

Hot Temperature and High Stakes Exams: Evidence from New York City Public Schools

Jisung Park*

March 20, 2018

Abstract

Understanding the link between temperature and educational outcomes is important in assessing the efficiency and equity implications of various schooling interventions and the potential welfare impacts of climate change. Using student-level administrative data for the largest public school district in the United States, I estimate the causal impact of hot temperature on high-stakes exam performance and subsequent educational attainment. Hot days reduce contemporaneous performance by up to 14% and lead to persistent impacts on high school graduation status. An analysis of teacher grade manipulation provides the first available evidence for ex post compensatory responses to hot temperature.

Keywords: temperature, education, climate change, adaptation

JEL Codes: I21, O18, Q54, Q56

*Park: Harvard Kennedy School of Government and University of California, Los Angeles, jisung-park@luskin.ucla.edu. Acknowledgements: The author would like to thank Larry Katz, Andrei Shleifer, Robert Stavins, Joe Aldy, Geoffrey Heal, Raj Chetty, Claudia Goldin, Edward Glaeser, Melissa Dell, Michael Kremer, Josh Goodman, Jonah Rockoff, Jeff Miron, Max Auffhammer, Olivier Deschenes and numerous seminar participants at Harvard, Columbia, UCLA, Duke, Maryland, UCSB, UC Berkeley, Oxford, IZA, Seoul National, the NYC Department of Health, the NBER Summer Institute and the Bill and Melinda Gates Foundation for valuable comments and feedback. Thanks also to the NYC Department of Education for data access, and to Nicolas Cerkez and Rodrigo Leal for excellent research assistance. All remaining errors are my own. Funding from the Harvard Environmental Economics Program, the National Science Foundation, the Harvard Climate Change Solutions Fund, and the Harvard Kennedy School of Government Environment and Natural Resources Program are gratefully acknowledged.

1 Introduction

Whether and how temperature enters the human capital production function is an important yet unresolved economic question. While emerging evidence finds that extreme temperature can have adverse biological effects on students (Schoer and Shaffran, 1973; Graff Zivin et al., 2017), the nature and magnitude of associated welfare impacts remain unclear. This is in large part due to the fact that welfare will depend not only on the biological relationship between temperature and student outcomes, but also the interaction between educational policies and environmental shocks as well as defensive investments made in response to such shocks (Graff Zivin and Neidell, 2014). This paper’s primary objective is to shed light on the potential interactions between hot temperature and educational policies and how they might affect the realized impact of ambient temperature on student welfare.

Ambient environmental conditions in schools simultaneously constitute a global externality and a local public good problem. The increased frequency and severity of heat waves associated with climate change are determined by global carbon emissions, and thus beyond any individual student’s control. In principle, adaptations locally (e.g. school air conditioning) might mitigate the realized educational impacts of hotter ambient temperature, regardless of whether the underlying climate externality has been resolved globally. However, even these more local defensive investments often require collective action, and students may be constrained in the avoidance behaviors they can take conditional on a given level of local public good provision.¹ Importantly, given the ubiquity of high stakes exams that act as hurdles between various stages of the human capital production process, it seems possible that even short instances of heat exposure may have lasting educational consequences.

To explore the impact of hot temperature on educational outcomes, I link local daily weather data to test scores of 1 million New York City public high school students taking synchronized high stakes exams, as well as to administrative data on subsequent high school graduation status. Student fixed effects regressions identify the causal impact of heat exposure on exam performance and eventual educational attainment by exploiting quasi-random variation in temperature across multiple exam dates and times. The research design is based on a simple premise: that within-student variations in day-to-day temperature are not caused by unobserved determinants of educational performance.

The setting allows for an analysis of how temperature affects student outcomes in a high

¹For instance, in most school environments, it is costly for a student to skip a class or an exam because of inclement weather, unless there is a school-sponsored coordination mechanism across students and teachers. The extent to which individual students can climate-control their learning spaces also seems limited given current technologies. Even when air conditioning exists at the school or classroom level, there may be gaps between individual student needs and realized ambient temperature, due, for instance, to underlying medical conditions such as obesity or asthma.

