The long term economic consequences of having a bad day: How high-stakes exams mismeasure potential

VICTOR LAVY
AVRAHAM EBENSTEIN
SEFI ROTH

June 2016

SUMMARY

- Examination candidates can have a bad day for many reasons unrelated to knowledge, skill, or cognitive ability. Possible causes include minor infection, migraine, hay fever, menstruation, disturbed sleep, and atmospheric pollution. When the stakes are high, as with exams used to rank students or admit them to further training or employment, there can be permanent consequences for the individual and for society.

- There is a lack of evidence on these consequences. Empirical challenges include the difficulty in identifying the return to cognitive ability separately from the return to doing well on the examination.

- Our solution to this problem is to examine the consequences of fluctuations in a random factor on exam performance. We use fluctuations in air pollution for this purpose. When the same student takes multiple exams, and exposure to ambient pollution varies randomly from day to day, we can use the associated variation in performance to measure the component of a student's score which is related entirely to luck. The results show that transitory ambient pollution exposure is associated with a significant decline in student performance.

- We then examine these students during adulthood (8-10 years after the exams) and we find that pollution exposure during exams causes lasting damage to post-secondary educational attainment and earnings later in life.

- Our analysis highlights that high-stakes exams provide measures of student quality that may be imprecise and misleading. These measures can lead to allocative inefficiency because students who have had a bad day because of factors outside their control are mis-ranked. After that, they are inefficiently assigned to further training and to different occupations and this reduces labour productivity overall.

- As well as illustrating these problems with high-stakes exams, our findings also expand understanding of the costs of pollution. They imply that a narrow focus on traditional health outcomes, such as hospitalisation and increased mortality, is not enough. The full cost of pollution includes loss of mental acuity, which is essential to productivity in most professions. The use of high-stakes exams to measure mental acuity then multiplies this loss over many years.

ABOUT THE SERIES

CAGE and Social Market Foundation (SMF) policy briefings in the Global Perspectives Series are non-technical summaries of one or more academic research papers intended for distribution among policy makers and journalists. They are available to the public on the CAGE and SMF websites. Policy briefings are not intended as independent scholarly publications. When referring to the results described here, please cite the original academic papers, if possible in their latest or published version, as insights may have changed as research progressed.

ABOUT CAGE

Established in January 2010, CAGE is a research centre in the Department of Economics at the University of Warwick. CAGE is carrying out a ten year programme of innovative research. The centre's research programme is focused on how countries succeed in achieving key economic objectives such as improving living standards, raising productivity, and maintaining international competitiveness, which are central to the economic wellbeing of their citizens. Our research analyses the reasons for economic outcomes both in developed economies like the UK and emerging economies such as China and India. We aim to develop a better understanding of how to promote institutions and policies which are conducive to successful economic performance and endeavour to draw lessons for policy makers from economic history as well as the contemporary world. Research at CAGE examines how and why different countries achieve economic success. CAGE defines 'success' in terms of well-being as well as productivity. The research uses economic analysis to address real-world policy issues. The centre is distinctive in providing a perspective that draws on economic history as well as economic theory and is applied to countries at various different stages of economic development. www.warwick.ac.uk/cage

ABOUT THE SOCIAL MARKET FOUNDATION

The Social Market Foundation is a non-partisan think tank. We believe that fair markets, complemented by open public services, increase prosperity and help people to live well. We conduct research and run events looking at a wide range of economic and social policy areas, focusing on economic prosperity, public services and consumer markets. The SMF is resolutely independent, and the range of backgrounds and opinions among our staff, trustees and advisory board reflects this. www.smf.co.uk

How high-stakes exams mismeasure potential

INTRODUCTION

Although many countries use high-stakes testing to rank students – in the UK, high-stakes exams include the Eleven Plus, A Levels and bespoke university admissions tests - the consequences of this policy are largely unknown. Does a particularly good or bad performance on a high-stakes exam have long-term consequences for test takers, after accounting for a student's cognitive ability? Insofar as there are permanent wage consequences to variation induced by completely random shocks to student performance, 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 of student quality, since it may lead to poor matching between students and occupations, and an inefficient allocation of labour. In spite of a dearth of evidence regarding the consequences of these tests, they are used extensively globally to rank students and allocate opportunity by acting as a gatekeeper in admissions. This briefing paper evaluates this policy question by examining the potential long-term effect of random transitory disturbances to cognitive performance during high-stakes exit exams in Israeli high schools.

