

Avraham Ebenstein, Victor Lavy and Sefi Roth

The long run economic consequences of high-stakes examinations: evidence from transitory variation in pollution

**Article (Published version)
(Refereed)**

Original citation:

Ebenstein, Avraham, Lavy, Victor and Roth, Sefi (2016) The long run economic consequences of high-stakes examinations: evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*, 8 (4). pp. 36-65. ISSN 1945-7782
DOI: [10.1257/app.20150213](https://doi.org/10.1257/app.20150213)

© 2016 by the American Economic Association

This version available at: <http://eprints.lse.ac.uk/65743/>
Available in LSE Research Online: November 2016

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution[†]

By AVRAHAM EBENSTEIN, VICTOR LAVY, AND SEFI ROTH*

Cognitive performance during high-stakes exams can be affected by random disturbances that, even if transitory, may have permanent consequences. We evaluate this hypothesis among Israeli students who took a series of matriculation exams between 2000 and 2002. Exploiting variation across the same student taking multiple exams, we find that transitory $PM_{2.5}$ exposure is associated with a significant decline in student performance. We then examine these students in 2010 and find that $PM_{2.5}$ exposure during exams is negatively associated with postsecondary educational attainment and earnings. The results highlight how reliance on noisy signals of student quality can lead to allocative inefficiency. (JEL I21, I23, I26, J24, J31, Q51, Q53)

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? Cognitive acuity can be affected temporarily by a variety of factors, including the intake of caffeine, nicotine, sleep deprivation, and noise (Jarvis 1993; Angus, Heslegrave, and Myles 1985; and Goriounova and Mansvelder 2012). Insofar as there are permanent consequences to variation induced by completely random shocks to student performance, it suggests that the use of high-stakes testing as a primary method for ranking students may be unfair. Furthermore, aggregate welfare may be reduced by relying too heavily on examinations that provide noisy measures

*Ebenstein: Department of Environmental Economics and Management, Hebrew University of Jerusalem, Mount Scopus, Jerusalem (e-mail: avraham.ebenstein@mail.huji.ac.il); Lavy: Department of Economics, University of Warwick, Coventry, CV4 7AL, United Kingdom and Department of Economics, The Hebrew University, Jerusalem, Israel (e-mail: v.lavy@warwick.ac.uk); Roth: Department of Geography and Environment, London School of Economics and Political Science, London, WC2A 2AE, United Kingdom (e-mail: s.j.roth@lse.ac.uk). This paper merges together Lavy, Ebenstein and Roth's working papers (Lavy, Ebenstein, and Roth 2014a, b). Excellent research assistance was provided by Boaz Abramson, Elior Cohen, Eyal Frank, Michael Freedman, Michal Hodor, Susan Schwartz, and Ben Raven. We thank Ran Abramitzky, Josh Angrist, Sascha Becker, Michael Boskin, Andrea Ichino, Caroline Hoxby, Eric Maurin, Erik Plug, Fabian Waldinger, seminar participants at Hebrew University, Stanford University, University of London Royal Holloway, University of Oslo, RES annual meeting 2014, CAGE May 2014 conference on Education, Human Capital and Labor Market Outcomes and the NBER Labor Studies Fall 2014 conference for useful comments and suggestions. We thank Israel's National Insurance Institute (NII) for allowing restricted access to data in the NII protected research lab. Research of the second author is supported by European Research Council (ERC) Advance Grant No. 323439. The authors declare that they have no relevant or material financial interests that relate to the research described in the paper.

[†]Go to <http://dx.doi.org/10.1257/app.20150213> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

of student quality, since it may lead to poor matching between students and occupations, and an inefficient allocation of labor. In the United States, the continued reliance on the Scholastic Aptitude Tests (SATs) for college admissions has generated a great deal of controversy. 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.¹ 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 (Lewin 2014). Nevertheless, the SAT remains a critical component of college admissions in the United States, and similar tests are used worldwide. In spite of a dearth of evidence regarding the consequences of these tests, they continue to play a crucial role in college admissions, and as a result, may affect long-term schooling and labor market outcomes.

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 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 high-stakes exams in other countries, such as the aforementioned SATs or A-levels in England. In Israel, 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. Furthermore, admission decisions in Israel are based almost entirely on concrete measures of student performance, with no weight assigned to extracurricular activities or student essays. As a consequence, Bagrut scores can affect an individual's entire academic career, and subsequent labor market outcomes.

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.² 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 potential candidate is fluctuation in air pollution that might have an effect on cognitive acuity and test scores, and generate plausibly random variation in a given student's outcome.³ Air pollution has been demonstrated to adversely

¹ In 2001, the president of University of California famously threatened to remove the SAT requirement for admission, leading to a redesign 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).

² 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 subpopulations (Cawley, Heckman, and Vytlačil 2001).

³ In China, parental concern over this has led to complaints to local officials to restrict traffic and other emission sources on the day of China's Higher Education Entrance Examination. According to media sources, officials in Yangzhou monitor and publicize air pollution the day of the exam, and larger cities such as Shanghai and Beijing restrict construction and traffic on the day of the examination. <http://www.bjjs.gov.cn/publish/portal0/tab662/info89715.htm>.

affect human productivity across a variety of tasks (Ham, Zweig, and Avol 2011; Graff Zivin and Neidell 2012; and Chang et al. 2014) and may influence cognitive performance on high-stakes exams. Since students are assigned to test sites without prior knowledge of pollution or the ability to reschedule, pollution is unlikely to be correlated with student quality. If transitory pollution exposure does indeed affect student performance, variation in pollution can be exploited to examine whether the component of a student's score, which is related entirely to luck, affects long-term schooling and economic outcomes.

In this paper, we present empirical evidence that exposure during exams to $PM_{2.5}$ is associated with a decline in student performance on high-stakes examinations, and that the variation in scores induced by pollution has a significant effect on long-term educational attainment and adult wages. Our focus is on 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. This is an almost ideal context for several reasons. First, we are able to access a complete record of all Bagrut exams taken between 2000 and 2002 and the date and location they were given, providing us with a large sample of high-stakes exams in which we can observe test outcomes as well as pollution. Second, Israel's $PM_{2.5}$ levels are highly variable due to a variety of factors, including forest fires and sandstorms that affect countries throughout the Middle East and extend into Europe.⁴ As a result, we are able to exploit short-term episodes of pollution that generate a first-stage relationship between pollution exposure and a student's Bagrut scores, which are plausibly unrelated to student quality. Third, Israel's national registration system allows us to match the universe of students who take the Bagrut with their completed post-high school education and their wages in 2010, after most have entered the labor force. Therefore, we are able to examine both whether short-term pollution affects exam performance, and whether the variation in scores generated by pollution has meaningful economic consequences in the long run.

In the first part of our analysis, we examine the impact of fine particulate matter ($PM_{2.5}$) on exam outcomes across a sample of over 400,000 administrations of the Bagrut. Our identification strategy exploits the fact that students take the Bagrut over several days, which enables us to examine the relationship between $PM_{2.5}$ exposure and scholastic performance across the same student's exams. The identification assumption is that variation in pollution exposure across exams is not correlated with student ability (or other unobserved factors affecting student performance), which is a plausible assumption because the tests are completely compulsory without any opportunity for rescheduling; the dates of national Bagrut exams are determined by the Ministry of Education years before the actual exam; and all students must take their exam at their local high school. This provides a context for analysis in which $PM_{2.5}$ variation across student tests is essentially random.

⁴Note that forest fires are very common in Western states of the United States, and most southern and central European countries are seriously affected by the same sandstorms that originate from the Sahara. See, for example, media coverage of early April 2014 episodes of extremely high pollution levels that extend even into England: <http://www.bbc.com/news/uk-26844425>.

In our preferred specification which includes student fixed effects, we find that a relative to a day with average air quality, a 1 standard deviation increase in the $PM_{2.5}$ (Air Quality Index (AQI)) is associated with a decline in student performance of 0.93 points, or 3.9 percent of a standard deviation ($\sigma_{Bagrut} = 23.74$).⁵ In a set of placebo exercises, we verify that pollution levels on days other than the actual exam are not correlated with performance. We find that the effect is transitory and concentrated on the day of the exam, with no meaningful relationship found between pollution and test scores in the days before or after the exam. The estimated magnitude is larger for boys, weaker students, and students from lower socioeconomic background. In light of the responsiveness of scores to pollution and Israel's periodically high pollution levels, it is likely that some students are materially affected by their good or bad luck by having or not having an extreme pollution event occur on the date of a Bagrut exam.

In the second part of our analysis, we examine the relationship between average pollution exposure during the Bagrut and long-term academic and economic outcomes in the same sample of students observed in 2010. In this part of our analysis, we exploit variation in pollution exposure during Bagrut testing among students at the same school. This variation is considerable, since students take Bagrut exams at the end of tenth, eleventh, and twelfth grades. Furthermore, each student is required to take examinations in several elective subjects, and the dates of exams vary by subject. Therefore, two students at the same school may experience different pollution levels due to being in a different birth cohort or by choosing a different set of elective subjects. This generates significant variation in pollution among students at the same school, enabling us to estimate models with school fixed effects.