stakes, centrally administered, high income setting. This has three major advantages. Because the exams are high stakes and data can be linked to graduation status, it is possible to assess persistent impacts on educational attainment. Because exams cannot be rescheduled and dates and times are set over a year in advance, I am able to effectively shut down the channel of student selection on the extensive margin of exam taking. Because the New York City public school system represents one of the most resource intensive public learning environments in the world, it seems likely that, unlike many developing country settings, the primary constraints to avoidance behavior will not be income, meaning that results can shed light on potential non-market barriers to adaptation and defensive investment.²

The paper presents three primary findings. The first main result is that hot temperature during high-stakes exams exerts a causal and economically meaningful impact on student performance. Each New York City public school student takes a series of mandatory exams in June which are spread over the course of two weeks and feature harmonized timing and pre-determined testing sites. Because I am able to link multiple exam records for each student and school location, and to match these records to local ambient temperature on the day of each subject exam, the analyses presented here likely identify the causal impact of hot temperature on contemporaneous exam performance. Hot temperature during an exam results in reduced performance: an approximately linear decline of -0.2 percentiles per °F above room temperature (72°F). This implies that taking an exam on a 90°F day reduces performance by 14 percent of a standard deviation relative to a more optimal 72°F day. For a sense of magnitude, the within-school Black-White achievement gap is approximately 25 percent of a standard deviation. At least 18% of the students in the study sample experience an exam with ambient temperatures exceeding 90°F.

Second, I find that heat exposure during these exams subsequently affects a student's chances of graduating from high school. Depending on the degree of institutional flexibility, the costs of retaking exams, or the presence of dynamic complementarities in the human capital production process, short instances of heat stress may or may not have lasting economic consequences. Consistent with the inflexible administration (i.e. no rescheduling) and high stakes nature of these exams, I find that hot temperature during a test reduces a student's likelihood of graduating from high school on time or at all. For the median student, taking an exam on a 90°F day leads to a 10.9% lower likelihood of passing a particular subject (e.g. Algebra), which in turn affects probability of graduation.³ An analysis of exam-time temperature exposure and graduation status that controls for number of exam takes reveals

²At approximately \$18,000, per-pupil spending in NYC public schools was the highest in the United States as of 2015 (Census-Bureau, 2017).

³Heat exposure also substantially reduces chances of achieving key performance thresholds that are used by local universities in college admissions decisions.

that one standard deviation increase in average exam-time temperature reduces a student’s likelihood of graduating on time by roughly 2.5 percentage points. This is despite the fact that students are able and often encouraged to retake failed exams during the ensuing summer and following school years, suggesting potential dynamic complementarities as in Cunha and Heckman (2007) or Diamond and Persson (2016).⁴

Third, I provide evidence of ex post compensatory behavior by teachers, who appear to selectively upward manipulate grades for students who score just below passing thresholds and who experienced particularly hot exam sittings. Using a subject, school, and date-specific bunching estimator at pass-fail cutoffs adapted from previous work (Dee et al., 2016), and relating the extent of bunching to temperature on the day of an exam, I show that teachers manipulated grades more frequently for hot exam takes. The amount of excess bunching is beyond what would result from mechanical correlation between temperature-induced performance declines and an increase in the proportion of scores in the manipulable zone, suggesting teachers are responding to hot exam-day temperature.

While it is difficult to infer teachers’ intentions, these patterns are consistent with a pedagogical view that transitory score shocks are not reflective of underlying human capital, and that the resulting educational and economic consequences would be inefficient and/or unfair. This ex post behavior is also consistent with the possibility that, in many public schools, there may be non-trivial constraints to ex ante defensive investments such as classroom air conditioning. According to 644 school-level air conditioning records scraped from the web, I find that fewer than 62% of NYC public schools had any form of air conditioning as of 2012, compared to air conditioning penetration rates of 90% or above in private residential and commercial buildings.

This paper contributes specifically to the literature on temperature and human capital. Recent work finds that hot ambient temperature can affect student cognition and potentially the rate of learning. Hot temperature reduces student performance on low-stakes (10-minute) cognitive assessments administered in US homes (Graff Zivin et al., 2017), and similar low-stakes assessments in India (Garg et al., 2017). Cumulative heat exposure during the summer months has been shown to potentially reduce winter-time exam performance in Korea (Cho, 2017). This study is the first to study the contemporaneous impact of hot temperature on high stakes exam performance, and to document precise and significant effects.⁵ It is also the

⁴Persistent impacts of shocks to high stakes exam scores have been found in the context of air pollution (Ebenstein et al., 2016) and teacher grade manipulation (Dee et al., 2016; Diamond and Persson, 2016). For instance, Ebenstein et al. (2016) find that air pollution exposure during high-stakes exams leads to lower post-secondary schooling attainment and reduced earnings.

