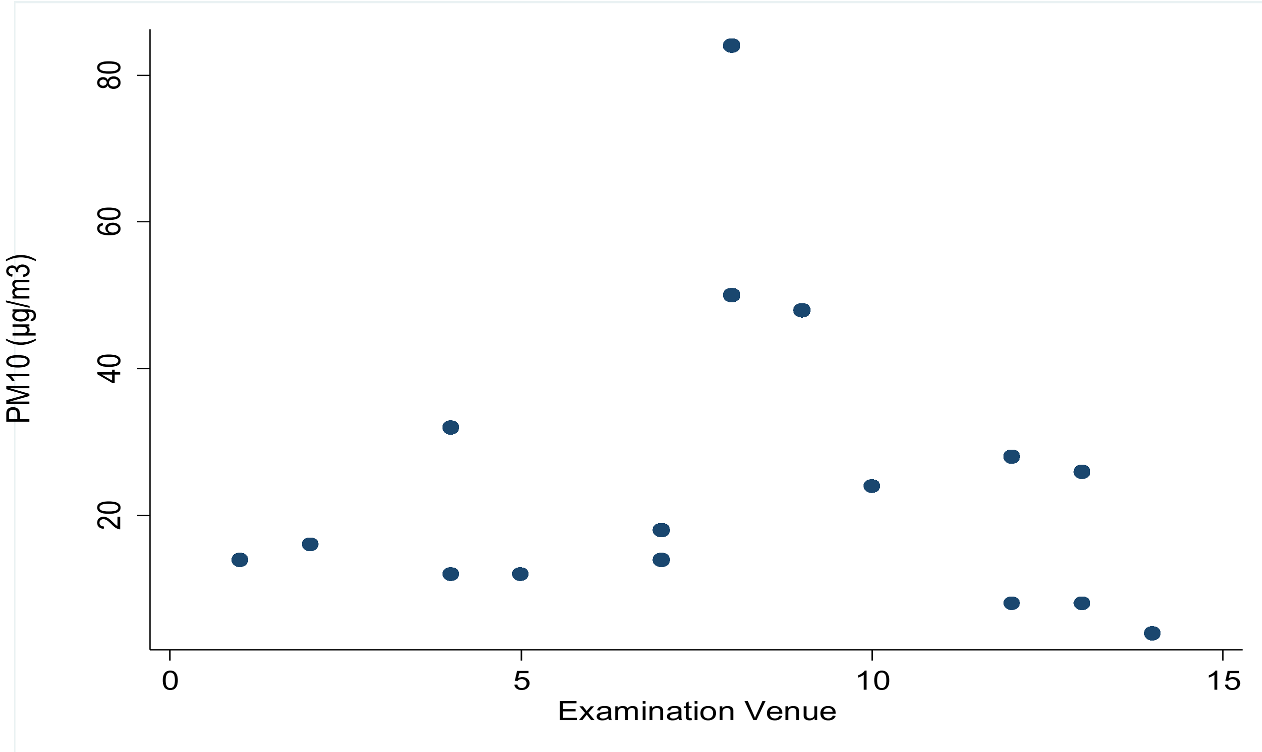
Notes: Each observation is a student and pollution is averaged over all of the tests taken. Standard error are heteroskedastic-consistent and clustered at deaprtment and year level.

Table 2.7
Placebo and Robustness Tests

	Pooled OLS		Fixed Effects	
	No Controls	Controls	Subject	Student
	(1)	(2)	(3)	(4)
<u>Panel A: Previous Exam</u>				
PM_{10} (µg/m3)	-0.086***	-0.056**	-0.027	-0.008
	(0.025)	(0.022)	(0.020)	(0.022)
Observations	9,268	9,079	9,079	9,079
<u>Panel B: UCAS Tariff Points</u>				
PM_{10} (µg/m3)	-0.982**	-1.490***	-0.675	
	(0.445)	(0.381)	(0.831)	
Observations	2,438	2,438	2,438	
<u>Panel C: Prior Pollution</u>				
PM_{10} (µg/m3)	-0.039	-0.041	0.035	
	(0.045)	(0.034)	(0.031)	
Observations	2,462	2,458	2,458	

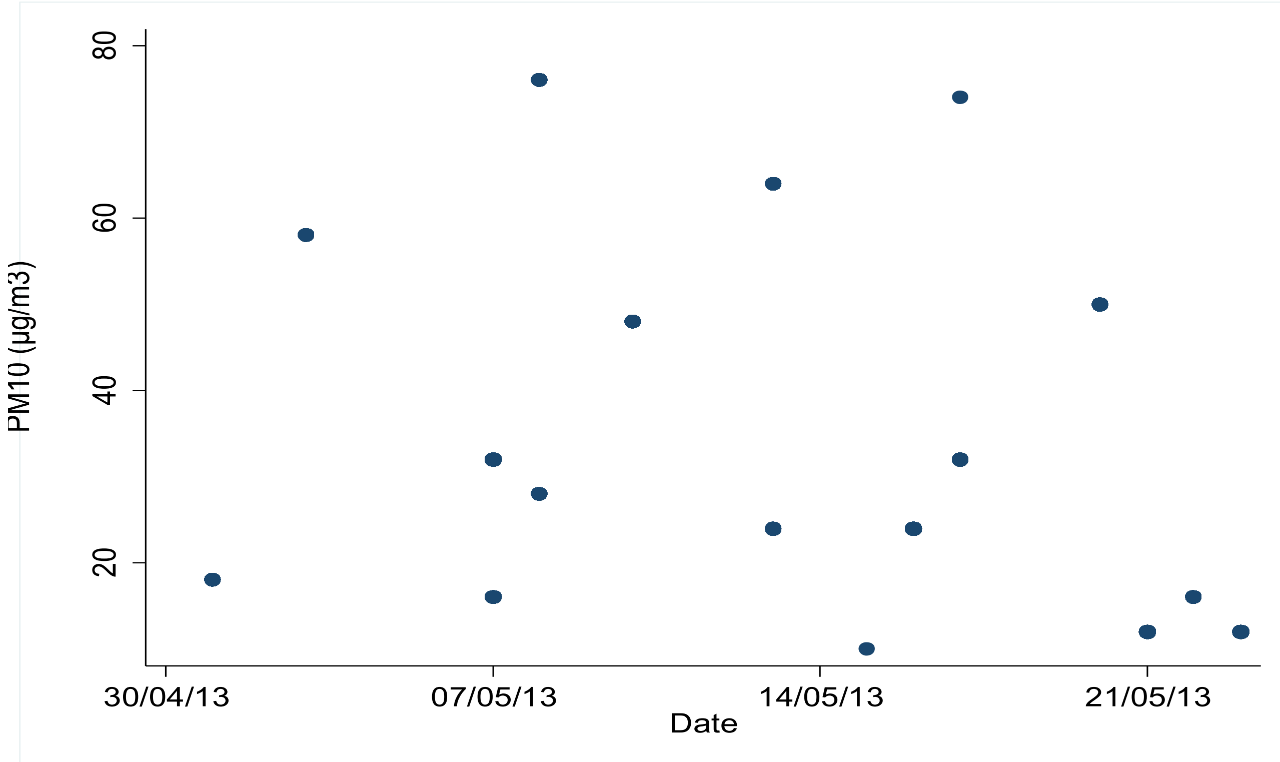
Notes: See Tables 2.2 and 2.6. In panel A, I assign PM10 to each exam using the reading of PM10 for the previous exam of the same student to the actual exam. In Panel B, I use my ex-ante measure of ability as the dependent variable. In Panel C, My dependent variable is the final test score and the independent variable is the average pollution level from all previous exams of the same student.

Figure 2.1
Within Day Variation



Notes: Example of variation in PM10 within a day across venues

Figure 2.2
Variation Across Days



Notes: Example of variation in PM10 across days within one examination venue

Chapter 3

The Long Run Human Capital and Economic Consequences of High-Stakes Examinations

3.1 Introduction

Cognitive performance is critical to scholastic achievement and to successful performance in many occupations. Cognitive acuity can be affected temporarily by a variety of factors, including the intake of caffeine, nicotine, and sleep deprivation (Jarvis 1993, Angus, Heslegrave and Myles 1985). Factors that induce variation in the cognitive performance of students during high-stakes exams may be of particular interest because they potentially affect long-term schooling attainment, and therefore can have permanent effects on wages and earnings. In this paper, we examine the potential long-term effect of transitory disturbances to cognitive performance during high-stakes exit exams in Israeli high schools. The exams are known as a the *Bagrut* and are a critical component of Israel's college admissions system, acting as a gatekeeper for the most selective schools, similar to the role played by Scholastic Aptitude Tests in the United States or A-levels in England and other European countries. Access to college majors is also determined by *Bagrut* performance, with many lucrative professional programs requiring minimum overall average scores for admission, such as law and medicine. As a consequence, *Bagrut* scores can affect an individual's entire academic career, and subsequent labor market outcomes.

Although many countries use high-stakes testing to rank students for college admission, the consequences of this policy are largely unknown. Does having a particularly good or bad performance on a high-stakes examination have long-term consequences for test takers, after accounting for a student's cognitive ability? Insofar as there are permanent wage consequences to variation induced by completely random shocks to student performance on a test like the *Bagrut*, it suggests that the use of high-stakes testing as a primary method for ranking students may be inefficient. Aggregate welfare may also be reduced by relying too heavily on examinations that provide noisy measures of student quality, since it may lead to poor matching between students and occupations, and an inefficient allocation of labor. Recent debate over the planned redesign of the SATs has been in part motivated by concerns that the current version is highly random and does not represent a fair measure of student quality (New York Times 2014).[46] In spite of a dearth of evidence regarding the consequences of these tests, they are used extensively globally to rank students and allocate opportunity by acting as a gatekeeper in admissions.