Using this design, we present evidence that random variation in pollution exposure during the Bagrut has a long-term impact on both academic and economic outcomes. We estimate that an additional ten units of $PM_{2.5}$ (AQI) exposure across the student's exams is associated with a 1.64 unit decline in a student's Bagrut composite score, a 0.15 decline in years of education at a university, and a 109 Israeli shekels (\$30) decline in monthly salary.⁶ We complement our reduced form results by examining our other academic outcomes using 2SLS, treating the Bagrut composite score as the endogenous regressor and using pollution as our instrument. We find that each additional instrumented point increases postsecondary academic enrollment by 1.9 percentage points, postsecondary education by 0.092 years, and 66 shekels (or 1.3 percent) in wages. Interestingly, we also find there is virtually no effect of pollution on noncompetitive forms of higher education (e.g., technical schools). 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 competitive postsecondary education.

⁵Our results are all presented in terms of $PM_{2.5}$ (AQI). Note that alternatively, we could report our results in cubic micrograms ($\mu\text{g}/\text{m}^3$) of $PM_{2.5}$. Our results are largely unchanged using either metric, as the correlation between $PM_{2.5}$ in $\mu\text{g}/\text{m}^3$ and AQI is 0.9855. We report our results using AQI so they can more easily be interpreted in terms of air quality, where 100 is the World Health Organization standard for unhealthy for sensitive groups.

⁶The standard deviation in average $PM_{2.5}$ (AQI) in our sample is 16.7, so these magnitudes can be considered roughly 60 percent of a standard deviation change in average pollution.

In the last section, we examine heterogeneity in the return to a point on the Bagrut across subpopulations in Israel using our 2SLS strategy. We find that the return to a Bagrut point is larger for boys than for girls (78 shekels versus 59 shekels), for stronger students than weaker students (124 shekels versus 80 shekels), and for higher socioeconomic status students than lower socioeconomic status 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 longer academic programs or professions that require long (and poorly paid) internships, like law or medicine. It is worth noting that the lifetime income effects may ultimately be very different than what we estimate, since our cohort of students are only 28–30 years old in 2010, and are observed relatively early in their careers. Over the course of a worker's career, it is possible that wages will depend more on actual quality and less on signals of quality from academic performance. However, we can conclude that students who took their Bagrut on very polluted days have significantly worse academic and economic outcomes even a decade after the exam. Furthermore, insofar as students are denied access to more lucrative occupations due to a poor Bagrut score, the wage effects may persist over the course of an individual's career.

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. Furthermore, by temporarily lowering the productivity of human capital, high pollution levels may lead to allocative inefficiency. If students with higher human capital are assigned a lower rank than their less qualified peers due to random chance, this may result in an inefficient allocation of workers across occupations and a less productive workforce overall. While high-stakes exams may serve a critical role in enabling comparisons across students at schools with different grading standards, our results suggest that these tests may provide a somewhat noisy measure of student quality, and should therefore be used judiciously. An additional implication is that if these tests are going to be heavily relied upon, students should be given reasonable accommodations for those who wish to retake exams.⁷ Furthermore, our results that pollution can affect performance suggest that significant resources should be directed towards limiting pollution near test sites, rescheduling high-stakes examinations when conditions are particularly severe, or allowing for students to retake exams.

The rest of the paper is laid out as follows. In the first 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 II, we present our data and empirical strategy. In Section III, we examine the impact of pollution on exam outcomes using exam-level data. In Section IV, we explore whether the variation in exam outcomes generated by pollution has long-term consequences on schooling and earnings. We conclude in Section V.

⁷As we will later discuss, Israeli students can only retake the exam years after the course, and so preparing for retaking the exam is extremely expensive, requiring preparatory coursework.

I. Background

A. Air Pollution and Cognitive Performance

Previous research has documented a relationship between short-term exposure to particulate matter and increased risk of illness including heart disease, stroke, and lung cancer (Pope, Bates, and Raizenne 1995; Dockery and Pope 1996; Chay and Greenstone 2003; and Arceo, Hanna, and Oliva 2016). Exposure to fine particulate matter is particularly dangerous since these small particles penetrate deep into the lungs effecting blood flow and oxygen circulation, which may also affect other aspects of human life (Pope and Dockery 2006). Mills et al. (2009) propose two possible mechanisms by which fine particulate matter affects the circulatory system: inhaled particles may provoke an inflammatory response in the lungs (with consequent release of prothrombotic and inflammatory cytokines into the circulation), or particles directly translocate into the circulatory system.⁸ Since the brain consumes a large fraction of the body's oxygen needs, any deterioration in oxygen quality can in theory affect cognitive performance (Clark and Sokoloff 1999; Calderón-Garcidueñas et al. 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 (Ham, Zweig, and Arol 2011; Graff Zivin and Neidell 2012; and Chang et al. 2014). Scholars have also identified that long-term exposure to pollution is associated with decline in cognitive acuity among the elderly (Ailshire and Clarke 2015; Wilker et al. 2015) and during preadolescence (Ham et al. 2011). However, to our knowledge, no previous study has examined how pollution affects short-term cognition as it would relate to high-stakes examination performance. This is a potentially important context for evaluating the link between pollution and cognition in light of the critical nature of these tests to determining access to higher education and higher wage occupations.

B. 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. The great weight placed on such exams 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 these

⁸ It is worth noting that in related work we have documented that exposure to other pollutants, such as CO, also inhibit cognitive function and influence test scores (Lavy, Ebenstein, and Roth 2014a). Ideally, we would be able to separately map out how each pollutant affects test performance. However, we focus on PM_{2.5} since this pollutant is monitored most extensively by the Israel Environmental Protection Agency (EPA) and empirically, CO is not highly correlated with PM_{2.5}, suggesting the bias of focusing exclusively on PM_{2.5} is limited.

scores are sensitive to random shocks and whether bad draws have long-run consequences. Since this would be an extremely challenging question to address in the United States, where SAT score data is fiercely guarded and generally not available for matching to adult outcomes, the Israeli Bagrut represents a novel opportunity to examine this question.

The Bagrut exams take place over a number of days, and are predominantly administered at the conclusion of the academic school year following tenth, eleventh, and twelfth grades.⁹ The exams focus on seven mandatory subjects and one or more elective subjects, and are held at the student's high school without opportunity for rescheduling or changing the testing site. Since students generally take between 8–10 separate exams, there is significant variation in pollution exposure across the same student's different tests, enabling us to estimate models with student fixed effects. Our design is also aided by the fact that retaking Bagrut exams is costly. Since most exams are given at the end of twelfth grade, and Israelis begin a period of compulsory military service (three years for boys and two years for girls) after high school graduation, retaking the exam is only possible for most students several years after the relevant coursework and would require many additional days of testing. Therefore, relative to the SATs (which is given on a single day), the fact that the exam is held over several days provides students a chance to more easily recover from a single bad performance, but makes it more difficult to simply retake the exam, and it is rare that students retake any section of the exam. As such, a negative Bagrut outcome during a student's first attempt is likely to have a significant effect on a student's postsecondary academic options.¹⁰

Passing the Bagrut exams awards a student a Bagrut (matriculation) certificate, which is a prerequisite for study at universities and most academic and teachers' colleges.¹¹ 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 of passing Bagrut exams, and also by affecting the average score on these exams. In summary, the mechanisms by which pollution can affect long-term economic outcomes is through its effect on the probability of pursuing higher education; affecting the type of higher education pursued; and the quality of higher education institution ultimately attended.

⁹A small number of exams are taken near the end of the first term in January (less than 2 percent of our sample).

¹⁰Note that since students may retake the Bagrut in a subsequent year, our estimates of pollution exposure's impact on long-term outcomes should be interpreted as an intent-to-treat measure.

¹¹The postsecondary 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 offer masters degrees), and a set of nonuniversity 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.

II. Evaluating the Consequences of High-Stakes Examinations

A. Data

Our dataset is generated by combining three primary data sources: Israeli test score data from 2000–2002, measures of air pollution and weather from the days of the exams, and completed education and wages for our sample of test takers when they are observed in 2010. The Bagrut exam information and demographic information for each test taker is 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 date of the test and the precise location of the school where the exam is administered, 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 particulate matter less than 2.5 microns in width, or $PM_{2.5}$ ($\mu\text{g}/\text{m}^3$) at 139 monitoring stations throughout Israel for the sample period (see Figure A1 in the online Appendix). Readings are taken at five minute intervals and averaged over the course of the day. Each test site is assigned the average pollution reading on the day of the exam for all monitoring stations within 2.5 kilometers of the city boundary in which the school is contained. 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 and there is also direct evidence that outdoor air quality affects the productivity of those indoors (Braniš, Pavla, and Donasová 2005; Chang et al. 2014). Schools that had no monitoring station within the city limits or 2.5 kilometers of the city limits were dropped from the sample.¹² 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. A histogram of our pollution readings is reported in Figure A2 in the online Appendix.