⁵In addition to the potential for selective sorting based on unobservable student characteristics, survey-based analyses such as Graff Zivin et al. (2017) face an additional challenge due to the fact that hot temperature may lead to systematic under-reporting of data by administrators. For instance, a sub-

first to document persistent impacts on educational attainment, and the first to document ex post compensatory behavior. The finding that short-run environmental shocks during high stakes exams can have persistent educational and economic consequences echoes findings from Ebenstein et al. (2016), who study air pollution in Israel, and Isen et al. (2017), who study in-utero heat shocks and their impact on later-life outcomes.⁶ This study however is the first to link short-run heat exposure during exams to long-run human capital outcomes, which has distinct implications for optimal carbon policy and educational policy as described below.

The findings are also relevant for the literature on dynamic complementarities in human capital investment. They are consistent with a world in which a hot exam day nudges some students to achieve less schooling overall due to institutional rigidities and opportunity costs of time similar to those documented by Dee et al. (2016) and Ebenstein et al. (2016), or dynamic complementarities in the human capital production process whereby idiosyncratic shocks discourage students or lead to subsequent tracking by teachers and peers (Diamond and Persson, 2016; Shah and Steinberg, 2017).

Finally, this paper is broadly related to a growing literature exploring the welfare impacts of hotter temperature and potential adaptations to such impacts. Hot temperature has been found to affect GDP per capita at the macro level (Hsiang, 2010; Dell et al., 2012; Heal and Park, 2013), as well as health (Deschênes and Greenstone, 2011; Anderson et al., 2013; Barreca et al., 2016), labor supply (Graff Zivin and Neidell, 2014), and labor productivity (Deryugina and Hsiang, 2014; Behrer and Park, 2017) at the micro level. Recent studies also explore adaptive responses to such impacts, and show that existing technologies can be effective if properly utilized. For instance, air conditioning has been shown to reduce health and labor productivity impacts considerably (Barreca et al., 2016; Behrer and Park, 2017), but such responses have also been shown to be limited by low income levels and higher marginal utility of consumption (Davis and Gertler, 2015) or liquidity constraints (Gertler et al., 2016). This study builds on these findings by exploring the effect of hot temperature on human capital production in a setting where one might expect such income or liquidity constraints not to be first order. The fact that air conditioning levels are so much lower than private market settings and that teachers appear to engage in such strategic ex post compensatory behavior may be indicative of additional constraints to efficient adaptation in

stantial proportion of NLSY surveys are missing cognitive (PIAT) assessments, or show incomplete reports, which may be due to heat-fatigued surveyors selectively skipping sections of the assessment. (See: <https://www.nlsinfo.org/content/cohorts/nlsy97/topical-guide/education/piat-math-test>).

⁶In related work using weather shocks as instruments for educational impacts, Goodman (2014) shows that snowfall can result in disruptions to learning by increasing absenteeism selectively across different student groups. A growing number of studies explores the impact of air pollution on student outcomes (Currie et al., 2009; Roth, 2016), and consistently find large impacts on absenteeism and exam performance.

school settings. More research is needed to better understand the nature and magnitude of such potential constraints, and the policy responses they may or may not justify.

I discuss specific welfare and policy implications in the conclusion. In brief, the findings suggest that ambient temperature may be an important factor to consider when designing education policy or allocating public resources, especially in contexts where heat exposure is frequent, high-stakes exams pose hurdles to further schooling, and where differentials in realized classroom temperature can be substantial. This paper also provides empirical support for the view that a hotter climate may adversely affect human capital accumulation, which has potential implications for the optimal social cost of carbon. Importantly, the findings underscore the importance of the interaction between environmental stress and educational institutions – which determine exam format as well as to a large degree the built environment experienced by students – as opposed to the presence or absence of hot temperature *per se*.

The rest of this paper is organized as follows. Section 1 presents relevant stylized facts and a simple conceptual framework that guides the empirical analysis. Section 2 describes the data and institutional context and presents key summary statistics. Section 3 presents the main results and various sensitivity analyses. Section 4 discusses implications and concludes.