Assessing the consequences of using high-stakes examinations for ranking students is challenging. First, large data samples are generally not available with standardized test scores and wages during adulthood for a representative population.[47] Second, since higher-ability students presumably perform better on high-stakes tests, it is difficult to separately distinguish the return to cognitive ability from the return to doing well on the examination. One possible solution is to examine the consequences of fluctuations in a random component affecting performance on these tests. A candidate is fluctuation in air pollution that might have an effect on cognitive acuity and test scores, therefore generating plausibly random variation in a given student's outcome. Air pollution has been demonstrated to adversely affect human productivity

[46] In a recent discussion of the planned revisions to the SATs, the president of the College Board stated that "only 20 percent…see the college-admission tests as a fair measure of the work their students have done." (New York Times 2014)

[47]Note that in the United States, Educational Testing Service (ETS) is notoriously private and no scholarship (to our knowledge) has been carried out linking SAT scores to adult outcomes for even small subsets of the population. For military recruits, the ASVAB has been made available but it is unclear how relevant this is for other sub-populations (Cawley et al. 2001).

across a variety of tasks (Graff Zivin and Neidell 2012, Chang et al. 2014). Since students are assigned to test sites without prior knowledge of pollution or the ability to reschedule, it represents an exogenous factor affecting performance. This may enable direct measurement of the return to the component of a student's score which is related entirely to luck, and provide evidence regarding whether these tests do or do not have long-term consequences.

In this paper, we examine the long-term consequences of high-stakes examinations using exogenous variation in scores generated by pollution exposure. We examine student performance on the *Bagrut*, a series of examinations across different subjects that Israeli students must pass as a prerequisite for entry into elite universities. The analysis presented here builds on earlier work where we estimated the reduced form causal impact of various pollutants on cognitive performance (Lavy et al. 2014). In Lavy et al. (2014), we exploit exam-level data to demonstrate that pollution harms student performance, even in models with student-fixed effects where we exploit variation in pollution across test administrations for the same student. We find significant effects of particulate matter ($PM_{2.5}$) and carbon monoxide (CO) on student performance, and demonstrate in a set of placebo exercises that the effects are very short-term: no significant effect is found of pollution on days before or after the exam. In this paper, we focus more narrowly on fine particulate matter ($PM_{2.5}$), considered to be particularly harmful to human health and cognition, and use it as an instrumental variable for a student's average *Bagrut* score across the series of examinations. Since Israel has frequent sandstorms, we are able to exploit severe short-term pollution which generates a first-stage relationship between pollution exposure and a student's *Bagrut* scores.[48] Another feature that makes Israel a suitable context for answering this research question is data availability. Israel's national registration system allows us to match the

[48] We note here that most of the Southern and central European countries, even England, are seriously affected by the same sand storms that originate from the Sahara. See for example media coverage of early April 2014 episodes of extremely high pollution levels in this region at: http://www.bbc.com/news/uk-26844425.

universe of students who take the *Bagrut* with their completed post-high school education and their wages in adulthood, enabling us to examine the impact of these tests in the long run for every test taker.

In the first part of our analysis, we examine the relationship between average pollution exposure during the *Bagrut* exams and long-term academic and economic outcomes. We exploit variation across students at the same school in their average pollution exposure, which stems from differences in the exam dates across different subjects. The identification assumption is that variation across students from the same school in the dates of *Bagrut* exams is not correlated with potential outcomes, which is a plausible assumption because dates of national *Bagrut* exams are determined by the Ministry of Education, and students choose their *Bagrut* study program years before the dates of exams are determined.[49] Using this design, we present evidence that completely random variation in *Bagrut* scores has a long term impact on both academic and economic outcomes. We estimate that an additional 10 units of $PM_{2.5}$ (AQI) in average exposure is associated with a .023% decline in a student's *Bagrut* composite score, a 0.03% percent decline in *Bagrut* matriculation certification, a .019% decline in enrollment in post-secondary education, and a 0.15 decline in years of education at a university. The wage consequences are also significant, with an additional 10 units of $PM_{2.5}$ in average pollution exposure lowering monthly income by 109 Israeli shekel ($29), roughly a 2% decline. This suggests that students who take an exam during a severe pollution episode experience non-trivial long run consequences, both academically and economically.

[49] In support of this assumption, we provide the results of balancing tests that demonstrate that pollution does not appear to be correlated with observable features of the student, after conditioning on the student's school. As additional support for our approach, we cite evidence presented in Lavy, Ebenstein and Roth (2014) where we use exam-level data to estimate the effect of pollution on test scores, allowing us to simultaneously control for time-invariant features of localities, schools and individual students. We present evidence that the results are broadly similar with student fixed effects or school fixed effects, supporting our identification strategy.

We then exploit the strong first-stage relationship between average pollution exposure during *Bagrut* exams and *Bagrut* composite scores to estimate the economic return to each additional point on the *Bagrut*. Using 2SLS, we estimate that each point is worth (in 2010) 66 shekels ($18) in additional monthly earnings. Since the standard deviation on *Bagrut* composite scores is roughly 24 points, this implies that there are significant wage consequences to the exam, even for relatively small deviations in one's score. We examine mechanisms for this result by examining our other academic outcomes, treating the *Bagrut* composite score as the endogenous regressor and using pollution as our instrument. We find that each additional instrumented point increases the probability of receiving a matriculation certificate by 2 percentage points, enrollment rates in post-secondary academic schooling by 1.9 percentage points, and post-secondary education by .092 years. This suggests that the mechanism for the *Bagrut*'s impact on student outcomes is through the posited channel of affecting a student's prospects for post-secondary education.

In light of the strong relationship between *Bagrut* pollution exposure and post-secondary education, we are able to use 2SLS to estimate the implied return to education. We estimate that an additional post-secondary year of schooling is worth 707 shekels ($191) per month, an implied return to college education of 14%. This rate is only marginally higher relative to existing estimates found in Israel and elsewhere for the return to post-secondary education (Frish 2009, Oreopoulos and Petronijevic 2013, Angrist and Chen 2011, Angrist and Lavy 2013).[50] As we discuss, the exclusion restriction may be violated in this context, but the similarity of our estimates to existing estimates of the return to post-secondary education is supporting evidence

[50] For example, Angrist and Lavy (2009) estimate that *Bagrut* holders earn 13 percent more than other individuals with exactly 12 years of schooling.

that the magnitude of our estimated effects of $PM_{2.5}$ on schooling and economic outcomes are reasonable.[51]

In the last section, we examine the heterogeneity of pollution's effect on scores and wages across sub-populations in Israel. We find that our first-stage relationship is strongest among populations of Israel which have the highest rates of asthma, suggesting a physiological mechanism for the observed relationship between pollution and exam outcomes. Our estimated effects are four times larger for boys than girls, twice as large for weaker students relative to stronger students, and a fifth larger for students from lower SES relative to students from high SES background. We then examine the impact of pollution on long run outcomes for these groups. We estimate that a point on the *Bagrut* is worth more for boys than girls (78 shekels versus 59 shekels), for stronger students than weaker students (124 shekels versus 80 shekels), and for higher SES students relative to lower SES students (105 shekels versus 56 shekels). These magnitudes suggest that the return to an extra point is quite substantial, especially for already-strong students or students from privileged backgrounds, who presumably can capitalize on the opportunity of gaining admission to a longer academic programs or professions that require long (and poorly paid) internships, like law or medicine.

Our analysis highlights a major drawback of using high-stakes examinations to rank students. If completely random variation in scores can still matter ten years after a student completes high school, this suggests that placing too much weight on high-stakes exams like the *Bagrut* may not be consistent with meritocratic principles. A second implication of this finding is that by temporarily lowering the productivity of human capital, high pollution levels lead to allocative inefficiency as students with higher human capital may be assigned a lower rank than

[51] Since the Bagrut composite score directly affects the post-secondary education options available to a student, 2SLS models of the return to post-secondary education using pollution as an instrument will be biased by the omission of the Bagrut composite score.

their less qualified peers. This may lead to inefficient allocation of workers across occupations, and possibly a less productive workforce. The results highlight the danger in assigning too much weight to a student's performance on a high-stakes exam, rather than their overall academic record. Our results also contribute to existing evidence that transitory random shocks that are unrelated to levels of human capital can have long-term implications for earnings (Oreopoulos et al. 2012). Our study also highlights the role of luck in determining wages, a pattern noted by other scholars (Bertrand and Mulinathan 2001).

The rest of the paper is laid out as follows. In the second section, we present relevant background information on air pollution and cognition, and on the controversial use of high-stakes examinations in college admissions both in Israel and abroad. In Section III, we present our empirical results for the overall sample and examine heterogeneity in the results across sub-populations in Section IV. We conclude in Section V.

3.2 Background and Data

3.2.1 Air pollution and Cognitive Performance

The existing literature provides compelling evidence that cognition may be affected by pollution, as a result of pollution's effect on the respiratory system. Researchers have documented that short-term acute exposure to particulate matter leads to increased risk of illness including heart disease, stroke, and lung cancer, in addition to increased hospitalization rates (Pope et al. 1995, Dockery and Pope 1996; Schlenker and Walker 2011, Chay and Greenstone 2003). Exposure to fine particulate matter is particularly dangerous since these small particles penetrate deep in to the lungs effecting blood flow and oxygen circulation, which may also affect other aspects of human life, and in particular, cognition (Pope and Dockery 2006). Since the brain consumes a large fraction of the body's oxygen needs, any deterioration in oxygen quality can in theory affect cognition (Clark and Sokoloff 1999, Calderón-Garcidueñas et al. 2008). Air pollution can also impact the nervous system, leading to symptoms such as memory disturbance, fatigue and blurred vision (Kampa and Castanas 2008). As a result of these physiological effects, a recent literature has been able to document that pollution significantly lowers labor productivity in a variety of contexts (Graff Zivin and Neidell 2012, Chang et al. 2014).