Our information on postsecondary 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 postsecondary enrollment, the number of years of postsecondary schooling, annual earnings, and number of months employed among all individuals in our study in 2010. The NII was able to successfully locate and match every student in our sample of test takers with their data. The youngest cohorts in our sample are already 28-years-old, implying that even after accounting for compulsory military service, most students who enrolled in

¹²Since Israel's population is densely concentrated in several metropolitan areas, this led to the dropping of less than 5 percent of schools.

postsecondary education, including those who continued on to graduate school, will have graduated by 2010–2011.¹³

The summary statistics for our sample are presented in Table 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 55,796 students at 626 schools throughout Israel between 2000 and 2002. 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 for each exam using the student's performance over the course of the school-year, and on an exam similar to the Bagrut, making its composite average over all exams taken by the student a natural candidate for stratifying the sample by student quality.¹⁴ As shown in Table 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. 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 students with higher average Magen scores also have higher Bagrut scores.

In panel B of Table 1, 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 the impact of pollution on 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 percent of girls receive a matriculation certificate, compared to only 64 percent of boys. Interestingly, however, girls earn lower earnings than their male counterparts. Boys on average earn 5,531 New Israeli shekels (NIS) versus 4,699 for girls ($\$1 \approx 3.75\text{NIS}$). In columns 4 and 5, we observe higher rates of matriculation certification (91 percent versus 48 percent) and wages (5,352 versus 4,867) in the group of high achievement students, consistent with our expectations. Almost two-thirds (63 percent) of the students enrolled in postsecondary studies; 27 percent enrolled in universities, and 25 percent enrolled in academic colleges. Note that we are able to match the entire universe of student test takers with their long-term outcomes, a particularly desirable feature of our data relative to panel datasets that face attrition.

B. Empirical Strategy I: Examination Performance and $PM_{2.5}$ Exposure

In the first section of our analysis, we examine the partial correlation between $PM_{2.5}$ and test scores in our sample of exam-level data. For identification, we 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.

¹³ Boys serve for three years in the military and girls for two (longer if they take a commission).

¹⁴ The date on which the Magen exam is given is usually up to a few weeks before the Bagrut, exam but the exact date is unavailable, precluding a direct analysis of these scores.

TABLE 1—SUMMARY STATISTICS: PARTICULATE MATTER EXPOSURE AND ISRAELI BAGRUT SCORES

Variable	All (1)	By sex		By Magen score (course grade ^a)	
		Boys (2)	Girls (3)	Low scores (4)	High scores (5)
<i>Panel A. Exam-level data</i>					
Pollution measures					
PM _{2.5} (μg/m ³)	21.05 (10.86)	20.89 (10.57)	21.18 (11.10)	21.15 (10.88)	20.96 (10.87)
PM _{2.5} (AQI)	59.74 (22.81)	59.47 (22.50)	59.98 (23.08)	60.01 (22.89)	59.51 (22.75)
Examination outcomes					
Bagrut exam score (1–100 points)	70.76 (23.74)	68.91 (24.86)	72.33 (22.64)	53.22 (30.69)	77.10 (22.18)
Failed Bagrut exam (1 = yes)	0.19 (0.39)	0.21 (0.41)	0.17 (0.37)	0.33 (0.47)	0.04 (0.19)
Magen score (1–100 points)	75.45 (21.37)	73.27 (22.50)	77.30 (20.19)	64.09 (23.25)	86.93 (10.47)
Climate controls					
Temperature (Celsius)	23.81 (2.61)	23.81 (2.61)	23.82 (2.62)	23.84 (2.66)	23.83 (2.50)
Relative humidity	69.86 (14.71)	70.01 (14.52)	69.74 (14.87)	69.83 (15.07)	69.90 (14.35)
Observations	415,219	190,410	224,809	206,571	204,527

(continued)

Since we observe the students taking multiple exams, we can include student fixed effects. Formally, the models we estimate are of the following form:

$$(1) R_{ist} = \beta PM_{st} + f(Temp_{st}, RH_{st}) + X_{it}\Pi + C_t + M_t + DOW_t + L_l + I_i + \varepsilon_{ist},$$

where R_{ist} is the test score (out of 100 points) of student i at school s at time t ; PM_{st} is our measure of air pollution (PM_{2.5}) at school s at time t , which is measured in units of AQI; $TEMP_{st}$ is the mean temperature at school s at time t in degrees Celsius; RH_{st} is the relative humidity measure at school s at time t ;¹⁵ X_{it} is a vector of observable individual characteristics possibly related to test outcomes, in which we include parental education in years and a dummy for sex; C_t , M_t , DOW_t , and L_t are cohort, month, day of the week, and exam proficiency level fixed effects, respectively; I_t is our fixed effect for the individual; and ε_{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, and in specifications with individual fixed effects our individual-level controls are obviously dropped.

The key identifying assumption for inferring a causal relationship between pollution and test scores estimated by equation (1), β , is that unobserved determinants

¹⁵Relative humidity is defined as the ratio of the water vapor density (mass per unit volume) to the saturation water vapor density expressed in percent. In the empirical analysis, we include linear and quadratic terms in both temperature and relative humidity, and linear and quadratic interaction terms of the two variables.

TABLE 1—SUMMARY STATISTICS: PARTICULATE MATTER EXPOSURE AND ISRAELI BAGRUT SCORES (*continued*)

Variable	All (1)	By sex		By Magen score (course grade ^a)	
		Boys (2)	Girls (3)	Low scores (4)	High scores (5)
<i>Panel B. Student-level data</i>					
Demographic information					
Mother's education (years)	11.44 (5.04)	11.60 (5.09)	11.30 (5.00)	10.79 (4.87)	12.08 (5.13)
Father's education (years)	11.62 (5.03)	11.83 (5.02)	11.44 (5.03)	10.85 (4.84)	12.39 (5.10)
Number of siblings	2.02 (1.58)	1.95 (1.49)	2.07 (1.65)	2.03 (1.61)	2.00 (1.55)
Bagrut outcomes and matriculation certification rates					
Bagrut composite score	70.76 (23.74)	68.91 (24.86)	72.33 (22.64)	53.22 (30.69)	77.10 (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)
Postsecondary enrollment rates					
Any postsecondary	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 and semi-engineering	0.070	0.063	0.076	0.078	0.059
Other ^b	0.046	0.036	0.055	0.046	0.047
Postsecondary schooling in years					
Any postsecondary	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 and 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 (NIS 2010) ^c	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 AQI value for each PM_{2.5} reading is calculated from a formula that converts micrograms ($\mu\text{g}/\text{m}^3$) into a 1–500 index value.

^aThe Magen score is composed of a score given to students based on coursework throughout the academic year and an end-of-year exam. The sample is split by whether the student's average Magen score over all subjects was above or below the median. In panel B, each observation represents a student. Receiving a matriculation certificate is determined by a combination of the student's average Bagrut and Magen score, and is a prerequisite for university enrollment.

^bThe other programs include technical schools, nonacademic colleges, and smaller schools.

^cWages are reported 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 are taken from the Israeli National Insurance Institute (*Bituach Leumi*).

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, ordinary least squares (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 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.

It is also worth noting that while we treat temperature and humidity as control variables, they could in theory be interesting in their own right. Extreme weather could also influence student performance, and represent an alternative ‘natural experiment’ affecting student performance. Empirically, we find that our coefficients on temperature and humidity are much smaller in magnitude than our coefficients on pollution, possibly since exam locations are required to have air conditioning, removing this as a channel influencing student performance.^{16,17}

C. Empirical Strategy II: The Long-Run Consequences of High-Stakes Examinations

Our analysis on long-term outcomes 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, as mentioned earlier, 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 levels on different exam dates is random, and therefore variation in average pollution exposure is also random.

¹⁶In our data, neither temperature nor humidity are statistically significant when used to predict test scores. Therefore, weather seems much less important as a source of test score variation than pollution, and we proceed with treating them only as control variables.

¹⁷Note that air conditioning can also serve to filter particulate matter, though air conditioning units are much less effective than stand-alone filters (Batterman et al. 2012). However, air conditioning does serve to filter out some particulate matter and it could be our results would be even larger in settings where air conditioning is not universally provided.

We estimate models relating the average pollution during the examinations to the student's composite score, after accounting for other observable factors that could influence scores. Formally, the first stage model that we estimate is of the following form:

$$(2) \quad R_{is} = \alpha \overline{PM}_{is} + f(\overline{Temp}_{is}, \overline{RH}_{is}) + X_{is}\Gamma + S_s + C_t + \varepsilon_{is},$$

where R_{is} is the Bagrut composite test score of student I at school s ; \overline{PM}_{is} is average air pollution exposure of student i at school s across the examinations; f is a flexible function of the mean temperature \overline{TEMP}_{is} and average humidity \overline{RH}_{is} across the examinations,¹⁸ X_{it} includes controls for the father's and mother's years of education and a dummy for the student's sex; S_s is a school fixed effect; C_t is a cohort fixed effect; and ε_{is} is a disturbance term.