2 Stylized Facts and Conceptual Framework

2.1 Relevant Stylized Facts

Three stylized facts from the existing literature are of relevance in thinking about the impact of temperature on human capital production. First: heat stress directly affects physiology in ways that can be detrimental to cognitive performance. Second: most individuals demonstrate a revealed preference for mild temperatures close to room temperature, commonly taken to be between 65°F and 74°F (18°C and 23°C). Third: the inverted U-shaped relationship between temperature and performance documented in the lab has been confirmed in a range of field settings including mortality and labor supply, but evidence of impacts on human capital remains thin, particularly in high-stakes school environments.⁷

⁷See Dell et al. (2014) and Heal and Park (2016) for reviews of the related literature. See Mackworth (1946); Seppanen et al. (2006) on the physiology of heat exposure, Roback (1982); Sinha et al. (2015) for examples of hedonic analyses and the revealed preference for mild temperatures, and Grether (1973); Sudarshan and Tewari (2013); Graff-Zivin and Neidell (2012) for impacts of temperature on labor and task productivity.

2.2 Conceptual Framework and Empirical Predictions

The basic conceptual framework is a Mincerian human capital model where the marginal value of student (and/or teacher) effort is a function of temperature, and both exam score and cumulative human capital stock can be affected. The intuition is that students can invest time and effort e in order to accumulate human capital h . At any given point in time t , hot temperature, T_t (expressed as a deviation from optimum or room temperature), may affect the productivity of this investment $a(T)$, and may affect both a given exam score $s_t(T_t)$ as well as the overall amount of learning achieved over a relevant time period $h_T(\sum_t^T T_t)$.

The main empirical predictions from the model, which is described more formally in the appendix, are as follows:

1. We expect acute heat exposure for student i in time t to reduce contemporaneous exam performance, $\frac{\Delta s_{it}}{\Delta T_{it}} < 0$, if any of (a) direct flow utility, (b) marginal cost of effort, or (c) cognitive performance are adversely affected by temperature, since they may affect time-use, effort, and productivity of time/effort respectively.
2. Short instances of heat exposure during an exam should not in principle reduce the stock of human capital h_{it} , but might nevertheless reduce overall human capital attainment $\frac{\Delta h_{it+\tau}}{\sum_t^{t+\tau} \Delta T_{it}} \leq 0$, if the schooling environment features high costs of retaking exams, or in the presence of dynamic complementarities due to policies such as tracking.
3. Any combination of (a) information asymmetries between students and school administrators or teachers and school administrators, (b) coordination problems in adjusting the timing of exams, or (c) liquidity constraints or collective action problems in the context of air conditioning or school infrastructure investment, can lead to socially inefficient levels of ex ante avoidance behavior. Tiebout sorting on income can moreover lead to a correlation between neighborhood income and extent of infrastructure investment, depending on the nature in which public school financing decisions are made (Tiebout, 1956; Cellini et al., 2010).

3 Institutional Context, Data, and Summary Statistics

3.1 New York City Public Schools

The New York City public school system (NYCPS) is the largest in the United States, with over 1 million students as of 2012.⁸ The average 4-year graduation rate, at 68%, is below the

⁸Approximately one fifth of NYC students attend private schools which are not included in our sample.

national average but comparable to other large urban public school districts. System-wide averages mask considerable discrepancies in achievement across neighborhoods. Schools in the predominantly Black or Hispanic neighborhoods of Brooklyn and the Bronx have four-year graduation rates as low as 35% per year (Figure 3a and 3b).

3.2 New York State Regents Exams

Each June, students in the state of New York take a series of high-stakes exams called “Regents exams”. These standardized subject assessments are administered by the New York State Education Department (NYSED) and are harmonized in administration across the state.

Regents exams carry important consequences. Students are required to meet minimal proficiency status – usually a scale score of 65 out of 100 – in five “core” subject areas to graduate from high school.⁹ Many local universities including City University of New York (CUNY) use strict Regents score cutoffs in the admissions process as well: for instance, requiring that students score above 75 on English and Math simply to apply. These exams are therefore pivotal for the median student in determining high school diploma eligibility and college admissions.

The vast majority of students take their Regents exams during a pre-specified two-week window in mid-to-late June each year.¹⁰ The test dates, times, and locations for each of these Regents exams are fixed over a year in advance by the NY State education authority (NYSED), and synchronized across schools in the NYC public school system to prevent cheating. Each exam is approximately 3 hours long, with morning and afternoon sessions each day, and are taken at the student’s home school.¹¹ Figures 1a and 1b provide a sample exam schedule and cover sheet.