In a related paper (Lavy, Ebenstein and Roth, 2014), we examine the effect of pollution on cognition in the Israeli context, examining the relationship between test scores and exposure to particulate matter and carbon monoxide. We analyze the impact of pollution in a panel data set where we observe the same student taking multiple exams with different levels of pollution on each day, enabling us to estimate models with student fixed effects. For particulate matter exposure, we find that a one standard deviation increase in the $PM_{2.5}$ AQI value (σ =22.81) on the day of an exam is associated with a .65 point decrease in score, or 2.8% of a standard deviation.[52] We also find that our results are largely driven by poor performance of test takers on extremely polluted days, with an AQI reading above 101 for $PM_{2.5}$ associated with a decline in test score of 1.95 points, or 8.2% of a standard deviation. These results suggest that modest pollution levels have only a marginal impact, but extremely polluted days can have much larger impacts, suggesting a non-linearity in pollution's relationship with cognitive performance.

In this paper, we focus our attention exclusively on particulate matter ($PM_{2.5}$): a complex mixture of solid and liquid microscopic droplets found in the air that consists of various components including acids, metals, dust particles, organic chemicals and allergens. This is a natural choice since it is arguably the most harmful pollutant for human health, and presumably

[52] We find the effect is concentrated on the day of the exam, with insignificant effects on the day before and day after the exam. Results can be found in Ebenstein, Lavy, and Roth (2013).

most likely to affect cognition. We also focus on student-level data instead of exam-level data, since our interest here is understanding how a particular student is affected by variation in their composite *Bagrut* score, which has more relevance for understanding the long-term consequences of the exam.

3.2.2 High-Stakes Examinations in Israel and Abroad

Since the Scholastic Aptitude Test's (SAT) first administration in 1926, it has been taken by millions of test-takers and has been used to rank students applying for college in the United States, and similar tests are used around the globe. However, the use of these tests is extremely controversial. Numerous concerns have been voiced by both popular and academic sources, including allegations of racial bias, arguments that test prep courses give privileged students an unfair advantage, and suggestions that the test places too much emotional stress on students.[53] Nevertheless, the SAT remains a critical component of college admissions in the US and, just as the *Bagrut* does in the Israeli educational system. The great weight placed on an exam given on one particular day has the benefit of being a cost-effective way of comparing students across schools with a similar metric, but may also represent a noisy measure of student quality. Many factors can affect student performance that are unrelated to cognitive ability, including how a student slept the previous night, whether the testing room has a comfortable temperature, and potentially, exposure to ambient air pollution. In light of the great weight placed on test scores in admissions processes at many elite schools, it is worth knowing whether (a) these scores are sensitive to random shocks and (b) whether bad draws have long-run consequences. Since this would be an extremely challenging question to address in the US, where SAT score data is

[53] In 2001, the President of University of California famously threatened to remove the SAT requirement for admission, leading to a re-design of the examination and the introduction of a writing section. However, the writing section later came under fire for rewarding students simply for lengthy essays (Winerip 2005).

fiercely guarded and generally not available for matching to adult outcomes, the Israeli *Bagrut* represents a novel opportunity to examine this question.

All universities and most academic and teachers' colleges require a *Bagrut* certificate for enrollment, while entry requirements for other post-secondary education establishments, such as academic colleges, are lower.[54] For a given field of study, it is typically more difficult to be admitted to a university than to an academic college, since universities are both more prestigious and more heavily subsidized by the government. Students earn a *Bagrut* certificate by passing a series of national exams (*Bagrut* tests) in core and elective subjects following tenth and eleventh grade, and then passing a larger set of exams following twelfth grade. The exam focuses on seven mandatory subjects and one or more elective subjects, allowing us to observe students completing exams with separate grades for each examination and over a range of subjects. About 52 percent of high-school graduates in 2002 and 46 percent of the overall cohort received matriculation certificates, a figure that is publicized in the national media in light of the great significance of passing the *Bagrut* in the Israeli educational system

Students are admitted to university programs on the basis of their average *Bagrut* scores and a separate psychometric examination. Each university ranks applicants according to the same formula, thus producing an index based on a weighted average of the student's average score on all his or her *Bagrut* exams and the psychometric examination. This ranking determines students' eligibility for university admission, and even which major they can choose within the university. Therefore, pollution levels can affect students' university schooling by affecting their probability

[54] The post-secondary education system in Israel consist of eight universities that grant PhDs (as well as other degrees), approximately 50 academic colleges which offer undergraduate degrees (of which a very limited subset which offer masters degrees), and a set of non-university institutions of higher education that confer teaching and vocational certificates. Practical engineering colleges run two-year programs awarding degrees (or certificates) in fields like electronics, computers, and industrial production. An additional two years of study in an engineering school is required in order to complete a BSc in engineering.

of passing *Bagrut* exams, and also by affecting the average score in these exams. The first channel will affect eligibility for university admission while the second will affect which programs (or majors) will be available to the student.

3.2.3 Data

Our data set is generated by combining Israeli test score data with air pollution and meteorological data for 2000-2002, and subsequent higher education and earnings outcomes from 2010-2011. The *Bagrut* exam information and demographic information for each test taker are provided by the Israeli Ministry of Education. These files also contain rich demographic information on the student and the student's family, such as parental education level, number of siblings, country of origin, and ethnicity. For each exam, we also know the exact date of the test and the precise location of the testing site, allowing us to assign pollution measures to each test administration. Our pollution data are taken from files published by the Israeli Ministry of Environmental Protection, which reports daily mean readings of particular matter less than 2.5 microns in width, or $PM_{2.5}$ ($\mu g/m^3$) at 139 monitoring stations throughout Israel for the sample period (see Figure 1.1). Readings are taken at 5 minute intervals and averaged over the course of the day. Each test site, which is at the student's school, is assigned the average pollution reading for all monitoring stations within 2.5 kilometers of the city limits of the school. Since Israeli cities are not very large, we generally are taking readings from stations very close to the schools. While we ideally would have a measure of pollution inside the test room, the air quality inside a test site is presumed to be highly correlated with the ambient reading outdoors (Branis et al. 2005). Schools that had no monitoring station within the city limits or 2.5 kilometers of the city limits were dropped from the sample.[55] These monitoring stations also record temperature and relative humidity, which are also assigned in a similar manner to pollution and are used as control variables. We use the daily average reading of pollution, temperature, and humidity at the monitoring stations in our analysis. The pollution measure is then converted into units of Air Quality Index (AQI) using a formula specified by the US EPA.

Our information on post-secondary enrollment and earnings is taken from administrative records provided by the National Insurance Institute of Israel (NII). In order to facilitate the analysis presented here, the NII Research and Planning Division constructed an extract containing indicators of post-secondary enrollment, the number of years of post-secondary schooling, annual earnings, and number of months employed among all individuals in our study. This file was then merged with our original sample and analyzed at a secure research lab at the NII headquarters in Jerusalem, Israel. Our administrative sample includes information for each student on whether they have ever enrolled in higher education for each of the aforementioned institution types, and the number of years they were enrolled in higher education by institution type. This information is available for all students, and is recorded as of the 2010-2011 academic year. The youngest cohorts in our sample are already 28 years old at this time, implying that even after accounting for compulsory military service, most students who enrolled in post-secondary education, including those who continued on to graduate school, will have graduated by 2010-2011.[56]

The summary statistics for our sample are presented in Table 3.1 in two panels; Panel A reports sample means of our exam-level data, and Panel B reports sample means of our student-level data. The sample is composed of 415,219 examinations taken by 55,873 students at 712

[55] Since Israel's population is densely concentrated in several metropolitan areas, this led to the dropping of less than 5% of schools.
[56] Boys serve for three years in the military and girls for two (longer if they take a commission).

schools throughout Israel. In columns (2) and (3) we stratify the sample by sex, and in columns (4) and (5), we stratify by a measure of achievement known as the *Magen* score. The *Magen* score is calculated using the student's performance over the course of the school-year, and on exam similar to the *Bagrut*, making it a natural candidate for stratifying the sample by student quality.[57] As shown in Table 3.1, for each *Bagrut* examination we observe the exam score, the pollution the day of the exam ($PM_{2.5}$), and the average temperature and humidity that day.[58] The table reveals that students face average pollution levels (AQI) that appear balanced along observables, with similar average readings among boys and girls (59.5 versus 59.9), and among higher/lower achievement students (60.0 versus 59.5). The sample means also reflect that girls perform better than boys, and student with higher *Magen* scores also have higher *Bagrut* scores.

In Panel B, we report our student-level means, which includes demographic information on the student, the education of both parents, and the student's earnings in 2010. Since our analysis of long-term economic outcomes will rely on school fixed effects, it is particularly important that we are able to include this rich set of control variables. The sample means also reveal several interesting patterns, including the higher achievement of girls: roughly 71% of girls receive a matriculation certificate, compared to only 64% of boys. Interestingly, however, girls earn lower earnings than their male counterparts. Boys earn on average 5,531 New Israeli Shekels (NIS) versus 4,699 for girls ($1≈ 3.75NIS). In columns (4) and (5), we observe higher rates of matriculation certification (91% versus 48%) and wages (5,352 versus 4,867) in the group of high achievement students, consistent with our expectations. Almost two thirds (63%) of a high school cohort are enrolled in post-secondary studies, 27% in universities, and 25% in academic colleges. The sample also reveals that we are able to match the entire universe of

[57] The date on which the *Magen* exam is given is unavailable, precluding a direct analysis of these scores.
[58] These variables are not shown due to space considerations but are used in regressions available in the appendix.

student test takers with their long-term outcomes, a particularly desirable feature of our data relative to panel data sets that face attrition.

3.3 Empirical Strategy

Our analysis focuses on student-level data, where we exploit variation across students in their average level of pollution across all their *Bagrut* tests. In this setup, the endogenous regressor is the student's *Bagrut* composite score, which is calculated as the average score across the *Bagrut* examinations. The identification assumption in the 2SLS analysis is that variation in the timing of *Bagrut* exams is not correlated with potential outcomes, after conditioning on a student's school. This is a plausible assumption because dates of national *Bagrut* exams are determined by the Ministry of Education, and students choose their *Bagrut* study program years before the dates of exams are determined. The realization of pollution level on different exam dates is random, and therefore variation in average pollution exposure is also random. In support of this assumption, we provide balancing estimates that demonstrate that, conditioned on locality or school, pollution does not appear to be correlated with observable features of the student.