The second stage equation is as follows:

$$(3) \quad O_{is} = \gamma \hat{R}_{is} + f(\overline{Temp}_{is}, \overline{RH}_{is}) + X_{is}\Lambda + S_s + C_t + v_{is},$$

where O_{is} represents a long-term academic or economic outcome, and \hat{R}_{is} represents the fitted values from estimating (1). Our long-term outcomes include Bagrut matriculation, postsecondary enrollment, postsecondary years of schooling, and monthly earnings, all measured at ages 28–30. As discussed, our exclusion restriction is that the residual variation in pollution across students within the same school (our instrument) affects our outcomes only through its relationship to Bagrut performance. Provided that this condition is satisfied, our empirical setup will allow us to generate unbiased 2SLS estimates of the influence of an additional point on the Bagrut (γ) on the long-term academic and economic outcomes of our sample of students.

III. Examination Performance and PM_{2.5} Exposure

A. Main Results

As a visual preview of our results, we present in Figure 1 a plot of Bagrut scores against PM_{2.5} (AQI) across over 400,000 exams. The plot is generated by regressing Bagrut scores and PM_{2.5} (AQI) on student fixed effects, calculating the residual, and averaging residual Bagrut scores over three unit bins of residual PM_{2.5} (AQI).¹⁹ We then examine the relationship between residual scores and pollution using lowess bandsmoother. The figure demonstrates that, on average, a student performs worse than his or her average when she faces pollution higher than her average pollution exposure across her exams. While clearly many other factors influence student performance, the plot suggests a robust negative relationship between pollution exposure and test scores, even when only exploiting variation within the same student's Bagrut examinations.

¹⁸In the empirical analysis, we include linear and squared terms of average temperature and average humidity (across the student's exams), and the interaction of the two variables.

¹⁹A version of this plot without binning is reported in online Appendix Figure A3.

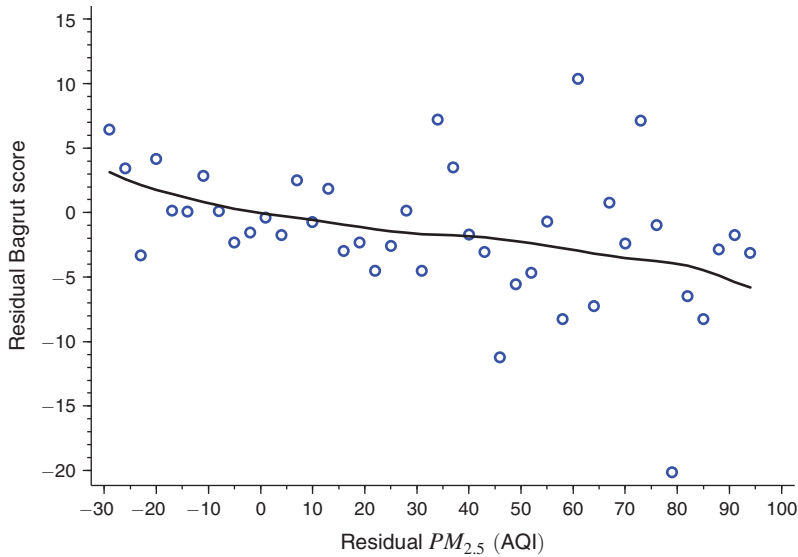


FIGURE 1. SCATTER PLOT OF RESIDUAL PM_{2.5} (AQI) AND BAGRUT TEST SCORES

Notes: The plot reports the relationship between residual Bagrut (test) scores and residual PM_{2.5} estimated by Lowess bandsmoother. Each observation on the plot is residual Bagrut scores averaged over bins of width three units residual PM_{2.5} (AQI). Residual Bagrut scores and residual PM_{2.5} are generated by regressing each variable on student fixed effects.

In Table 2, we report our baseline results of the relationship between the Air Quality Indicator values for PM_{2.5} 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 ten unit increase of PM_{2.5} (AQI) is associated with a 0.55 points decrease in a student's test score, significant at the 1 percent level. The results also indicate that a relatively small part of the variation in test scores ($R^2 = 0.003$) is explained by air pollution, as one would expect. 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 ten unit increase in pollution is associated with a 0.52 decrease in a student's score. Note that the sample with controls is roughly 20 percent smaller, as we have incomplete demographic information for these individuals (e.g., parental education). 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.²⁰

In columns 3–5 of Table 2, we take advantage of the panel structure of our data and include city, school, and student fixed effects, respectively. These account for

²⁰We also used the smaller sample to estimate the OLS regression without any controls and obtained estimates almost identical to those reported in column 1 of Table 2, which suggest the sample of students with some missing characteristics is not, on average, different from the rest of the sample.

TABLE 2—POOLED OLS AND FIXED EFFECT MODELS OF PARTICULATE MATTER'S IMPACT ON BAGRUT SCORES

	Pooled OLS		Fixed effects		
	No controls (1)	Controls (2)	City (3)	School (4)	Student (5)
PM _{2.5} (AQI) (10 units)	-0.55 (0.15)	-0.52 (0.11)	-0.70 (0.08)	-0.56 (0.07)	-0.40 (0.07)
Female (1 = yes)		3.37 (0.34)	3.44 (0.33)	2.89 (0.22)	
Mother's education		0.159 (0.065)	0.137 (0.063)	0.109 (0.036)	
Father's education		0.413 (0.06)	0.399 (0.06)	0.243 (0.03)	
R ²	0.003	0.055	0.059	0.159	0.510
Observations	415,219	380,435	380,435	380,435	380,435

Notes: The dependent variable in all regressions is Bagrut score (0–100). All regressions include suppressed controls for a linear and quadratic term in temperature and humidity, and the linear and quadratic interaction terms of the two variables. We additionally include day-of-the-week fixed effects, fixed effects for the level (difficulty) of the exam, gender, and the father and mother's education (except in models with student fixed effects). The coefficients are reported per ten units of PM_{2.5}(AQI). Standard errors are heteroskedastic-consistent and clustered by school.

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 -0.70 and -0.56 , respectively. Adding student fixed effects generates similar results, with our preferred estimate indicating that a ten unit increase in PM_{2.5} (AQI) is associated with a 0.40 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 0.083 ($0.040 \times (59.7 - 10.1)/23.7$) standard deviations.

The effect of PM_{2.5} on Bagrut scores for the ninety-ninth percentile of exposure in our sample (AQI = 137) is very large and implies a decline of roughly 0.13 of a standard deviation in scores relative to an average day's air quality. 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 score 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, Levy, and Lin 2008).

In Table 3, we report results where we examine whether pollution has a nonlinear impact on test takers using specifications where we include dummy variables for clean, moderately polluted, or very polluted days. For PM_{2.5}, we define moderately polluted days as days where the AQI score ranges from 51–74 (which the EPA defines as moderate pollution) and AQI scores above 75 as poor or very polluted days. The results largely point to a monotonic negative relationship between scores and pollution exposure, with very bad days being worse than only modestly polluted days. For example, column 5 of Table 3 indicates that having poor air quality

TABLE 3—PARTICULATE MATTER'S IMPACT ON BAGRUT SCORES ON POLLUTED AND EXTREMELY POLLUTED DAYS

	Pooled OLS		Fixed effects		
	No controls (1)	Controls (2)	City (3)	School (4)	Student (5)
Dummy for AQI > 50 and < 75	-1.97 (0.39)	-1.51 (0.50)	-2.29 (0.66)	-1.92 (0.64)	-1.51 (0.48)
Dummy for AQI ≥ 75	-3.32 (0.60)	-2.97 (0.69)	-3.77 (0.86)	-3.04 (0.81)	-2.25 (1.06)
R^2	0.003	0.055	0.059	0.159	0.511
Observations	415,219	380,435	380,435	380,435	380,435

Notes: The dependent variable in all regressions is Bagrut score (0–100). These regressions are estimated in the same manner as those in Table 2 (with the same controls) but we replace average $PM_{2.5}$ (AQI) with dummies for $PM_{2.5}$ (AQI) being less than 50, between 50 and 75, and above 75.

(AQI ≥ 75) from $PM_{2.5}$ exposure the day of the exam is associated with a 2.25 point decline in the student's Bagrut score, which is roughly 50 percent larger the size of the coefficient for moderately polluted days (1.51). 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.

B. 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 4, we compare the relationship between Bagrut scores with pollution on the day of the exam (row 1), the week before the exam (row 2), the month before the exam (row 3), and the year before the exam (row 4). Aside from row 1, the results in the rest of the table are not statistically different from zero, with a single exception that pollution a year before the exam is positively associated with scores in models with student effects.²¹ The lack of a significant effect in these placebo tests is reassuring that our results are not driven by a spurious correlation.