ASSESSING THE CONSEQUENCES OF USING HIGH STAKES EXAMINATIONS IS CHALLENGING

There are several challenges in assessing the consequences of using high-stakes examinations for ranking students. First, large data samples are generally not available with standardised test scores and wages during adulthood for a representative population.[1] Second, since higher-ability students presumably perform better on high-stakes tests, it is difficult to separately distinguish the return to cognitive ability from the return to doing well on the examination.

One possible solution is to examine the consequences of fluctuations in a random component affecting performance on these tests. A candidate is fluctuation in air pollution that might have an effect on cognitive acuity and test scores, therefore generating plausibly random variation in a given student's outcome. Air pollution has been demonstrated to adversely affect productivity across a variety of tasks (Graff Zivin and Neidell 2012, Chang et al. 2014). In addition, since students in Israel are assigned to test sites without prior knowledge of pollution or the ability to reschedule, it represents an exogenous factor affecting performance. This may enable direct measurement of the return to the component of a student's score which is related entirely to luck, and provide evidence regarding whether these tests do or do not have long-term consequences.

BACKGROUND ON THE ADVERSE EFFECTS OF AMBIENT AIR POLLUTION

Previous research has documented a robust 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, Arceo, Hanna, and Oliva 2015). 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.[2] 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 productivity in a variety of contexts (Graff Zivin and Neidell 2012, 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). 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.

ESTIMATING THE EFFECT OF AMBIENT AIR POLLUTION ON STUDENT PERFORMANCE

In this section we present empirical evidence that exposure during exams to $PM_{2.5}$ (our measure of pollution) is associated with a decline in student performance on high-stakes examinations. 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. As a result, we are able to exploit short-term episodes of pollution 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 labour 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 which will be discussed in the subsequent section.

To start off, we present in Figure 1 a plot of *Bagrut* scores against our pollution measure across over 400,000 exams[3]. 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.

Figure 1: Scatter plot of residual PM$_{2.5}$ (AQI) and *Bagrut* test scores
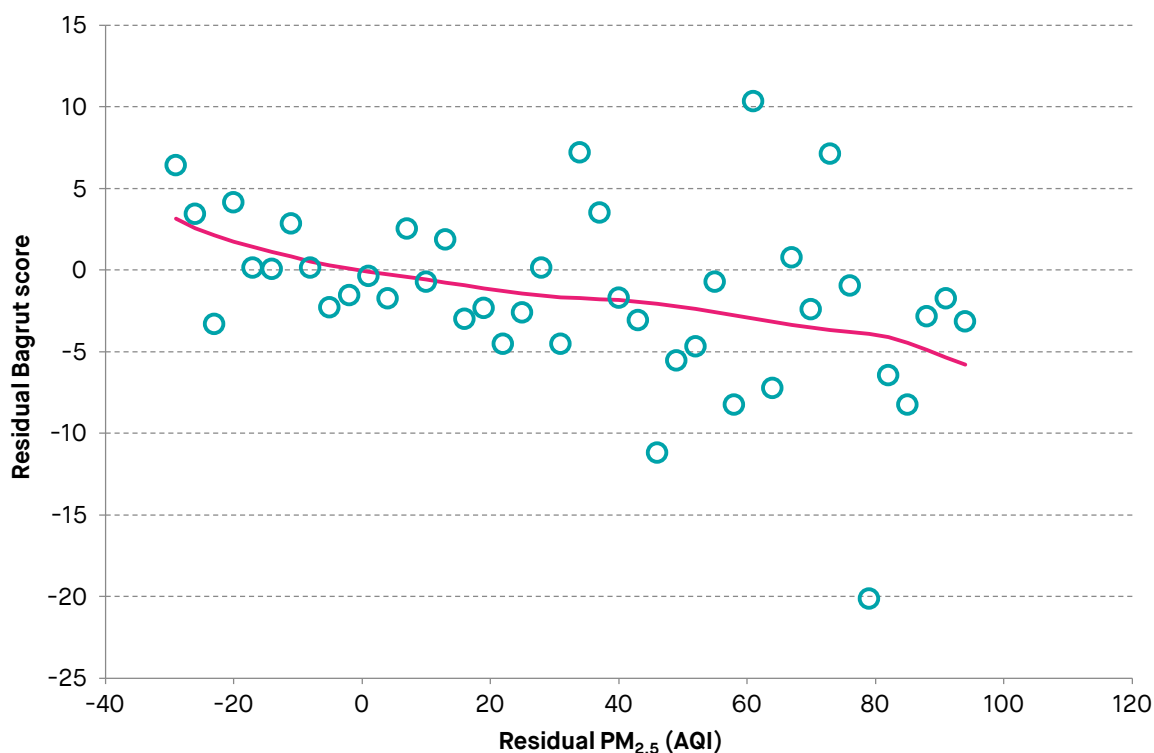
Notes: The plot reports the relationship between Residual *Bagrut* (test) scores and Residual PM$_{2.5}$ estimated by Lowess bandsmoother. Each observation on the plot is Residual *Bagrut* scores averaged over bins of width 3 units Residual PM$_{2.5}$ (AQI). Residual *Bagrut* scores and Residual PM$_{2.5}$ are generated by regressing each variable on student fixed effects.