A likely possibility is that pollution is correlated with time invariant features of a testing location or a particular student. For example, if poorer schools are located in more polluted parts of cities, OLS will likely overstate the causal link between pollution and test scores. Conversely, if schools in denser (and wealthier) cities have more pollution exposure, OLS might understate the true cost of pollution, as it is mitigated by other compensating factors (e.g. tutoring). More generally, endogenous sorting across schools, heterogeneity in avoidance behavior, or measurement error in assigning pollution exposure to individuals will all bias results that do not properly account for unobserved factors correlated with both our outcome of interest and ambient pollution (Moretti and Neidell 2011). In our setup, since we account for time-invariant features of schools and students with fixed effects, the challenge relevant to our estimation is to account for omitted variables that are varying over time but are potentially correlated with pollution and *Bagrut* scores. For example, if weather or traffic the day of the exam is correlated with pollution, our fixed effects models will fail to identify the true effect. In our empirical analysis, we include controls for time-varying factors that could be contemporaneous with pollution, such as daily temperature and relative humidity, but of course it is untestable whether there are factors that are unobserved that are both correlated with pollution and *Bagrut* exam scores. In Lavy et al. (2014), we conduct a rich set of robustness checks and placebo tests documenting the impact of pollution on test performance. [59] In this paper, we are interested in exploiting this relationship to examine the consequences of high-stakes exams.

Our estimation strategy is relatively straightforward. We estimate models relating the average air quality conditions during the examination to the student's composite score. Formally, the first stage model that we estimate is of the following form:

$$(1)\ R_{is} = \beta_0 X_{is} + \beta_1 \overline{POL}_{is} + \beta_2 \overline{Temp}_{is} + \beta_3 \overline{RH}_{is} + S_s + \varepsilon_{is}$$

where R_{is} is the *Bagrut* composite test score of student *i* at school *s*; ; X_{it} is a vector of individual characteristics possibly related to test outcomes, such as parental education: $\overline{POL}_{is}$ is average air pollution exposure of student *i* at school *s* across the examinations; $\overline{Temp}_{is}$ is the mean temperature for student *i* at school *s* across the examinations; $\overline{RH}_{is}$ is the average

[59] In Lavy et al. (2014), we estimate models with higher order polynomial functions of temperature. The results are very similar, which is logical, since most classrooms where the exams are held have air conditioners.

humidity measure of student *i* at school *s* across the examinations; S_s and e_{ist} is an idiosyncratic error term. The second stage equation is as follows:

$$(2)\ O_{is} = \gamma_0 X_{is} + \gamma_1 \hat{R}_{is} + \gamma_2 \overline{Temp}_{is} + \gamma_3 \overline{RH}_{is} + S_s + \varepsilon_{is}$$

where O_{is} are the other long-term outcomes: *Bagrut* matriculation, post-secondary enrollment, post-secondary years of schooling, and monthly earnings, all measured at age 30 and S_s is a school-fixed effect.

3.4 Empirical Results

3.4.1 Estimates of Pollution's Impact on Long-Term Outcomes

In this section, we first examine the reduced form relationship between average $PM_{2.5}$ exposure during the *Bagrut* exams and long-term outcomes. For identification we rely on variation in average pollution exposure across students within school which is driven by the timing of different exams across different subject areas. As previously discussed, this is unlikely to be correlated with student potential outcomes. We empirically support this in balancing exercises presented in Table A3.1, where we demonstrate that, conditioned on city or school, pollution does not appear to be correlated with observable features of the student. We find that the balancing coefficients estimates are of small magnitude and not significantly different than zero. As further evidence in support of our identification strategy, we also note that in results presented in Lavy et al. (2014), the estimated effect of pollution was not systematically sensitive to adding student fixed effects, suggesting that our results are robust to the type of fixed effects included.

In Table 3.2, we present the reduced form effect of average $PM_{2.5}$ on several academic outcomes related to the *Bagrut*. including average score (composite score), passing rates, and proportion of students who receive matriculation certification. In the first row, we report the impact of pollution on a student's *Bagrut* composite score; we estimate that an additional 10 units of $PM_{2.5}$ (AQI) is associated with a 2.66 and a 1.64 unit reduction in a student's composite score in our models with city and school fixed effects respectively, an estimated effect of roughly 13% and 20% of a standard deviation respectively (σ=23.7). In rows 2-4, we examine how pollution affects students who are closer to the margin in terms of continuing on to higher education. In particular, in our preferred models with school fixed effects, we find that pollution exposure of an additional 10 units of $PM_{2.5}$ (AQI) raises the total number of failed exams by 0.11, raises the proportion of failed exams by 1.5 percentage points and lowers matriculation certification rates by 3.3 percentage points. Since matriculation certification is required by many elite post-secondary academic institutions in Israel, it is likely that students which suffer a negative shock that lowers their certification probability will ultimately impact their prospects for higher education. In rows 5 and 6, we present the estimated effect of average pollution exposure on two longitudinal educational outcomes: enrollment in post-secondary institution (1=yes), and years of postsecondary schooling attained. Indeed, we find that enrollment rates in higher education decline by 3.1 percentage points and schooling declines by 0.15 years when a student is exposed to an additional 10 units of $PM_{2.5}$ (AQI). All estimates are statistically significant at the 5% level, and suggest that taking *Bagrut* exams in highly polluted days can have long-lasting effects on schooling attainment. In row 7, we present the reduced form effect of average $PM_{2.5}$ on average monthly earnings. In our preferred specification with school fixed effects in column 3, we estimate that a student exposed during the *Bagrut* exams dates to an additional 10 units of $PM_{2.5}$ (AQI) is associated with an average monthly earnings decline at age 28 of 109 shekels ($30). This estimate is also precisely estimated, with a T statistic greater than three.

3.4.2 The Long-Term Consequences of the *Bagrut*

In Table 3.3, we use as a first-stage relationship the highly significant reduced form effect of pollution on a student's *Bagrut* composite score to examine the long-term consequences of the examination. In Panel A, we estimate the return to an additional point on the *Bagrut* composite score using 2SLS. In the first row, we reproduce the relationship between the *Bagrut* composite score and $PM_{2.5}$ shown in Table 3.2 that is used here as our first-stage. Exploiting the relationship between scores and pollution, we find using 2SLS that an additional point is worth 45 shekels in monthly earnings in models with city fixed effects, and 66 shekels in our preferred specification including school fixed effects. Since the standard deviation of the *Bagrut* composite score is roughly 23 points, these estimates imply that even small deviations from a student's "average" score can have significant consequences on adult income.

In Panel B, we use the first-stage relationship between pollution and the *Bagrut* composite score to examine the mechanisms underlying the strong relationship between scores and earnings. Since the *Bagrut* composite score is an important factor in gaining admission into lucrative courses of study, it is logical to examine whether the instrumented score is correlated with subsequent educational outcomes. As shown in Panel B, we find that each additional instrumented point increases the probability of receiving a matriculation certificate by 2 percentage points, enrollment rates in post-secondary schooling by 1.9 percentage points, and post-secondary educational attainment by .092 years. This indicates that an additional *Bagrut* point has a tremendous economic value to a student, and even random variation in scores can have important consequences for a student's future attainment of post-secondary schooling.

In Panel C, we exploit the relationship between pollution exposure and the *Bagrut* composite score to estimate the return to an additional year of post-secondary schooling. It is worth noting that this strategy does not identify 'cleanly' the rate of return to schooling since the *Bagrut* score can directly affect earnings, and therefore its omission might violate the exclusion restriction. However, as way of benchmarking our results, we wish to compute the return to education and compare our estimates to those found in the existing literature. Treating post-secondary schooling as the endogenous regressor and $PM_{2.5}$ as the instrument, we estimate using 2SLS that each additional year of post-secondary schooling is worth 707 ($191) shekels. This estimate implies a rate of return to college education of 14%, which is somewhat higher in comparison with recent estimates in Israel and elsewhere. For example, Angrist and Chen (2011) exploit variation in veteran status and the GI Bill to estimate a return to education of roughly 9%.

While our IV approach will not be valid if pollution exposure on exam dates leads to a permanent diminution of intellectual ability, in light of the highly temporary nature of pollution in Israel, such as sandstorms, we think it is unlikely that our effects are picking up such an effect. In Lavy et al. (2014), we present evidence that exposure to pollution in our data is only related to temporary variation in performance. We find that a student's average pollution exposure during the *Bagrut* period (May-July) in 11th grade has no correlation with his or her average *Bagrut* scores on exams taken in May-July of 12th grade, suggesting that the effects we estimate are capturing the consequences of short-term random shocks, rather than reflecting long-term cognitive diminution.

We examine possible mechanisms for our results by examining how pollution affects the probability of a student matriculating at different types of post-secondary institutions. If our results are operating through a mechanism in which the *Bagrut* is a gatekeeper to lucrative occupations, we should find that our results are driven by large estimated effects for universities, and milder effects for academic colleges. In fact, it may be that for students who attend technical schools, there is no financial value to passing the *Bagrut*, insofar as they purse a profession of a technical nature. This could similarly be true for students planning to be small business managers, which is common in Israel, especially among the Israeli-Arab population, who generally have more limited access to lucrative professions.[60] As reported in Table A3.3, this is indeed the case, with our effects significant and negative only for the probability of attending a university. In fact, interestingly, the impact of pollution is *positive* (though imprecisely measured) for the less competitive programs, such as teacher's colleges and semi-engineering programs, possibly due to students being shifted out of universities or academic colleges and into these less selective programs.