In Figure 2, we examine the impact of $PM_{2.5}$ on test scores where we use pollution from the three days prior to the exam, the day of the exam, and the three days following the exam on test scores. As shown in the figure, an additional ten units of AQI on the day of the exam is associated with a 1.5 point decrease in student performance. Furthermore, while pollution the day of a test has a large impact on test takers, pollution levels on other days of the week of the exam are almost unrelated to performance. This plot supports the claim that the effect identified is a transitory effect of pollution, with the effect driven primarily by exposure on the day of the exam.

²¹ Note that our sample drops significantly in this specification, since we have no pollution data for 1999, which may explain this result.

TABLE 4—MEASURING THE RELATIONSHIP BETWEEN BAGRUT PERFORMANCE AND PARTICULATE MATTER ON THE ACTUAL TEST DAY AND IRRELEVANT DAYS

	Pooled OLS		Fixed effects		
	No controls (1)	Controls (2)	City (3)	School (4)	Student (5)
Day of exam	-0.55 (0.15)	-0.52 (0.11)	-0.70 (0.08)	-0.56 (0.07)	-0.40 (0.07)
Previous week	0.13 (0.12)	0.01 (0.07)	-0.09 (0.12)	-0.09 (0.11)	-0.02 (0.11)
Previous month	0.08 (0.12)	-0.01 (0.15)	-0.07 (0.18)	-0.10 (0.14)	0.08 (0.11)
Previous year	-0.19 (0.14)	0.25 (0.18)	0.11 (0.26)	0.16 (0.20)	0.31 (0.11)

Notes: The dependent variable in all regressions is Bagrut score (0–100). Each cell in the table represents a separate regression. The regressions are estimated in the manner described in Table 2. In the first row, exam scores are matched to our $PM_{2.5}$ (AQI) on the day of the actual exam. In rows 2–4, we assign $PM_{2.5}$ (AQI) to each exam using the reading of $PM_{2.5}$ (AQI) for the week, month, or year prior to the actual exam. The coefficients are reported per ten units of $PM_{2.5}$ (AQI).

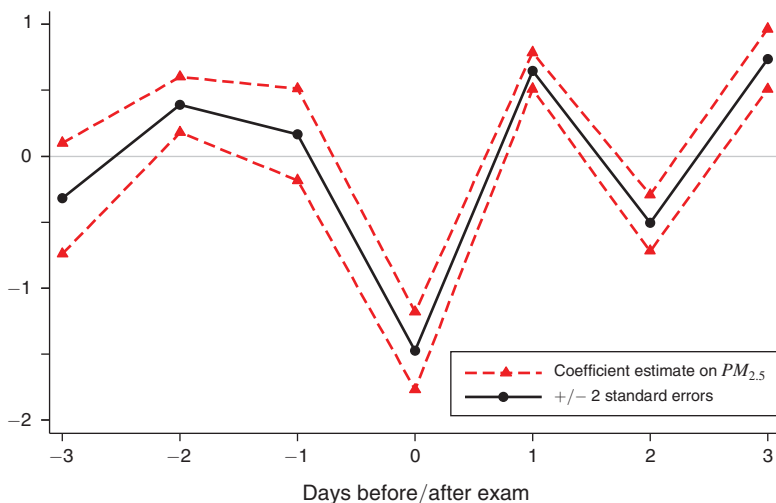


FIGURE 2. IMPACT OF $PM_{2.5}$ ON BAGRUT TEST SCORES IN THE DAYS BEFORE AND AFTER THE EXAM

Notes: The figure plots the coefficients from a regression of Bagrut test scores on $PM_{2.5}$ (AQI) readings in the days prior to, the day of (Day = 0), and the days following the examination, estimated in a single regression. Standard errors are clustered by school. Effects are reported in terms of change in score per ten additional units of $PM_{2.5}$ (AQI).

C. Heterogeneity

In this section, we examine heterogeneity in the treatment effects reported in Table 2. Our interest is twofold. First, we wish to identify whether there are subpopulations 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 $PM_{2.5}$ will have

a larger impact on groups who are more sensitive to poor air quality. 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, Cohen, and Danon (1993) the rate of asthma incidence in Israel is 25 percent higher among boys. Second, children of lower socioeconomic status are known to have higher rates of asthma and respiratory illnesses (Hedlund, Eriksson, and Rönmark 2006; Basagaña et al. 2004). Our third comparison is between stronger and weaker students as measured by their course grade (Magen). While we do not have a strong prior on who should be more affected, it may be that weaker students are less able to cope with the negative effects of pollution.

In Table 5, we examine our results separately by gender, student quality, and parental socioeconomic status. The results in panel A 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 two and four times larger than among women. For example, in models with student fixed effects, we estimate that an additional ten units of $PM_{2.5}$ (AQI), which is roughly the standard deviation of $PM_{2.5}$ (AQI), is associated with a 0.62 point decline (2.5 percent of a standard deviation) among men and a 0.24 point decline (1.1 percent of a standard deviation) 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 panel B of Table 5, 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. When we split the sample by whether their Magen score is above or below the median, our estimated treatment effects for ten additional units of $PM_{2.5}$ (AQI) are more than four times larger among those classified as low quality: a -0.66 point (2.2 percent of standard deviation (std)) versus a -0.14 point (0.6 percent of std) effect. In panel C, we examine the students separately by our measure of parental socioeconomic status (SES).²² 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). As mentioned earlier, poorer children also have higher incidence of asthma. Indeed, we find modestly larger effects for an additional 10 units of $PM_{2.5}$ (AQI): a decline of 0.46 for low SES students versus 0.30 for high SES students, suggesting a somewhat larger impact on students from poorer backgrounds. This could be driven by higher incidence of health issues, or through a correlation with being a weaker student (who appear more sensitive to pollution), as shown in panel B.

²²In a complementary exercise, we estimate the impact of pollution on Bagrut failure separately by Magen decile. The strongest relationship is found among students in the bottom decile (see online Appendix Figure A4).

TABLE 5—HETEROGENEITY IN PARTICULATE MATTER'S IMPACT ON BAGRUT SCORES ACROSS SUBPOPULATIONS

	Pooled OLS		Fixed effects		
	No controls (1)	Controls (2)	City (3)	School (4)	Student (5)
<i>Panel A. Boys and girls</i>					
Boys	-0.87 (0.15)	-0.73 (0.11)	-0.91 (0.18)	-0.80 (0.16)	-0.62 (0.20)
Girls	-0.31 (0.09)	-0.34 (0.09)	-0.52 (0.11)	-0.37 (0.12)	-0.24 (0.13)
<i>Panel B. Low and high course grades</i>					
Low course grades	-0.75 (0.12)	-0.72 (0.09)	-0.77 (0.15)	-0.72 (0.14)	-0.66 (0.19)
High course grades	-0.27 (0.05)	-0.07 (0.09)	-0.19 (0.10)	-0.13 (0.10)	-0.14 (0.13)
<i>Panel C. Low and high socioeconomic status (SES)</i>					
Low SES	-0.67 (0.12)	-0.66 (0.08)	-0.77 (0.12)	-0.64 (0.13)	-0.46 (0.16)
High SES	-0.36 (0.11)	-0.34 (0.14)	-0.50 (0.16)	-0.40 (0.16)	-0.30 (0.19)

Notes: The dependent variable in all regressions is Bagrut score (0–100). Each cell in the table represents a separate regression. The regressions are estimated in the same manner as those in Table 2. A student is determined to have low/high course grades if her average Magen score is below/above the sample median. A student is determined to have low/high socioeconomic status if her father's education is below/above the sample median. The coefficients are reported per ten units of $PM_{2.5}$ (AQI).

IV. The Long-Term Consequences of Pollution Exposure during Examinations

A. The Reduced Form Impact of Particulate Matter Exposure during the Bagrut 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 the same school. As previously discussed, this is driven by student choice of elective subjects for exams as well as variation across cohort (which take different exams at the end of each year). Within-school variation in pollution is unlikely to be correlated with student potential outcomes since neither students nor schools have control over the timing of test. We present empirical evidence that within-school variation in pollution is not related to student quality in balancing exercises presented in Table A1 in the online Appendix. The results indicate that, conditioned on city or school, pollution does not appear to be correlated with observable features of the student. We also verify that our results are not related to long-term harm from pollution in Table A2 in the online Appendix, in which we see that pollution from eleventh grade is uncorrelated with test scores in twelfth grade. Therefore, the results we present in this section are likely to be the direct consequence of transitory pollution exposure during the Bagrut exams, rather than related to omitted variables or due to permanent cognitive damage from pollution.