In Table 1, we report our baseline results of the relationship between the Air Quality Indicator (AQI)[4] 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) without city, school or student fixed effects. In column (1), we estimate that a 10 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% level. The results also indicate that a relatively small part of the variation in test scores (R-squared = 0.003) is explained by air pollution, 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 10 unit increase in pollution is associated with a 0.52 decrease in a student's score.

Table 1: 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-squared	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 10 units of $PM_{2.5}$(AQI). Standard errors are heteroskedastic-consistent and clustered by school.

In columns (3)-(5) of Table 1, we take advantage of the panel structure of our data and include city, school, and student fixed effects, respectively. These account for variation in time-invariant unobserved heterogeneity that could be correlated with ambient pollution. The estimates from a regression with city or school fixed effects in columns (3) and (4), are somewhat larger, with estimated coefficients of -.70 and -0.56 respectively.

Adding student fixed effects generates similar results, with our preferred estimate indicating that a 10 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 (.040*(59.7-10.1)/23.7) standard deviations.

The effect of $PM_{2.5}$ on *Bagrut* scores for the 99th percentile of exposure in our sample (AQI=137) is very large and implies a decline of roughly 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 scores gains associated with paying teachers large financial bonuses based on their students' test scores (Lavy, 2009). Unfortunately, days with elevated levels of particulate matter are not unusual in Israel and in neighboring countries in the Middle East, as they are often the result of sandstorms that originate in the Sahara desert and are relatively common in the spring and summer months, with serious health effects (Bell, Levy and Lin 2008).

We also find two additional important findings. First, the effect is very temporary. As shown in Figure 2, the impact of pollution is pronounced the day of the exam with much smaller effects observed the day before or after. Second, 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.