3.4.3 Heterogeneity

In this section, we examine heterogeneity in the relationship between the average *Bagrut* score and long-term schooling and economic outcomes using the variation generated by pollution. We stratify our data by comparing three groups: boys and girls, academically stronger and weaker students, and students from high and low socioeconomic background. We begin with an analysis of the first stage relationship between pollution and test scores in the three sub-samples. Our motivation for this stratification is twofold. First, we have a prior that asthma rates

[60] Willis and Rosen (1979) find that, in a sample of World War II veterans, comparative advantage dictates whether people sort into higher education. This is consistent with our findings, which indicate that there is almost no marginal value of academic achievement for the lower ability students.

are higher among boys and among those of lower SES. Insofar as we observe larger effects for these groups, it sheds light on a possible mechanism for our finding, and implies that our result is being driven partly by a physiological mechanism. Second, in terms of fairness, it is worth considering how these exams affect different students. Since these exams are often the gatekeeper for prized occupations in Israel, it is worth investigating how different students are able to capitalize on these forms of achievement.

3.4.4 Heterogeneity in the Responsiveness to Pollution

We first examine whether there is heterogeneity in the relationship between scores and pollution by building on a set of stylized facts regarding which groups would be most sensitive to poor air quality, taken from an extensive medical literature. First, Israeli boys are more likely to be asthmatic than Israeli girls (Laor et al. 1993). Second, children of lower economic status are known to have higher rates of asthma and respiratory illnesses (Eriksson et al. 2006, Basagana et al. 2004). Third, Laor et al. (1993) also found that *Ashdkenazic* Jews (ethnic origin from America and Europe) have 63% higher incidence of these illnesses than *Sephardim* (ethnic origin from Africa and Asia). This gives a rich set of potential comparisons for gauging whether asthma (or other respiratory illnesses) is a mechanism for the observed reduced form relationship between pollution and exam outcomes.

We estimate models in the same manner as those reported in Table 3.2, based on panel data at the exam level, separately by sex, student quality, and student background (SES). These results are presented in Table A3.4. The evidence in Panel A suggest that boys are between 2 and 4 times more sensitive to pollution than girls. We posit that the difference could be partly generated by the different asthma rates in these cohorts. Another possibility is that male students are more likely to be affected by small cognitive decline and distraction, consistent with higher rates of Attention Deficit Disorder in males (Biederman et al. 2002). As shown in Panel B, a similar pattern is found when comparing between academically weaker and stronger students. When we stratify the students by whether their *Magen* score is above or below the median, our estimated treatment effects for $PM_{2.5}$ are more than two times larger among those classified as low achievers. The results also indicate a similarity between results estimated with school and student fixed effects, which is reassuring since our results on earnings are estimated at the student level and rely on school fixed effects. Lastly, in Panel C, we stratify the sample by SES. It may be that poorer families are more affected by air pollution as well, due to lower ability to engage in compensating behavior (Neidell 2004). Poorer children also have higher incidence of asthma (Basagana et al. 2004, Eriksson et al. 2006). We proxy SES with the student's father's education, with high SES being students with a father above the median value in our sample. We again find evidence in support of an interpretation that the relationship between pollution and exam performance is driven by a physiological mechanism.[61]

3.4.5 Heterogeneity in the Long-Term Consequences of the *Bagrut*

In this section, we examine the long term schooling and economic consequences of the *Bagrut* for different sub-populations in Israel. We repeat the group breakdowns used earlier, dstratifying students by sex, quality, and socio-economic background. In Table 3.4, columns 1

[61] In Lavy, Ebenstein and Roth (2013), we examine the relationship between air pollution exposure and the probability of failing a *Bagrut* exam for each of these sub-populations. The results indicate that boys are more sensitive to $PM_{2.5}$ than girls, lower quality students are more sensitive than stronger students, and students from lower SES backgrounds are more sensitive than those from higher SES backgrounds. For example, raising the fraction of days with very polluted air by 10 percentage points is associated with a .57 percentage point increase for boys in the chance of failing a particular *Bagrut* in models with student fixed effects. Girls appear largely unaffected, with the increased chance of not passing being statistically indistinguishable from zero. The gap is even more striking for student with low *Magen* scores: a 10 percentage point increase in the fraction of days with very polluted air is associated with a .59 percentage point increase in failure probability.

and 2 of Panel A, we present estimates of the return to an additional point on the *Bagrut* using 2SLS, where the *Bagrut* composite score is treated as the endogenous regressor and $PM_{2.5}$ is the instrument. Our results by student sex are reported in columns 1 and 2, and indicate that the return to an additional point is roughly 60% higher for boys than girls: 78 shekels vs 59 shekels ($21 vs $16). One explanation is that women choose less financially rewarding fields of study than men, even when they have similar qualifications, as a result of gender attitudes in Israel. It is also worth noting that although female labor force participation rates are relatively similar to the US, Israeli women have much higher fertility than their American counterparts.[62] This may lead Israeli women to choose less lucrative professions than men and often work part time, which would be reflected in a lower payoff per additional year of higher education. In our context, this is plausible, since many Israelis work in government jobs which are lower-wage, offer more flexible work schedules, and have generous maternity leave policies. A second explanation is that this is driven by discrimination against women in the labor market, resulting in a lower payoff to an additional year of schooling.

In columns 3 and 4, we find larger returns to a point among higher achievement students. Specifically, stronger students experience a 124 shekel return to each point, compared to only a 80 shekel return among lower quality students ($34 vs $22). We offer two explanations for this finding. First, it may be related to the instrument we are using; insofar as our estimate is a local average treatment effect where the disturbance to a student's true potential is relatively small, the estimated return to an additional point on the *Bagrut* will be larger among those who could participate in lucrative occupations. For weaker students, pollution is not affecting their already-

[62] Average fertility rates in Israel are 3.0, roughly 50% higher than the US rate of 2.0 (World Bank, 2010). However, employment rates are relatively similar. Among women 25-45, the employment rates among Israeli men and women were 80% and 61% respectively (Israel 1995 census), compared to rates in the US of 86% and 69% (US 2010 census).

low chance of being accepted into a very lucrative profession. A second explanation is that there are heterogeneous returns to different types of higher education. So, for example, the return to an additional year of studying economics might be different than the return to philosophy. Since Israeli majors have different standards for admission, with humanities having lower standards generally, our estimates may be picking up the differences between the return to different majors, which provide an avenue to different occupations. Unfortunately, our data do not contain information on area of study, precluding further examination into this hypothesis.

In columns 5 and 6, we observe very large differences between students of high and low SES. The return to an additional point is 105 shekels ($28) among high SES, and roughly half that amount for low SES (56 shekels or $15). Interestingly, this gap is even larger than when we had previously split the sample by *Magen* score in columns 1 and 2. This suggests that coming from a wealthier background raises the return to education significantly, and in a more dramatic way than even stratifying by student quality. One possible explanation is that parental income enables students to undertake longer and more costly academic paths, but results in them landing ultimately in more lucrative positions. Having a non-binding funding constraint could be a partial explanation for the higher return to higher education. Another explanation is that credentials and connections are complements, so students with greater social capital *and* qualifications can capitalize on their qualifications more than students from less privileged background.

In Panel B of Table 3.4, we examine the mechanisms for the aforementioned results by estimating 2SLS models where we instrument for a point with pollution and treat the *Bagrut* composite score as our endogenous regressor. We repeat our earlier analysis performed on the overall sample, and examine 3 channels through which *Bagrut* scores may influence long run economic outcomes: by affecting the probability of receiving a matriculation certificate, by affecting enrollment rates in post-secondary institutions, and through its effect on total completed post-secondary education. Interestingly, we find that girls, weaker students, and lower SES background students are *more* affected by each additional instrumented point on the *Bagrut* than boys, stronger students, or higher SES background students. We interpret this as evidence that the stakes of each point is higher for students with lower labor-force attachment or less economic advantage – not necessarily economically in terms of the consequences for wages, but in terms of their likelihood of pursuing post-secondary education.

For example, if male students are committed to participating in a lucrative profession, they will perhaps be less dissuaded by a poor score; in Israel, students can attend a lower-ranked school in their field of study of choice, if they are denied admission in the top programs in Tel Aviv or Jerusalem. Likewise, stronger students and students from high SES background may proceed with their intended course of study in spite of a bad score. Greater commitment to the labor force or greater access financial resources may result in matriculation at campuses in other cities, such as Be'er Sheva, where students generally cannot live at home and need to pay for dormitories but often have lower *Bagrut* score thresholds for admission. Having sufficient economic resources could also facilitate a student retaking the exam in a subsequent year, lowering the stakes of a single bad outcome for students from privileged backgrounds.[63] Our analysis of mechanisms highlights the disruptive effect of a poor *Bagrut* outcome on female and weaker/low SES students, and is further evidence that reliance on the *Bagrut* has questionable properties in terms of supporting a meritocratic environment.

[63] Vigdor and Clotfelter found this to be important in the US context, where students from wealthier backgrounds are more likely to retake the SAT (2003). However, since Israeli high-school graduates immediately begin a period of military service (3 years for boys and 2 years for girls), retaking the exam is only possible with a long delay. Therefore, retaking the exam is relatively uncommon but could explain in part the weaker relationship between instrumented *Bagrut* scores and post-secondary education for students from high SES.

3.4.6 Heterogeneity in the Returns to Higher Education

In Table 3.5, we examine heterogeneity in the estimated return to education by sub-population in Israel using 2SLS, with pollution again as our instrument for *Bagrut* composite scores. Note that since pollution affects scores as well, this will not satisfy the exclusion restriction, but is worth exploring nonetheless to assess the economic magnitude of our estimated effects. In Panel A, where we stratify the sample by sex, we estimate that an additional 10 units of $PM_{2.5}$ reduces male and female post-secondary schooling by 0.17 and 0.14 years respectively. We estimate that the return to an additional year of schooling is 888 shekels and 564 shekels respectively ($240 versus $152), suggesting that male students are more able to capitalize on post-secondary education, possibly due to the choice of more lucrative majors and professions, discrimination in the labor force, or due to their stronger labor-force attachment. We also find that stronger students are able to capitalize more from higher education: the wage return to post-secondary schooling is nearly twice as high among stronger students, with each year increasing wages by 1,131 shekels per month for strong students and only by 698 for weaker students. This pattern is even more extreme when we consider students stratified by SES: an additional year is worth 1,264 shekels to a student of high SES background, more than twice the return to low SES students (580 shekels). Similar to our discussion of the return to a point on the *Bagrut*, this highlights the interplay between achievement and status: the results indicate that the return to post-secondary education is largest among those most able to leverage this achievement, highlighting an additional avenue by which high stakes examinations can affect the wage distribution and wage inequality.