All exams are written by the same state-administered entity and scored on a 0-100 scale, with scaling determined by subject-specific rubrics provided by the NYSED in advance of

⁹The core subject areas are English, Mathematics, Science, U.S. History and Government, and Global History and Geography. The passing threshold is the same across all core subjects. Students with disabilities take separate RCT exams, and are evaluated on more lenient criteria.

¹⁰For any given student, exam takes are spread out across multiple days and years though, in effect, most exams are taken junior and senior year. Apart from the fact that most students take English their junior year, and Living Environment and Global History prior to other “advanced” sciences and US History respectively, there do not seem to be any clear patterns in the timing of subjects throughout students’ high school careers. Some advanced students may take Regents subject exams during middle school or during early January waves, while failing students are required to retake exams in August.

¹¹Throughout the study period, students typically took Regents exams at the school in which they were enrolled unless they required special accommodations which were not available at their home school. Students who fail their exams (or are deemed unready by their teachers to progress to the next grade) are required to attend summer school, which occurs in July and August.

the exams each year. All scores are therefore comparable across schools and students within years, and the scaling designed in such a way that is not intended to generate a curve based on realized scores.¹² I use standardized performance at the subject level as the primary measure of exam performance in this study, though the results are robust to using scale scores. While centrally administered, exams were locally graded by teachers in the students' home schools, usually on the evening of the associated subject exam.

In summary, using scores from Regents exams offers several distinct advantages empirically. First, they are high-stakes exams used to determine diploma eligibility and possibly affecting college enrollment. This means that, in addition to direct welfare relevance, they may also provide information about compensating behavior that is not available in laboratory studies or voluntary cognitive assessments such as those in the NLSY. Second, they are offered at a time of year when temperatures fluctuate considerably, resulting in substantial variation within a relatively small geographic locale. Finally, Regents exams were centrally administered and compulsory for all public school students during the study period, meaning there is relatively little possibility of anticipatory alteration of exam timing based on weather forecasts, or for bias due to selection into taking the exam.

3.3 Student Outcome Data

I obtain individual exam-level information from the New York City Department of Education (NYC DOE). This includes records for the universe of NYC public school students who took one or more Regents exams over the period 1999 to 2011.¹³ Information on exam dates comes from archived Regents exam schedules from the NYC DOE database, which provide date and time information for each subject by year and month of administration. Graduation status by student is available in a separate data file, which can be linked to exam records using unique 10-digit student identifiers. These records include cohort and school information, as well as graduation status 4, 5, and 6 years post-matriculation, including the type of diploma received and whether the student dropped out. A detailed description of the matching procedures and subsequent sample restrictions are provided in the online appendix.

¹²In principle, scores are comparable across years as well, as psychometricians in the NYSED conduct difficulty assessments of each year's subject exams and engage in "equating" procedures prior to their release. The primary identification of short-run impacts include year fixed effects, and thus do not rely on this cross-year comparability.

¹³I also use data from standardized math and English language and arts (ELA) exams administered in 3rd through 8th grade from NYC DOE to provide a measure of previous ability. Specifically, I calculate the average combined z-score of each student for whom previous standardized ELA and math exam records are available. These exams are generally taken in January and March.

3.4 Weather Data

Weather data comes from the National Oceanic and Atmospheric Administration’s Daily Global Historical Climatology Network, which provides daily min, max, and mean temperatures, precipitation and dew point information from a national network of several thousand weather stations over the period 1950-2014. I take daily minimum and maximum temperature as well as daily average precipitation and dewpoint readings from the 5 official weather stations in the NYC area that provide daily data for the entirety of the sample period (1998-2011). I match schools to the nearest weather station, one for each of the five boroughs: The Bronx, Brooklyn, Manhattan, Queens, Staten Island. To account for spatial heterogeneity in outdoor temperature due to urban heat island effects, I also assign spatial correction factors generated by satellite reanalysis data. I impute test-time temperature – for instance, average outdoor temperature between 9:15am to 12:15pm for morning exams – by fitting a fourth-order polynomial in hourly temperature.¹⁴ Given existing evidence on the impact on air quality on student performance, I include controls for pm2.5 and ozone, taken from EPA monitoring data from Manhattan.¹⁵