In Table 6, we present the reduced form effect of average $PM_{2.5}$ on several academic outcomes related to the Bagrut, including average score (composite score),

TABLE 6—PARTICULATE MATTER'S IMPACT ON POSTSECONDARY EDUCATION AND ADULT EARNINGS

	Pooled OLS	Fixed effects	
	Controls (1)	City (2)	School (3)
Bagrut composite score	-0.67 (0.08)	-2.66 (0.13)	-1.64 (0.18)
Number of Bagrut failures	0.081 (0.084)	0.197 (0.011)	0.106 (0.020)
Proportion of Bagrut failures	0.008 (0.001)	0.027 (0.002)	0.015 (0.003)
Matriculation certification	-0.023 (0.002)	-0.053 (0.003)	-0.033 (0.005)
Enrolled in postsecondary institution (1 = yes)	-0.009 (0.002)	-0.050 (0.003)	-0.031 (0.004)
Completed years of postsecondary education	-0.067 (0.009)	-0.236 (0.013)	-0.152 (0.018)
Average monthly earnings (NIS)	-155 (33)	-120 (33)	-109 (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 using the student-level sample described in Table 1. The coefficients are reported per ten units of $PM_{2.5}$ (AQI). All regressions include suppressed controls for average temperature and humidity during the Bagrut, mother's and father's years of schooling, sex, and student's age in 2010. Standard errors are heteroskedastic-consistent, clustered by school, and are reported below the coefficients in parentheses.

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 ten 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 percent and 20 percent of a standard deviation, respectively ($\sigma = 23.7$). 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 ten units of $PM_{2.5}$ (AQI) raises the probability of failure on a given Bagrut exam by 2 percentage points, raises the total number of failed exams by 0.11, and lowers matriculation certification rates by 3 percentage points. The standard deviation of average $PM_{2.5}$ (AQI) is 15.5 units, so increasing average $PM_{2.5}$ (AQI) across a student's exams by a full standard deviation would raise these estimates by 55 percent (relative to any estimate reported per ten units of AQI). Since matriculation certification is required by many elite postsecondary academic institutions in Israel, it is likely that students who suffer a negative shock that lowers their certification probability will ultimately be impacted in their prospects for higher education.

In rows 5 and 6 of Table 6, we present the estimated effect of average pollution exposure on two longitudinal educational outcomes: enrollment in postsecondary institution (1 = yes), and years of postsecondary schooling attained. Indeed, we find that enrollment rates in higher education decline by 3 percentage points and schooling declines by 0.15 years when a student is exposed to an additional ten units of $PM_{2.5}$ (AQI). All estimates are statistically significant at the 5 percent level,

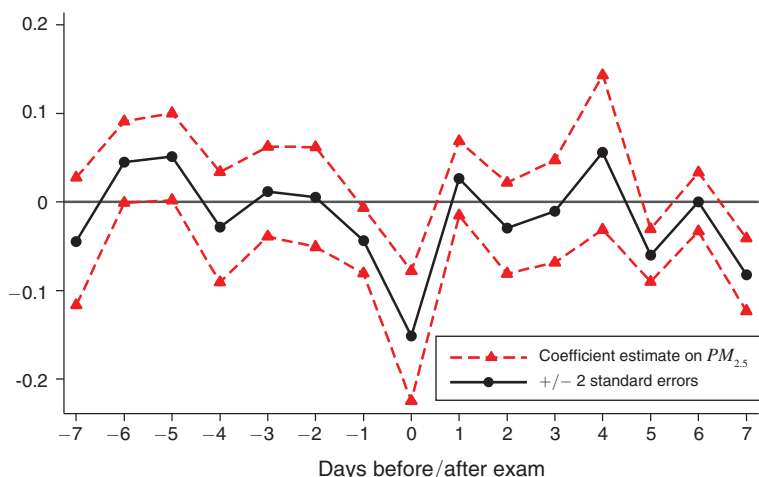


FIGURE 3. IMPACT OF $PM_{2.5}$ ON POSTSECONDARY EDUCATION USING AVERAGE $PM_{2.5}$ ON THE DAYS LEADING UP TO AND FOLLOWING THE EXAM

Notes: The figure plots the coefficients from a series of regressions where we predict postsecondary schooling using our standard controls from our analysis of long-term outcomes and average pollution exposure. In this figure, we replace average actual pollution on the day of the exam with the pollution average on the days leading up to and following the exam. Standard errors are clustered by school. Effects are reported in terms of change in enrollment probability per ten additional units of $PM_{2.5}$ (AQI).

and suggest that taking Bagrut exams in highly polluted days can have long lasting effects on schooling attainment.

In Figure 3, we complement these results with a placebo exercise estimating the relationship between postsecondary schooling and average pollution on days other than the actual exam. In particular, we estimate a modified version of equation (2), where we replace \overline{PM}_{it} with alternative measures of pollution that are generated from the pollution levels on days leading up to and following the exam. We generate this “mis-assigned” pollution level using readings from the days in the week before and after the actual exam, creating 14 additional placebo pollution measures. We report in Figure 3 the results of these 15 separate regressions (including the pollution on the day of the exam). As anticipated, the observed negative relationship between pollution and postsecondary schooling is much stronger using pollution from the day of the actual exam, and in most other instances, our estimates are not statistically different than zero. This is supportive evidence that our results on postsecondary education are driven by the transitory effect of pollution, rather than other explanations.

In row 7 of Table 6, 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 to an additional ten units of $PM_{2.5}$ (AQI) during the Bagrut exam dates is associated with an average monthly earnings decline at age 28 of 109 shekels (\$30), or 2.1 percent. This estimate is also precisely estimated, with a T -statistic greater than three. A visual complement to this result is presented in online Appendix Figure A5, where we demonstrate a

negative relationship between residual $PM_{2.5}$ exposure (after inclusion of school fixed effects) during the Bagrut and residual test scores across quintiles of pollution exposure, especially among test takers who took the test on very polluted days. Note that our effects on wages may be manifest either through lower hourly wages, fewer hours worked, or exit from the labor force entirely. Therefore, these results should be interpreted as the “reduced form” effect of having a lower score on wages, mediated through these potential channels. Our baseline results include those who are not working and have zero wages.²³ However, we do verify that our results are similar including or excluding observations with zero wages in online Appendix Table A3.

B. The Long-Term Consequences of Random Variation in Bagrut Scores

In Table 7, we use the highly significant reduced form effect of pollution on a student’s Bagrut composite score as a first stage to examine the long-term consequences of exogenous variation in exam scores. It is worth noting that our 2SLS results should be interpreted with caution, as other pollutants (e.g., CO) may be correlated with $PM_{2.5}$ and also influence scores, violating the exclusion restriction. However, we proceed with this analysis as a way of evaluating the plausibility of our reduced form results. In panel A, we estimate the economic return to an additional point on the Bagrut composite score using 2SLS. In the first row of Table 7, we reproduce the relationship between the Bagrut composite score and $PM_{2.5}$ shown in Table 6 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 45/66 shekels in monthly earnings in models with city and school fixed effects. Relative to the average wage in our sample (5,084 shekels), this implies each additional point is worth roughly a full percent of monthly salary. Since the standard deviation of the Bagrut composite score is roughly 13 points, our estimates imply that even modest declines in scores induced by pollution can have significant consequences on adult income.

In panel B of Table 7, 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 courses of study that lead to lucrative occupations, 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 matriculation certificate rates by 2.0 percentage points, enrollment rates in postsecondary schooling by 1.9 percentage points, and postsecondary educational attainment by 0.092 years. This indicates that each additional Bagrut point can have important consequences for a student’s future attainment of postsecondary schooling, highlighting the importance of the test in the Israeli educational system.

In panel C of Table 7, we exploit the relationship between pollution exposure and the Bagrut composite score to estimate the return to an additional year of postsecondary schooling. It is worth noting that this strategy does not identify

²³ Since delayed entry to the workforce or inability to find work are both possible consequences of a poor Bagrut score, we chose to keep those with zero wages in our baseline results.

TABLE 7—THE ECONOMIC AND ACADEMIC RETURN TO THE BAGRUT

	Pooled OLS		Fixed effects	
	Controls (1)	City (2)	School (3)	
<i>Panel A. Effect of the Bagrut composite score on adult earnings using $PM_{2.5}(AQI)$ as an IV</i>				
First stage	-0.67 (0.08)	-2.66 (0.13)		-1.64 (0.18)
Reduced form	-155 (33)	-120 (33)		-109 (34)
2SLS	229 (147)	45 (13)		66 (21)
<i>Panel B. Effect of the Bagrut composite score on follow-up academic outcomes using $PM_{2.5}(AQI)$ as an IV</i>				
Matriculation certification	0.034 (0.011)	0.020 (0.002)		0.020 (0.002)
Enrolled in postsecondary institution (1 = yes)	0.016 (0.006)	0.019 (0.002)		0.019 (0.002)
Completed years of postsecondary education	0.105 (0.026)	0.089 (0.006)		0.092 (0.009)
<i>Panel C. Estimated return to postsecondary education using $PM_{2.5}(AQI)$ as an IV</i>				
First stage	-0.067 (0.009)	-0.236 (0.013)		-0.152 (0.018)
Reduced form	-1,548 (326)	-1,199 (331)		-1,093 (344)
2SLS	2,278 (1,343)	509 (139)		707 (219)

Notes: Each cell in the table represents a separate regression. The regressions are estimated with the same set of control variables as in Table 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 IV. In panel B, we present 2SLS models of the relationship between the Bagrut composite score and other academic outcomes using $PM_{2.5}(AQI)$ as an IV. In panel C, we estimate the implied return to postsecondary schooling using $PM_{2.5}(AQI)$ as an IV. All first-stage F -statistics exceed ten. Standard errors are heteroskedastic-consistent, clustered by school, and are reported below the coefficients in parentheses.