3.5 Conclusion

This paper has examined the relationship between transitory shocks to performance in high stakes exams and their long term consequences for determining college schooling attainment and earnings. We exploited variation in ambient air pollution during high-stakes examinations as an instrumental variable, and demonstrated that pollution affects student test scores on very important high-school exit exams, which in turn affects post-secondary schooling and long run earnings. Our analysis consisted of two parts. First, using a large sample of Israeli high-school *Bagrut* examinations (2000-2002), we presented evidence that there is a robust negative relationship between academic outcomes and random fluctuations in ambient pollution concentrations. Among Israeli sub-populations with higher rates of asthma and respiratory illnesses, our estimated treatment effects for $PM_{2.5}$ are larger, suggesting that physiological impairment is a potential mechanism for our findings. Second, using matched data between students and their adult earnings, we find that ambient pollution exposure during the *Bagrut* has long-term consequences; students exposed to high levels of pollution during the *Bagrut* are less likely to receive a matriculation certificate, have fewer years of post-secondary schooling, and have lower wages when they are observed 8-10 years after high school graduation.

We argue that this is evidence that placing so much weight on *Bagrut* scores may reduce welfare for unlucky students and lead to allocative inefficiencies. As a result this will negatively affect the overall economy, as the mis-ranking of students due to variability in pollution exposure could result in poor assignment of workers to different occupations and reduce labor productivity. Our findings also suggest that the exams may be generating random variation in people's opportunities, consistent with recent concerns voiced by officials in the US regarding the reliance of the SATs for college admissions (Lewin 2014).

Table 3.1

Summary Statistics: Particulate Matter Exposure and Israeli *Bagrut* Scores

Variable	All (1)	By Sex: Boys (2)	By Sex: Girls (3)	By *Magen* Score (Course Grade[1]): Low Scores (4)	By *Magen* Score (Course Grade[1]): High Scores (5)
Panel A: Exam-Level Data					
Pollution Measures					
$PM_{2.5}$	21.05	20.89	21.18	21.15	20.96
($\mu g/m^3$)	(10.86)	(10.57)	(11.10)	(10.88)	(10.87)
$PM_{2.5}$	59.74	59.47	59.98	60.01	59.51
(AQI Index)	(22.81)	(22.50)	(23.08)	(22.89)	(22.75)
$PM_{2.5}$	0.05	0.05	0.05	0.05	0.05
(AQI $\geq$101)	(0.21)	(0.21)	(0.22)	(0.22)	(0.21)
Examination Outcomes					
Bagrut Exam Score	70.76	68.91	72.33	53.22	77.10
(1-100 points)	(23.74)	(24.86)	(22.64)	(30.69)	(22.18)
Failed a *Bagrut* Exam	0.19	0.21	0.17	0.33	0.04
(1=yes)	(0.39)	(0.41)	(0.37)	(0.47)	(0.19)
Magen Score	75.45	73.27	77.30	64.09	86.93
(1-100 points)	(21.37)	(22.50)	(20.19)	(23.25)	(10.47)
Climate Controls					
Temperature	23.81	23.81	23.82	23.84	23.83
(celsius)	(2.61)	(2.61)	(2.62)	(2.66)	(2.50)
Relative Humidity	50.90	50.86	50.94	50.98	50.95
(percent saturation)	(14.71)	(14.52)	(14.87)	(15.08)	(14.35)
Observations	415,219	190,410	224,809	206,571	204,527
Panel B: Student-Level Data					
Demographic Information					
Mother's Education	11.44	11.60	11.30	10.79	12.08
(years)	(5.04)	(5.09)	(5.00)	(4.87)	(5.13)
Father's Education	11.62	11.83	11.44	10.85	12.39
(years)	(5.03)	(5.02)	(5.03)	(4.84)	(5.10)
Number of Siblings	2.02	1.95	2.07	2.03	2.00
	(1.58)	(1.49)	(1.65)	(1.61)	(1.55)
Bagrut Outcomes and Matriculation Certification Rates					
Bagrut Composite Score	70.76	68.91	72.33	53.22	77.10
	(23.74)	(24.86)	(22.64)	(30.69)	(22.18)

Matriculation Certification Rate	0.68 (0.47)	0.64 (0.48)	0.71 (0.45)	0.48 (0.50)	0.91 (0.28)
Post-Secondary Enrollment Rates					
Any Post-Secondary	0.631	0.602	0.656	0.475	0.821
University	0.274	0.258	0.289	0.115	0.469
Academic Colleges	0.248	0.253	0.244	0.244	0.253
Teacher & Semi-eng.	0.070	0.063	0.076	0.078	0.059
Other[2]	0.046	0.036	0.055	0.046	0.047
Post-Secondary Schooling in Years					
Any Post-Secondary	2.25 (2.15)	2.05 (2.10)	2.42 (2.18)	1.45 (1.86)	3.23 (2.08)
University	1.03 (1.90)	0.95 (1.83)	1.10 (1.95)	0.35 (1.13)	1.85 (2.28)
Academic Colleges	0.83 (1.47)	0.80 (1.44)	0.85 (1.50)	0.73 (1.38)	0.95 (1.57)
Teachers & Semi-engineering	0.26 (0.87)	0.21 (0.68)	0.31 (1.00)	0.25 (0.82)	0.27 (0.92)
Adult Earnings					
Monthly Wages ([3]NIS 2010)	5,084 (4,515)	5,531 (5,198)	4,699 (3,788)	4,867 (4,053)	5,352 (5,013)
Observations	55,796	26,158	29,638	30,668	25,128

Notes : Standard deviations are in parentheses. In Panel A, each observation represents a *Bagrut* exam. The measure of pollution is particulate matter smaller than 2.5 microns, or $PM_{2.5}$. The AQI value for each reading is calculated from a formula that converts micrograms (μg/m3) into a 1-500 index value. Relative humidity is the amount of moisture in the air as a share of what the air can hold at that temperature. [1]The low and high subsamples were based on being above or below the median of a student's average *Magen* score over all subjects. In Panel B, each observation represents a student. The pollution and climate controls are averages over a student's exams. Receiving a matriculation certificate is determined by a combination of the student's average *Bagrut* and *Magen* score. [2]The other programs include technical schools, non-academic colleges, and smaller schools. [3]Wages are repoted in monthly New Israeli Shekels ($1=3.6 NIS) and are taken for 2010 from the students who took Bagrut examinations between 2000 and 2002. The schooling and wage outcomes were made available by the Israeli National Insurance Institute (Bituach Leumi). Each student's record contains whether they matriculated at a post-secondary institution, and the number of years they were enrolled at the institution.

Table 3.2

Reduced Form Effect of Particulate Matter On Post-Secondary Education and Adult Earnings

	Pooled OLS	Fixed Effects	
	Controls	City	School
	(1)	(2)	(3)
Bagrut Composite Score	-0.67	-2.66	-1.64
	(0.08)	(0.13)	(0.18)
Number of Bagrut Failures	0.081	0.197	0.106
	(0.084)	(0.011)	(0.020)
Proportion of Bagrut Failures	0.008	0.027	0.015
	(0.001)	(0.002)	(0.003)
Matriculation Certification	-0.023	-0.053	-0.033
	(0.002)	(0.003)	(0.005)
Enrolled in Post Secondary Institution (1=yes)	-0.009	-0.050	-0.031
	(0.002)	(0.003)	(0.004)
Completed Years of Post-secondary Education	-0.067	-0.236	-0.152
	(0.009)	(0.013)	(0.018)
Average Monthly Earnings (NIS)	-155	-120	-109
	(33)	(33)	(34)

Notes : Each cell in the table represents a separate regression. The table reports the relationship between average $PM_{2.5}$(AQI) during the *Bagrut* and the listed outcome, estimated using the student-level sample described in Table 3.1. All regressions include suppressed controls for average temperature and humidity during the *Bagrut* , mother's and father's years of schooling, sex, and age in 2010. The coefficients are reported per 10 units of AQI. Standard errors are clustered at the school level, are heteroskedastic-consistent, and are reported below the coefficients in parentheses.

Table 3.3

The Economic and Academic Return to the *Bagrut* Composite Score

	Pooled OLS	Fixed Effects	
	Controls	City	School
	(1)	(2)	(3)
Panel A: Effect of the Bagrut Composite Score on Adult Earnings using $PM_{2.5}$ (AQI) as an IV			
First Stage	-6.66	-26.57	-16.38
	(0.84)	(1.27)	(1.85)
Reduced Form	-1528	-1203	-1073
	(324)	(327)	(341)
2SLS	229	45	66
	(147)	(13)	(21)
Panel B: Effect of the Bagrut Composite Score on Follow Up Academic Outcomes using $PM_{2.5}$ (AQI) as an IV			
Matriculation Certification	0.034	0.020	0.020
	(0.011)	(0.002)	(0.002)
Enrolled in Post Secondary Institution (1=yes)	0.016 (0.006)	0.019 (0.002)	0.019 (0.002)
Completed Years of Post-secondary Education	0.105 (0.026)	0.089 (0.006)	0.092 (0.009)
Panel C: Estimated Return to Post-Secondary Education using $PM_{2.5}$ (AQI) as an IV			
First Stage	-0.67	-2.36	-1.52
	(0.09)	(0.13)	(0.18)
Reduced Form	-1548	-1199	-1093
	(326)	(331)	(344)
2SLS	2278	509	707
	(1343)	(139)	(219)

Notes : Each cell in the table represents a separate regression. The regressions are estimated in the same manner as those reported in Table 3.2. In Panel A, we present 2SLS models of the relationship between the *Bagrut* Composite Score and Adult Earnings using $PM_{2.5}$(AQI) as an instrumental variable. In Panel B, we present 2SLS models of the relationship between *Bagrut* Composite Score and other academic outcomes, using $PM_{2.5}$(AQI) as an instrumental variable. In Panel C, we estimate the implied return to post-secondary schooling using $PM_{2.5}$(AQI) as the instrument. Standard errors are clustered at the school level, are heteroskedastic-consistent, and are reported below the coefficients in parentheses.