Figure 2: Impact of PM$_{2.5}$ on *Bagrut* test scores in the days before/after exam
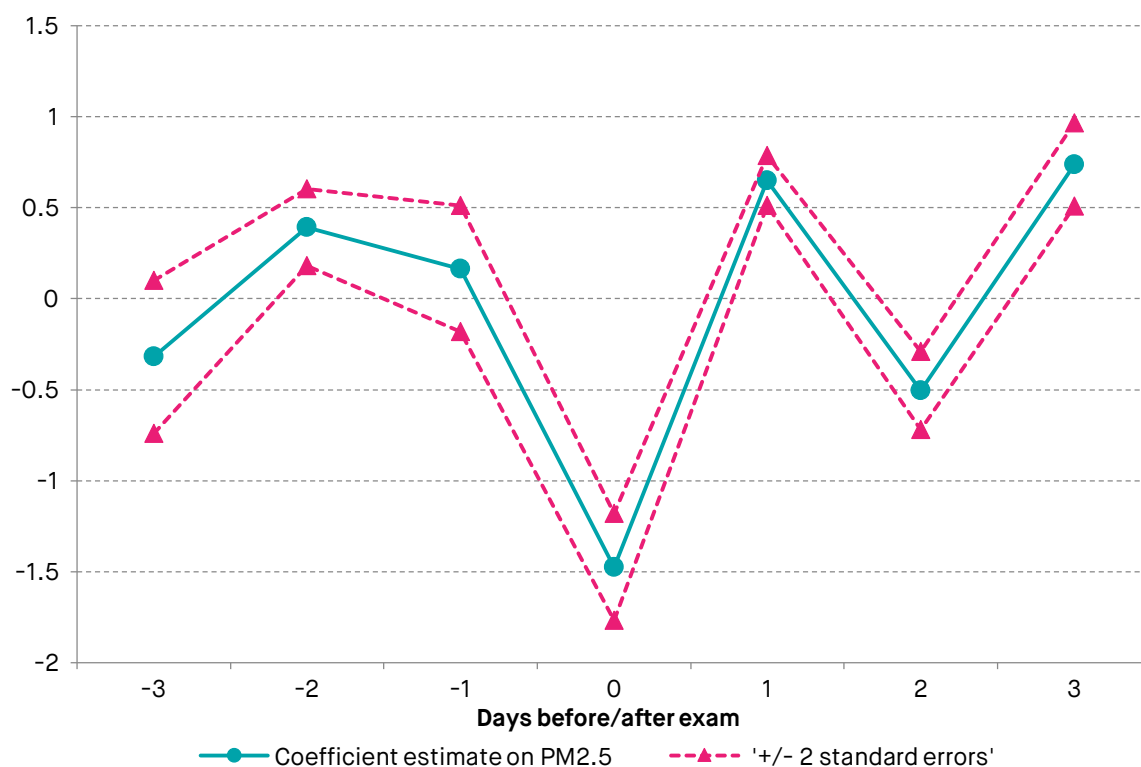
Notes: The figure plots the coefficients from a regression of *Bagrut* test scores on PM$_{2.5}$ (AQI) readings in the days prior to, the day of (Day=0), and the days following the examination, estimated in a single regression. Standard errors are clustered by school. Effects are reported in terms of change in score per 10 additional units of PM$_{2.5}$ (AQI).

The results indicate that there is indeed a strong causal link between pollution and cognitive performance. This relationship is an important policy result by itself as it implies that a narrow focus on traditional health outcomes, such as hospitalisation and increased mortality, may understate the true cost of pollution as mental acuity is essential to productivity in most professions.

THE LONG-TERM RAMIFICATIONS OF HAVING A 'BAD DAY' ON HIGH-STAKES EXAMS

Here we use the relationship between pollution and test score that we presented above to ask: what are the long run implications to a student of having a "bad day" on a high-stakes exam? We evaluate this question using the same sample of students, and their

average exposure to pollution across their exams. Since students at the same school take exams on different days, there is considerable variation in pollution exposure even after including school fixed effects.

Table 2: Particulate matter's impact on post-secondary education and adult earnings

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

Notes: Each cell in the table represents a separate regression. The table reports the relationship between average $PM_{2.5}$ (AQI) during the *Bagrut* and the listed outcome using the student-level sample described in Table 1. The coefficients are reported per 10 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.

Our results are highlighted in Table 2. We estimate that an additional 10 units of $PM_{2.5}$ (AQI) is associated with a 1.64 unit reduction in a student's composite score in our preferred specification, an estimated effect of roughly 13% (20%) of a standard deviation ($\sigma=23.7$). 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, we find that pollution exposure of an additional 10 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 .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% (relative to any estimate reported per 10 units of AQI). Since matriculation certification is required by many elite post-secondary academic institutions in Israel, it is likely that students which suffer a negative shock that lowers their certification probability will ultimately impact their prospects for higher education. In rows 5 and 6, we present the estimated effect of average pollution exposure on two longitudinal educational outcomes: enrollment in post-secondary institution (1=yes), and years of post-secondary 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 10 units of $PM_{2.5}$ (AQI). All estimates are statistically significant at the 5% level, and suggest that taking *Bagrut* exams in highly polluted days can have long-lasting effects on schooling attainment.