“cleanly” the rate of return to schooling since the Bagrut score can directly affect earnings, and therefore its omission might violate the exclusion restriction. Furthermore, the omission of other pollutants correlated with $PM_{2.5}$ may bias our results. However, as a 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 postsecondary schooling as the endogenous regressor and $PM_{2.5}$ as the instrument, we estimate using 2SLS that each additional year of postsecondary schooling is worth 707 (\$191) shekels. This estimate implies a rate of return to college education of 14 percent, 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 percent.

In Table 8, we examine possible mechanisms for our results by examining how pollution affects the probability of a student matriculating at different types of postsecondary 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

TABLE 8—PARTICULATE MATTER'S IMPACT ON POSTSECONDARY SCHOOLING BY TYPE

	LHS: Enrolled in postsecondary institution (1 = yes)		LHS: completed years of postsecondary education	
	City (1)	School (2)	City (3)	School (4)
Any postsecondary institution	-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 dependent 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 2, and the coefficients are reported per ten 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, clustered by school, and are reported below the coefficients in parentheses.

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 pursue 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.²⁴ As reported in Table 8, 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.

C. Heterogeneity in the Long-Term Consequences of Random Variation in Bagrut Scores

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 backgrounds. 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.

²⁴ 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.

TABLE 9—HETEROGENEITY IN THE ECONOMIC AND ACADEMIC RETURN TO THE BAGRUT

	By sex		By course grades		By socioeconomic status	
	Boys (1)	Girls (2)	Low (3)	High (4)	Low (5)	High (6)
<i>Panel A. Effect of the Bagrut composite score on adult earnings using $PM_{2.5}(AQI)$ as an IV</i>						
First stage	-1.97 (0.26)	-1.37 (0.19)	-0.66 (0.18)	-1.22 (0.16)	-1.32 (0.20)	-2.11 (0.25)
2SLS	78 (27)	59 (30)	80 (58)	124 (51)	56 (24)	105 (32)
<i>Panel B. Effect of the Bagrut composite score on follow-up academic outcomes</i>						
Matriculation certification	0.017 (0.002)	0.025 (0.003)	0.032 (0.008)	0.011 (0.004)	0.023 (0.003)	0.015 (0.002)
Enrolled in postsecondary institution (1 = yes)	0.018 (0.002)	0.020 (0.003)	0.029 (0.008)	0.019 (0.004)	0.021 (0.003)	0.013 (0.002)
Completed years of postsecondary education	0.087 (0.010)	0.108 (0.014)	0.108 (0.030)	0.118 (0.015)	0.096 (0.013)	0.085 (0.010)
<i>Panel C. Estimated return to postsecondary education using $PM_{2.5}(AQI)$ as an IV</i>						
First stage	-0.17 (0.02)	-0.14 (0.02)	-0.08 (0.02)	-0.13 (0.03)	-0.13 (0.02)	-0.17 (0.03)
2SLS	888 (296)	564 (265)	698 (484)	1,131 (454)	580 (235)	1,264 (405)

Notes: Each cell in the table represents a separate regression. All specifications include school fixed effects. A student is determined to have low/high course grades if her average Magen score is below/above the sample median. A student is determined to have low/high socioeconomic status if her father's education is below/above the sample median. All first-stage F -statistics exceed ten. Standard errors are heteroskedastic-consistent, clustered by school, and are reported below the coefficients in parentheses.

In panel A of Table 9, 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 percent higher for boys than girls: 78 shekels versus 59 shekels (\$21 versus \$16). One explanation is that women choose less financially rewarding fields of study than men, even when they have similar qualifications. It is also worth noting that although female labor force participation rates are relatively similar to the United States, Israeli women have much higher fertility than their American counterparts.²⁵ 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.

²⁵ Average fertility rates in Israel are 3.0, roughly 50 percent 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 percent and 61 percent, respectively (Israel 1995 census), compared to rates in the United States of 86 percent and 69 percent (US 2010 census).

In columns 3 and 4 of Table 9, 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 an 80 shekel return among lower quality students (\$34 versus \$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-low chance of being accepted into a very lucrative profession. This may help explain differences that are observed in columns 5 and 6, where 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). 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 nonbinding 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 a less privileged background.

The results in Table 8 are complemented by Figure 4, where we stratify our sample by decile of Magen score. In the figure, we report the coefficient from linear probability models of either university or college matriculation on $PM_{2.5}$ exposure during the Bagrut. Interestingly, the results indicate that at lower levels of student quality, pollution exposure is negatively associated with college matriculation but unrelated to university attendance. However, at higher deciles of student quality, pollution exposure is *positively* associated with attendance at less-prestigious colleges and negatively associated with university attendance. These results are reinforced by Figure A6 in the online Appendix, which indicates that the negative relationship between wages and Bagrut pollution exposure is largest among the top decile of students, suggesting that our effects are particularly relevant for students competing for the best schools and occupations.

In panel B of Table 9, we examine the mechanisms for the aforementioned results by estimating 2SLS models where pollution is our instrument and we treat the Bagrut composite score as our endogenous regressor. We repeat our earlier analysis performed on the overall sample, and examine three 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 postsecondary institutions, and by affecting total completed postsecondary 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 in terms of the consequences for wages, but in terms of their likelihood of pursuing postsecondary education. One explanation for this finding is that stronger students and students from wealthier backgrounds are more likely to retake the exam if they perform poorly. While retaking

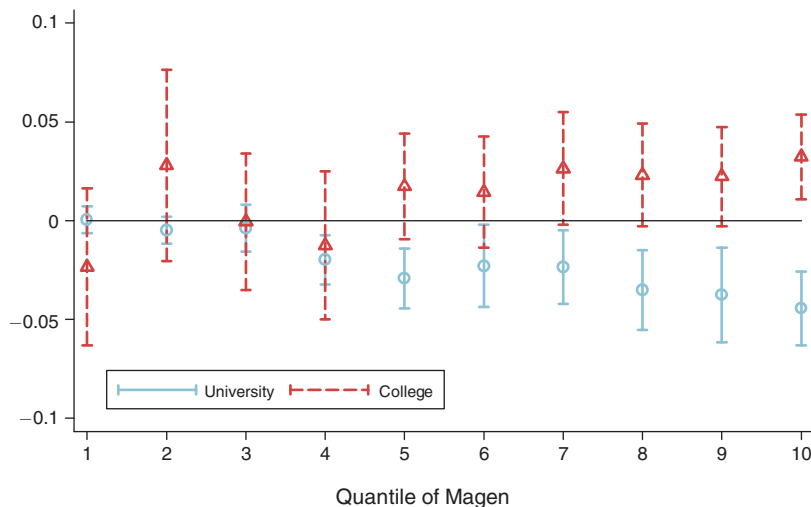


FIGURE 4. IMPACT OF $PM_{2.5}$ ON UNIVERSITY AND COLLEGE ENROLLMENT BY STUDENT QUALITY DECILE

Notes: The figure plots the coefficients from linear probability models of university (solid) or college (dashed) enrollment on $PM_{2.5}$ (AQI) with school fixed effects, separately by Magen (average course grade) decile. Standard errors are clustered by school. Effects are reported in terms of change in enrollment probability per ten additional units of $PM_{2.5}$ (AQI).

the examination is uncommon, it is allowed but generally involves participation in expensive preparatory classes for students to refamiliarize themselves with the material. Therefore, the consequences of a single bad outcome may be lower for students who anticipate doing well by retaking the exam, or those from privileged backgrounds who can more easily absorb the financial costs of retaking the exam.²⁶

In panel C of Table 9, we examine heterogeneity in the estimated return to education by sub-population in Israel using 2SLS, with pollution during the Bagrut serving as our instrument for postsecondary education. Again note that since pollution affects scores as well, this will not satisfy the exclusion restriction, but is worth exploring to assess the economic magnitude of our estimated effects. The results are similar to the patterns we find in the return to an additional point on the Bagrut. In the first two columns, where we stratify the sample by sex, 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 postsecondary 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 postsecondary 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

²⁶ Vigdor and Clotfelter (2003) found this to be important in the US context, where students from wealthier backgrounds are more likely to retake the SAT.

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 postsecondary 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.

V. Conclusions

This paper has examined the relationship between pollution exposure during Israeli matriculation exams, student exam performance, and long-run academic and economic outcomes. In the first section of our analysis, we demonstrate that exposure to $PM_{2.5}$ during Bagrut examinations has a statistically and economically significant effect on student performance. In the second section, we analyze this group of test takers a decade later and examined whether the exogenous variation in scores generated by $PM_{2.5}$ has long-term consequences. We find that pollution exposure during the exams leads to significant declines in postsecondary education and earnings, indicating that even random variation in test scores can influence a student's academic path and earnings potential.