Table 3.4

Heterogeneity in the Economic and Academic Return to the *Bagrut* Composite Score

	By Sex		By Student Quality		By Socio-Economic Status	
	Boys	Girls	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effect of the Bagrut Composite Score on Adults Earnings using PM $_{2.5}$ (AQI) as an IV						
First Stage	-19.72	-13.69	-6.61	-12.15	-13.22	-21.10
	(2.65)	(1.94)	(1.84)	(1.65)	(1.98)	(2.52)
Reduced Form	-1,540	-810	-530	-1,510	-747	-2,205
	(495)	(421)	(352)	(622)	(288)	(734)
2SLS	78	59	80	124	56	105
	(27)	(30)	(58)	(51)	(24)	(32)
Panel B: Effect of the Bagrut Composite Score on Follow up Academic Outcomes						
Matriculation Certification	0.017	0.025	0.032	0.011	0.023	0.015
	(0.002)	(0.003)	(0.008)	(0.004)	(0.003)	(0.002)
Enrolled in Post Secondary Institution (1=yes)	0.018	0.020	0.029	0.019	0.021	0.013
	(0.002)	(0.003)	(0.008)	(0.004)	(0.003)	(0.002)
Completed Years of Post-secondary Education	0.087	0.108	0.108	0.118	0.096	0.085
	(0.010)	(0.014)	(0.030)	(0.015)	(0.013)	(0.010)

Notes : Each cell in the table represents a separate regression. The regressions are estimated in the same manner as those reported in Table 3.2. Student quality is determined by whether the student's average Magen score was above or below the median. High SES is defined as children whose father was above the median level of education.

Table 3.5

Heterogeneity in the Estimated Return to Post-Secondary Education

	By Sex		By Student Quality		By Socio-Economic Status	
	Boys	Girls	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)
First Stage	-0.17	-0.14	-0.08	-0.13	-0.13	-0.17
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Reduced Form	-1,559	-839	-521	-1,568	-743	-2,251
	(498)	(428)	(355)	(617)	(288)	(747)
2SLS	888	564	698	1,131	580	1,264
	(296)	(265)	(484)	(454)	(235)	(405)

Notes : Each cell in the table represents a separate regression. Standard errors are clustered by school. Each estimate is in terms of a single year of additional post-secondary schooling, and the instrument is $PM_{2.5}$(AQI). Student quality is determined by whether the student's average Magen score was above or below the median. High SES is defined as children whose father was above the median level of education. Standard errors are heteroskedastic-consistent and are reported below the coefficients in parentheses.

Table A3.1

Balancing Tests: Assessing the Relationship between Students' Characteristics and Pollution

Variable	Pooled OLS (1)	School Fixed Effects (2)
Female (1=yes)	0.00 (0.00)	0.10 (0.00)
Father's Education	0.10 (1.00)	0.40 (0.50)
Mother's Education	0.30 (1.00)	-0.10 (0.60)
Number of Siblings	0.60 (0.30)	0.30 (0.10)
Ashkenazi (1=yes)	0.00 (0.00)	0.00 (0.00)
Sephardi (1=yes)	0.00 (0.00)	0.00 (0.00)
Father Born in Israel (1=yes)	0.00 (0.00)	0.00 (0.00)
Observations	54,294	54,294

Notes: Each cell in the table represents a separate regression, where the dependent variable is $PM_{2.5}$(AQI) and the independent variable is the covariate listed in the row. The regressions are estimated in the same manners as those presented in Table 3.2.

Table A3.2

Relationship Between Particulate Matter Exposure During Previous Exams and Average *Bagrut* Scores at Conclusion of 12th Grade

	Pooled OLS		Fixed Effects	
	No controls	Controls	City	School
	(1)	(2)	(3)	(4)
Panel A: All Students				
	-0.80	0.90	-0.40	1.70
	(2.90)	(2.80)	(3.50)	(2.10)
Panel B: By Sex				
Boys	-0.90	0.30	-2.40	-0.70
	(3.40)	(3.50)	(4.50)	(2.80)
Girls	-1.20	1.30	0.90	4.00
	(2.80)	(3.00)	(3.60)	(2.40)
Panel C: By Student Quality				
Low Achievement	2.60	3.30	0.20	2.60
Students	(2.50)	(2.50)	(3.30)	(2.30)
High Achievement	1.30	1.40	1.10	2.30
Students	(1.10)	(1.10)	(1.60)	(1.30)
Panel D: By Socio-Economic Status (SES)				
Low SES	-2.10	0.80	1.00	1.30
	(2.90)	(3.00)	(3.50)	(2.30)
High SES	1.10	0.10	-1.30	2.20
	(3.00)	(2.80)	(4.10)	(2.80)

Notes : Each cell in the table represents a separate regression. The regressions are estimated in the same manners as those presented in Table 3.2. Student quality is determined by whether the student's average *Magen* score was above or below the median. High SES is defined as children whose father was above the median level of education. Standard errors are heteroskedastic-consistent and are reported below the coefficients in parentheses.

Table A3.3

Particulate Matter's Impact on Post-Secondary Schooling by Type

	LHS: Enrolled in Post-Secondary Institution (1=yes)		LHS: Completed Years of Post-Secondary Education	
	City (1)	School (2)	City (3)	School (4)
All Post-Secondary Institutions	-0.049 (0.007)	-0.030 (0.004)	-0.235 (0.031)	-0.153 (0.018)
Universities	-0.054 (0.007)	-0.037 (0.004)	-0.222 (0.029)	-0.157 (0.018)
Academic Colleges	-0.009 (0.004)	0.002 (0.003)	-0.041 (0.013)	-0.004 (0.010)
Teacher and Semi-engineering	0.004 (0.003)	0.001 (0.002)	0.012 (0.007)	0.006 (0.005)

Notes : Each cell in the table represents a separate regression. In each regression, the dependent variable is either enrollment (columns 1 and 2) or years of schooling (columns 3 and 4) at the listed academic type. The depdent variable is the average $PM_{2.5}$ (AQI) exposure during the student's *Bagrut* examinations. The regressions are estimated with the same controls as those presented in Table 3.2, and the coefficients are reported per 10 units of $PM_{2.5}$ (AQI). The column title reports whether fixed effects are included at the city or school level. Standard errors are heteroskedastic-consistent and are reported below the coefficients in parentheses.

Table A3.4

Heterogeneity in the Impact of $PM_{2.5}$ on *Bagrut* Test Scores

	$PM_{2.5}$ (AQI Index)		$PM_{2.5}$ (AQI >=101)	
	School	Student	School	Student
	(1)	(2)	(3)	(4)
Panel A: By Sex				
Boys	-0.10	-0.08	-5.33	-4.10
	(0.01)	(0.01)	(0.82)	(0.87)
Girls	-0.04	-0.02	-0.66	-0.38
	(0.01)	(0.01)	(0.80)	(0.83)
Panel B: By Student Quality				
Low Achievement Students	-0.08	-0.06	-3.86	-3.49
	(0.01)	(0.01)	(1.04)	(1.10)
High Achievement Students	-0.03	-0.03	-0.93	-0.76
	(0.01)	(0.01)	(0.57)	(0.68)
Panel C: By Socio-Economic Status (SES)				
Low SES	-0.07	-0.05	-2.76	-2.07
	(0.01)	(0.01)	(0.74)	(0.77)
High SES	-0.06	-0.04	-2.40	-1.66
	(0.01)	(0.01)	(0.86)	(0.93)

Notes: Each cell in the table represents a separate regression. The regressions are estimated in the same manners as those reported in Table 3.2. Student quality is determined by whether the student's average *Magen* score was above or below the median. High SES is composed of children whose father was above the median level of education in the sample. The column title reports whether fixed effects are included at the school or student level. Standard errors are heteroskedastic-consistent and are reported below the coefficients in parentheses.

Conclusion

This thesis examined the relationship between pollution exposure, student exam performance, and long run academic and economic outcomes. The first two chapters demonstrate that exposure to ambient and indoor air pollution has a statistically and economically significant effect on student performance. The last chapter analyzes whether the exogenous variation in scores generated by ambient air pollution has long-term consequences. The analysis suggest that pollution exposure during the exams leads to significant declines in post-secondary education and earnings, indicating that even random variation in test scores can influence a student's academic path and earnings potential.

The results highlight how heavy reliance on noisy signals of student quality can lead to allocative inefficiency. The mis-ranking of students due to variability in pollution exposure could result in poor assignment of workers to different occupations and reduce labor productivity. While it is beyond the scope of this thesis to consider the aggregate efficiency loss associated with the current system, the reduced form evidence suggests that a structural approach could more precisely quantify the costs in foregone productivity due to worker misallocation, and these may be quite large. Furthermore, the results for the *Bagrut* may represent a "lower bound" on the negative consequences of high stakes exams; while the *Bagrut* is given over a series of days, enabling students to recover from a single poor performance, many high stakes exams (e.g. SATs) are administered on a single day, where random factors could materially affect a student’s future. The findings lend empirical support for the concern voiced by officials in the US regarding the reliance of the SATs for college admissions, and suggest that more stable measures of student quality should be given greater weight (Lewin 2014). Policymakers should consider adopting strict standards on exam days in the spirt of fairness to test-takers, and in order to reduce the noise in this important measure of student quality. Finally, the demonstrated link between pollution and cognitive performance also implies that a narrow focus on traditional health outcomes may understate the true cost of pollution as mental acuity is essential to productivity in most professions.

Bibliography

Angrist, J. D., & Chen, S. H. (2011). Schooling and the Vietnam-Era Gl Bill: Evidence from the Draft Lottery. *American Economic Journal: Applied Economics*, 96-118.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, *114*(2), 533-575.