In row 7, we present the reduced form effect of average $PM_{2.5}$ on average monthly earnings. In our preferred specification with school fixed effects in column 3, we estimate that a student exposed during the *Bagrut* exams dates to an additional 10 units of $PM_{2.5}$ (AQI) is associated with an average monthly earnings decline at age 28 of 109 shekels ($30), or 2.1%. This estimate is also precisely estimated, with a T statistic greater than three. A visual complement to this result in presented in Figure 3, 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. Overall, these results suggest that students who take an exam during a severe pollution episode experience long run consequences, both academically and economically.

Figure 3: Scatter plot of residual $PM_{2.5}$ quantile and wages
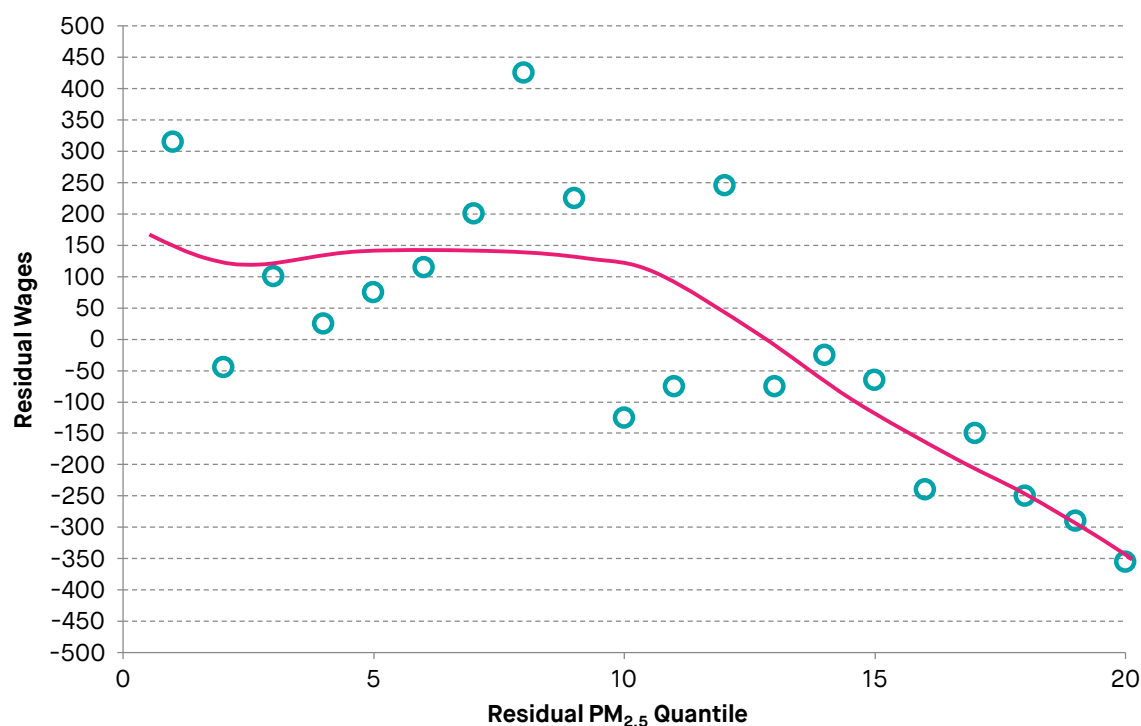
Notes: Each observation is a quantile of residual $PM_{2.5}$. Residual wages scores and Residual $PM_{2.5}$ are generated by regressing each variable on school fixed effects, and calculating the residual. The plot is generated using lowess bandsmoother.

We complement the above results by examining 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 post-secondary academic enrollment by 1.9 percentage points, post-secondary education by .092 years, and 66 shekels (or 1.3%) in wages. Interestingly, we also find there is virtually no effect of pollution on non-competitive 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 post-secondary education.

Finally, we examine heterogeneity in the return to a point on the *Bagrut* across sub-populations in Israel. 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 capitalise on the opportunity of gaining admission to a 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.

CONCLUSION

This briefing 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 analyse 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 post-secondary 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 labour 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 in the US) are administered on a single day, where random factors could materially affect a student's future. Our findings lend empirical support for the concern that there is too much reliance on high-stakes exams across educational levels, and suggest that more stable measures of student quality should be given greater weight (Lewin 2014).