Our results demonstrate that short-term cognitive function may be affected by pollution exposure and that in the context of high-stakes exams, this may have significant long-term consequences on test takers. More generally, 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 our scope to consider the aggregate efficiency loss associated with the current Israeli system, our 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, our 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. Our findings lend empirical support for the concern voiced by officials in the United States 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 also consider adopting strict standards on exam days in the spirit of fairness to test takers, and in order to reduce the noise in this important measure of student quality.

REFERENCES

- Ailshire, Jennifer A., and Philippa Clarke. 2015. "Fine Particulate Matter Air Pollution and Cognitive Function Among U.S. Older Adults." *Journal of Gerontology, Series B: Psychological Sciences* 70 (2): 322–28.
- Arceo, Eva, Rema Hanna, and Paulina Oliva. 2016. "Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City?" *Economic Journal* 126 (591): 257–80.

- Angrist, Joshua D., and Stacey Chen. 2011. "Schooling and the Vietnam-Era GI Bill: Evidence from the Draft Lottery." *American Economic Journal: Applied Economics* 3 (2): 96–118.
- Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (2): 533–75.
- Angus, Robert, Ronald J. Heslegrave, and W. Stewart Myles. 1985. "Effects of Prolonged Sleep Deprivation, With and Without Chronic Physical Exercise, on Mood and Performance." *Psychophysiology* 22 (3): 276–82.
- Basagaña, Xavier, Jordi Sunyer, Manolis Kogevinas, Jan-Paul Zock, Enric Duran-Tauleria, Deborah Jarvis, Peter Burney, et al. 2004. "Socioeconomic Status and Asthma Prevalence in Young Adults: The European Community Respiratory Health Survey." *American Journal of Epidemiology* 160 (2): 178–88.
- Batterman, S., L. Du, G. Mentz, B. Mukherjee, E. Parker, C. Godwin, J.-Y. Chin, et al. 2012. "Particulate matter concentrations in residences: An intervention study evaluating stand-alone filters and air conditioners." *Indoor Air* 22 (3): 235–52.
- Bell, Michelle L., Jason K. Levy, and Zhaohui Lin. 2008. "The effect of sandstorms and air pollution on cause-specific hospital admissions in Taipei, Taiwan." *Occupational and Environmental Medicine* 65 (2): 104–11.
- Biederman, Joseph, Erik Mick, Stephen Faraone, Ellen Braaten, Alysa Doyle, Thomas Spencer, Timothy E. Wilens, et al. 2002. "Influence of Gender on Attention Deficit Hyperactivity Disorder in Children Referred to a Psychiatric Clinic." *American Journal of Psychiatry* 159 (1): 36–42.
- Braníš, Martin, Pavla Řezáčová, and Markéta Domasová. 2005. "The effect of outdoor air and indoor human activity on mass concentrations of PM₁₀, PM_{2.5}, and PM₁ in a classroom." *Environmental Research* 99 (2): 143–49.
- Calderón-Garcidueñas, Lilian, Antonieta Mora-Tiscareño, Esperanza Ontiveros, Gilberto Gómez-Garza, Gerardo Barragán-Mejía, James Broadway, Susan Chapman, et al. 2008. "Air pollution, cognitive deficits and brain abnormalities: A pilot study with children and dogs." *Brain and Cognition* 68 (2): 117–27.
- Cawley, John, James Heckman, and Edward Vytlačil. 2001. "Three observations on wages and cognitive ability." *Labour Economics* 8 (4): 419–42.
- Chang, Tom, Joshua S. Graff Zivin, Tal Gross, and Matthew J. Neidell. 2014. "Particulate Pollution and the Productivity of Pear Packers." National Bureau of Economic Research (NBER) Working Paper 19944.
- Chay, Kenneth Y., and Michael Greenstone. 2003. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics* 118 (3): 1121–67.
- Clark, Donald D., and Louis Sokoloff. 1999. "Circulation and energy metabolism of the brain." In *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, edited by George J. Siegel, Bernard W. Agranoff, R. Wayne Albers, Stephen K. Fisher, and Michael D. Uhler, 637–70. Philadelphia: Lippincott-Raven.
- Dockery, Douglas, and Arden Pope. 1996. "Epidemiology of Acute Health Effects: Summary of Time Series Studies." In *Particles in Our Air*, edited by Richard Wilson and John D. Spengler, 123–48. Cambridge: Harvard University Press.
- Ebenstein, Avraham, Victor Lavy, and Sefi Roth. 2016. "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution: Dataset." *American Economic Journal: Applied Economics*. <http://dx.doi.org/10.1257/app.20150213>.
- Goriounova, Natalia A., and Huibert D. Mansvelde. 2012. "Short- and long-term consequences of nicotine exposure during adolescence for prefrontal cortex neuronal network function." *Cold Spring Harbor Perspectives in Medicine* 2 (12).
- Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review* 102 (7): 3652–73.
- Ham, John C., Jacqueline S. Zweig, and Edward Avol. 2011. "Pollution, Test Scores and the Distribution of Academic Achievement: Evidence from California Schools 2002–2008." http://www.iza.org/conference_files/TAM2012/ham_j1496.pdf.
- Hedlund, Ulf, Kåre Eriksson, and Eva Rönmark. 2006. "Socio-economic status is related to incidence of asthma and respiratory symptoms in adults." *European Respiratory Journal* 28 (2): 303–10.
- Israel Central Bureau of Statistics. 1995. *Census of Population and Housing 1995*. Jerusalem: Israel Central Bureau of Statistics.
- Jarvis, Martin J. 1993. "Does caffeine intake enhance absolute levels of cognitive performance?" *Psychopharmacology* 110 (1): 45–52.

- Laor, Arie, Leon Cohen, and Yehuda L. Danon.** 1993. "Effects of time, sex, ethnic origin, and area of residence on prevalence of asthma in Israeli adolescents." *British Medical Journal* 307 (6908): 841–44.
- Lavy, Victor.** 2009. "Performance Pay and Teachers' Effort, Productivity and Grading Ethics." *American Economic Review* 99 (5): 1979–2011.
- Lavy, Victor, Avraham Ebenstein, and Sefi Roth.** 2014a. "The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation." National Bureau of Economic Research (NBER) Working Paper 20648.
- Lavy, Victor, Avraham Ebenstein, and Sefi Roth.** 2014b. "The Long Run Human Capital and Economic Consequences of High-Stakes Examinations." National Bureau of Economic Research (NBER) Working Paper 20647.
- Lewin, Tamar.** 2014. "A New SAT Aims to Realign With Schoolwork." *New York Times*, March 5. http://www.nytimes.com/2014/03/06/education/major-changes-in-sat-announced-by-college-board.html?_r=0.
- Mills, Nicholas L., Ken Donaldson, Paddy W. Hadoke, Nicholas A. Boon, William MacNee, Flemming R. Cassee, Thomas Sandström, et al.** 2009. "Adverse cardiovascular effects of air pollution." *Nature Clinical Practice Cardiovascular Medicine* 6 (1): 36–44.
- Moretti, Enrico, and Matthew Neidell.** 2011. "Pollution, Health, and Avoidance Behavior Evidence from the Ports of Los Angeles." *Journal of Human Resources* 46 (1): 154–75.
- Neidell, Matthew J.** 2004. "Air pollution, health, and socio-economic status: The effect of outdoor air quality on childhood asthma." *Journal of Health Economics* 23 (6): 1209–36.
- Pope, C. Arden, III, David V. Bates, and Mark E. Raizenne.** 1995. "Health Effects of Particulate Air Pollution: Time for Reassessment?" *Environmental Health Perspectives* 103 (5): 472–80.
- Pope, C. Arden, III, and Douglas W. Dockery.** 2006. "Critical Review: Health Effects of Fine Particulate Air Pollution: Lines that Connect." *Journal of the Air and Waste Management Association* 56: 709–42.
- United States Census Bureau.** 2010. "2010 Census." Washington, DC: United States Census Bureau.
- Vigdor, Jacob L., and Charles T. Clotfelter.** 2003. "Retaking the SAT." *Journal of Human Resources* 38 (1): 1–33.
- Wilker, Elissa H., Sarah R. Preis, Alexa S. Beiser, Philip A. Wolf, Rhoda Au, Itai Kloog, Wenyuan Li, et al.** 2015. "Long-Term Exposure to Fine Particulate Matter, Residential Proximity to Major Roads and Measures of Brain Structure." *Stroke* 46 (5): 1161–66.
- Winerip, Michael.** 2005. "SAT Essay Test Rewards Length and Ignores Errors." *New York Times*, May 4. <http://www.nytimes.com/2005/05/04/education/04education.html>.
- Willis, Robert J., and Sherwin Rosen.** 1979. "Education and Self-Selection." *Journal of Political Economy* 87 (5, Pt. 2): S7–36.