Angus, R. G., Heslegrave, R. J., & Myles, W. S. (1985). Effects of prolonged sleep deprivation, with and without chronic physical exercise, on mood and performance. *Psychophysiology*, *22*(3), 276-282.

Basagaña, X., Sunyer, J., Kogevinas, M., Zock, J. P., Duran-Tauleria, E., Jarvis, D., ... & European Community Respiratory Health Survey. (2004). Socioeconomic status and asthma prevalence in young adults the European Community Respiratory Health Survey. *American Journal of Epidemiology*, *160*(2), 178-188.

Bell, M. L., Levy, J. K., & Lin, Z. (2007). The effect of sandstorms and air pollution on cause-specific hospital admissions in Taipei, Taiwan. *Occupational and environmental medicine*.

Bernstein, J. A., Alexis, N., Bacchus, H., Bernstein, I. L., Fritz, P., Horner, E., & Tarlo, S. M. (2008). The health effects of nonindustrial indoor air pollution. *Journal of Allergy and Clinical Immunology*, *121*(3), 585-591.

Bertrand, M., & Mullainathan, S. (2001). Are CEOs rewarded for luck? The ones without principals are. *Quarterly Journal of Economics*, 901-932.

Biederman, J., Mick, E., Faraone, S. V., Braaten, E., Doyle, A., Spencer, T., & Johnson, M. A. (2002). Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *American Journal of Psychiatry*, *159*(1), 36-42.

Braniš, M., Řezáčová, P., & Domasová, M. (2005). The effect of outdoor air and indoor human activity on mass concentrations of PM 10, PM 2.5, and PM 1 in a classroom. *Environmental Research*, *99*(2), 143-149.

Brown, P. (1995). Race, class, and environmental health: a review and systematization of the literature. *Environmental Research*, *69*(1), 15-30.

Calderón-Garcidueñas, L., Mora-Tiscareño, A., Ontiveros, E., Gómez-Garza, G., Barragán-Mejía, G., Broadway, J., ... & Engle, R. W. (2008). Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs. *Brain and cognition*, *68*(2), 117-127.

Cawley, J., Heckman, J., & Vytlacil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, *8*(4), 419-442.

Chang, T., Zivin, J. S. G., Gross, T., & Neidell, M. J. (2014). *Particulate pollution and the productivity of pear packers* (No. w19944). National Bureau of Economic Research.

Chatzidiakou, L., Mumovic, D., & Summerfield, A. J. (2012). What do we know about indoor air quality in school classrooms? A critical review of the literature. *Intelligent Buildings International*, *4*(4), 228-259.

Chay, K. Y., & Greenstone, M. (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*(3), 1121-1167.

Chay, K. Y., & Greenstone, M. (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy*, *113*(2).

Clarke, D. D., & Sokoloff, L. (1999). Circulation and energy metabolism of the brain. *Basic neurochemistry: molecular, cellular and medical aspects*, *6*, 637-669.

Currie, J., Neidell, M., & Schmieder, J. F. (2009). Air pollution and infant health: Lessons from New Jersey. *Journal of health economics*, *28*(3), 688-703.

Dockery, D., & Pope, A. (1996). Epidemiology of acute health effects: summary of time-series studies. *Particles in our air: Concentrations and health effects*, 123-147.

Dockery, D. W. (2009). Health effects of particulate air pollution. *Annals of epidemiology*, *19*(4), 257-263.

Donaldson, K., Gilmour, M. I., & MacNee, W. (2000). Asthma and PM10. *Respiratory Research*, *1*(1), 12.

Hedlund, U., Eriksson, K., & Rönmark, E. (2006). Socio-economic status is related to incidence of asthma and respiratory symptoms in adults. *European Respiratory Journal*, *28*(2), 303-410.

Frish, R. (2009). The economic returns to schooling in Israel. *Israel Economic Review*, *7*(1), 113-141.

Graff Zivin, J., & Neidell, M. (2012). The Impact of Pollution on Worker Productivity. *American Economic Review*, *102*(7), 3652-73.

Ham, J. C., & Zweig, J. S. Edward Avol. (2011). Pollution, Test Scores and the Distribution of Academic Achievement: Evidence from California Schools 2002-2008. IZA Working Paper.

Jackson, C. K. (2010). A Little Now for a Lot Later A Look at a Texas Advanced Placement Incentive Program. *Journal of Human Resources*, *45*(3), 591-639.

Jarvis, M. J. (1993). Does caffeine intake enhance absolute levels of cognitive performance?. *Psychopharmacology*, *110*(1-2), 45-52.

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, *151*(2), 362-367.

Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., & Engelmann, W. H. (2001). The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of exposure analysis and environmental epidemiology*, *11*(3), 231-252.

Laor, A., Cohen, L., & Danon, Y. L. (1993). Effects of time, sex, ethnic origin, and area of residence on prevalence of asthma in Israeli adolescents. *BMj*, *307*(6908), 841-844.

Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of political Economy*, *110*(6), 1286-1317.

Lavy, V. (2009). Performance Pay and Teachers' Effort, Productivity, and Grading Ethics. *American Economic Review*, *99*(5), 1979-2011.

Lavy, V., Ebenstein, A., & Roth, S. (2014). *The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation* (No. w20648). National Bureau of Economic Research.

Lin, T. Y., Young, L. H., & Wang, C. S. (2001). Spatial variations of ground level ozone concentrations in areas of different scales. *Atmospheric environment*, *35*(33), 5799-5807.

Lewin, Tamar. "A New SAT Aims to Realign With Schoolwork". *The New York Times* 5 March 2014; accessed at: http://www.nytimes.com/2014/03/06/education/major-changes-in-sat-announced-by-college-board.html?_r=0. 10 March 2014.

Madureira, J., Paciência, I., & Fernandes, E. D. O. (2012). Levels and indoor–outdoor relationships of size-specific particulate matter in naturally ventilated Portuguese schools. *Journal of Toxicology and Environmental Health, Part A*, *75*(22-23), 1423-1436.

Mendell, M. J., & Heath, G. A. (2005). Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature. *Indoor air*, *15*(1), 27-52.

Moretti, E., & Neidell, M. (2011). Pollution, health, and avoidance behavior evidence from the ports of Los Angeles. *Journal of human Resources*, *46*(1), 154-175.

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

Hanna, R., & Oliva, P. (2015). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *Journal of Public Economics*, *122*, 68-79.

Oreopoulos, P., & Petronijevic, U. (2013). *Making college worth it: A review of research on the returns to higher education* (No. w19053). National Bureau of Economic Research.

Piantadosi, C. A. (2002). Carbon monoxide poisoning. *New England Journal of Medicine*, *347*(14), 1054-1055.

Pope, C. A., Dockery, D. W., & Schwartz, J. (1995). Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation toxicology*, *7*(1), 1-18.

Pope 3rd, C. A., Bates, D. V., & Raizenne, M. E. (1995). Health effects of particulate air pollution: time for reassessment?. *Environmental health perspectives*, *103*(5), 472.

Pope III, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the air & waste management association*, *56*(6), 709-742.

Pope III, C. A., Ezzati, M., & Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine*, *360*(4), 376-386.

Suglia, S. F., Gryparis, A., Wright, R. O., Schwartz, J., & Wright, R. J. (2008). Association of black carbon with cognition among children in a prospective birth cohort study. *American journal of epidemiology*, *167*(3), 280-286.

Schlenker, W., & Walker, W. R. (2015). *Airports, air pollution, and contemporaneous health*. *Review of Economic Studies, forthcoming*.

Seaton, A., Godden, D., MacNee, W., & Donaldson, K. (1995). Particulate air pollution and acute health effects. *The Lancet*, *345*(8943), 176-178.

Shu, W., Zhang, J., Zeng, X., Zeng, Y., & She, W. (2009). Association of traffic-related air pollution with children's neurobehavioral functions in Quanzhou, China. *Environ Health Perspect*, *117*, 1612-1618.

Son, J. Y., & Bell, M. L. (2013). The relationships between short-term exposure to particulate matter and mortality in Korea: impact of particulate matter exposure metrics for sub-daily exposures. *Environmental Research Letters*, *8*(1), 014015.

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

Suglia, S. F., Gryparis, A., Wright, R. O., Schwartz, J., & Wright, R. J. (2008). Association of black carbon with cognition among children in a prospective birth cohort study. *American journal of epidemiology*, *167*(3), 280-286.

Vigdor, J. L., & Clotfelter, C. T. (2003). Retaking the SAT. *Journal of Human Resources*, *38*(1), 1-33.

Von Wachter, T., & Heisz, A. (2012). The short-and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, *4*(1), 1-29.

Shu, W., Zhang, J., Zeng, X., Zeng, Y., & She, W. (2009). Association of traffic-related air pollution with children's neurobehavioral functions in Quanzhou, China. *Environ Health Perspect*, *117*, 1612-1618.

Weinmayr, G., Romeo, E., De Sario, M., Weiland, S. K., & Forastiere, F. (2010). Short-term effects of PM10 and NO2 on respiratory health among children with asthma or asthma-like symptoms: a systematic review and meta-analysis. *Environ Health Perspect*, *118*(4), 449-57.

WHO (2011). Methods for Monitoring Indoor Air Quality in Schools. World Health Organisation, WHO Regional Publications, Bonn, Germany. European Series.

WHO (2006). Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Copenhagen: WHO Regional Office for Europe.

Winerip, M. (2005). "SAT Essay Test Rewards Length and Ignores Errors". New York Times, May 5th.

Willis, R. J., & Rosen, S. (1979). Education and Self-Selection. *Journal of Political Economy*, *87*(5 pt 2).

Wu, X., Apte, M. G., & Bennett, D. H. (2012). Indoor particle levels in small-and medium-sized commercial buildings in California. *Environmental science & technology*, *46*(22), 12355-12363.

Yang, X., Zhang, Y. P., Chen, D., Chen, W. G., & Wang, R. (2001). Eye irritation caused by formaldehyde as an indoor air pollution-a controlled human exposure experiment. *Biomedical and environmental sciences: BES*, *14*(3), 229-236.