REFERENCES

Angrist, J.D., and S.H. 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, and Victor Lavy. 2009. "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial." American Economic Review, 99(4): 1384-1414.

Ailshire, Jennifer and Philippa Clarke. 2015. "Fine Particulate Matter Air Pollution and Cognitive Function Among U.S. Older Adults." Journal of Gerontology B Psychological Sciences 70(2): 322-328.

Arceo, Eva, Rema Hanna, and Paulina Oliva. 2011. "Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City". Forthcoming, Economic Journal.

Bell, Michelle, 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: 104-111.

Calderón-Garcidueñas, L., Mora-Tiscareño, A., Ontiveros, E., Gómez-Garza, G., Barragán-Mejía, G., Broadway, J., Chapman, S., Valencia-Salazar, G., Jewells, V., Maronpot, RR., Henríquez-Roldán, C., Pérez-Guillé, B., Torres-Jardón, R., Herrit, L., Brooks, D., Osnaya-Brizuela, N., Monroy, M., González-Maciel, A., Reynoso-Robles, R., Villarreal-Calderon, R., Solt. A., and R. Engle. 2008. "Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs." Brain and Cognition 68(2): 117-127.

Cawley, J, J. Heckman and E. Vytlacil. 2001. "Three observations on wages and cognitive ability." Labour Economics 8:419-442.

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):1121-116

Chang, Tom, Joshua Graff Zivin, Tal Gross, Matthew Neidell. 2014. "Particulate Pollution and the Productivity of Pear Packers". NBER WP 19444.

Clark, Donald. and Louis Sokoloff. 1999. "Circulation and energy metabolism of the brain." in G. Siegel, B. Agranoff, R. Albers, S. Fisher and M. Uhler, eds., Basic Neurochemistry. Molecular, Cellular and Medical Aspects, Lippincott-Raven, pp. 637–670.

Dockery, Douglas, and C Arden Pope. 1996. "Epidemiology of Acute Health Effects: Summary of Time Series Studies." in R. Wilson and J. Spengler, eds., Particles in Our Air, Harvard University Press, 123-148.

Graff Z., J. S. and M. J. Neidell. 2012. "The Impact of Pollution on Worker Productivity." American Economic Review 102:3652-3673.

Lavy, Victor, Avraham Ebenstein and Sefi Roth. 2014 "The Long Run Human Capital and Economic Consequences of High-Stakes Examinations." NBER WP 20647.

Lavy, Victor, Avraham Ebenstein and Sefi Roth. 2014 "The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation." NBER WP 20648.

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

Mills, Nicholas, Ken Donaldson, Paddy W Hadoke, Nicholas A Boon, William MacNee, Flemming R Cassee, Thomas Sandström, Anders Blomberg and David E Newby. 2009. Nature Clinical Practice Cardiovascular Medicine 6(1):36-44.

Pope, C. Arden. III, David Bates, and Mark Raizenne. 1995. "Health Effects of Particulate Air Pollution: Time for Reassessment?" Environmental Health Perspectives 103: 472-480.

Pope, C. Arden III, and Douglas Dockery. 2006. "Critical Review–Health effects of fine particulate air pollution: Lines that connect." Journal of the Air and Waste Management Association 56:709-742.

Wilker, Elissa, Sarah Preis, Alexa Beiser, Philip Wolf, Joel Schwartz, Petros

Koutrakis, Rhoda Au, Sudha Seshadri, Murray Mittelman. 2015. "Long-Term Exposure to Fine Particulate Matter, Residential Proximity to Major Roads and Measures of Brain Structure." Stroke 46(5):1161-1166.

ENDNOTES

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

[2] 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 2014). 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 EPA and empirically, CO is not highly correlated with $PM_{2.5}$, suggesting the bias of focusing exclusively on $PM_{2.5}$ is limited.

[3] 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 3 unit bins of residual $PM_{2.5}$ (AQI). We then examine the relationship between residual scores and pollution using lowess bandsmoother.

[4] We report our results using AQI so they can more easily be interpreted in terms of air quality, where 100 is the EPA standard for unhealthy for sensitive groups.