3.4.1 School Air Conditioning Information

Information on building-level air conditioning equipment comes from records originally compiled by New York City School Construction Authority (SCA), which administers detailed, building-level surveys for NYC public schools. While a centralized database was not publicly available, a web-scrape of individual school websites resulted in matchable records for 644 middle and high school buildings in the study sample. The records include information on air conditioning equipment presence and maintenance status as of the year 2012. Unfortunately, the data does not provide AC installation status by year, nor does it provide information regarding where within a school AC was available or whether existing units were actually utilized. As such, this data does not lend itself well to precise comparative analysis of the effectiveness of AC. The primary purpose of this data is therefore to provide descriptive

¹⁴Further details regarding these corrections are presented in the online appendix. The primary results reported below are not sensitive to either of these corrections. The corrections reduce standard errors but leave implied point estimates relatively unchanged.

¹⁵Air pollution in NYC during this period is relatively low, compared, for instance, to the levels found to affect Israeli student performance (Ebenstein et al., 2016). The maximum recorded value of pm2.5 in my data is 38 micrograms per cubic meter, compared to readings that regularly went above 120 micrograms per cubic meter in Ebenstein et al. (2016). The air quality controls used here are nevertheless crude, especially for localized pollutants such as ozone. Given the focus of the study, the relatively low levels of particulate matter during the sample period, and the high correlation between ozone and summertime temperature, I run analyses with and without controls for air quality but do not attempt to separately identify or interpret causal effects of fine particulates or ozone.

analyses, though I present regression results that use school-level AC installation status as of 2012 as a rough proxy for classroom AC utilization in the appendix.

3.5 Summary Statistics

The final working dataset consists of 4,509,102 exam records for 999,582 students. It includes data from 91 different exam sessions pertaining to the core Regents subjects over the 13 year period spanning the 1998-1999 to 2010-2011 school years.

Table 4 presents summary statistics for the key outcome variables that form the basis of this analysis. The student body is 40% Latino, 31% African American, 14% Asian and 13% White, and approximately 78% of students qualify for federally subsidized school lunch. The median student scores just around the passing cutoff, with a score of 66 (sd = 17.9), though there is considerable heterogeneity by neighborhood as well as demographic group.

Students take on average 7 June Regents exams over the course of their high school careers, and are observed in the Regents data set for roughly 2 years, though some under-achieving students are observed for more than 4 years, as they continue to retake exams upon failing. Fewer than 0.2% of students are marked as having been absent on the day of the exam, corroborating the high-stakes, compulsory nature of these exams.

Figure 2 illustrates the source of identifying variation for short-run temperature impacts, with temperatures weighted by exam observation and school location. Outdoor temperature during exams range from a low of 60°F to a high of 98°F. Day-to-day variation within the June exam period can be considerable, as suggested by Figure 2b, which shows the variation in outdoor temperature by school and exam take across two consecutive test dates within the sample period.

4 Empirical Strategy and Primary Results

4.1 Hot Temperature and High Stakes Exam Performance

Figure 5a presents a visual depiction of performance and temperature that motivates the analysis that follows. It shows a binned scatterplot of standardized exam score by percentile of observed exam-day temperature, plotting residual variation after controlling for school fixed effects and average differences across subjects and years. Exams taken on hot days clearly exhibit lower scores.

To further isolate the causal impact of short-run temperature fluctuations on student performance, I exploit quasi-random variation in day-to-day temperature across days within student-month-year cells, focusing on the main testing period in June. While it is unlikely

that temperature is endogenous to student behavior, nor is it likely for students to select into different temperature treatments given the rigidity of exam schedules, time-varying unobservables may still be correlated with weather realizations. For instance, if certain subjects tend to be scheduled more often in the afternoon when students are relatively fatigued (as in Sievertsen et al. (2016)) or toward the end of the exam period (Thursday as opposed to Monday), we may expect mechanical correlation between temperature and test scores that is unrelated to the causal effect of temperature on student cognition or effort. This motivates a baseline specification that includes year, time of day, and day of week fixed effects:

$$Y_{ijsty} = \gamma_{iy} + \eta_s + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \beta_3 Time_{sty} + DOW_{sty} \beta_4 + \epsilon_{ijsty} \quad (1)$$

Here, Y_{ijsty} denotes standardized exam performance for student i taking an exam in subject s in school j on date t in year y . The terms γ_{iy} and η_s denote student-by-year and subject fixed effects respectively. T_{jsty} is the outdoor temperature in the vicinity of school j during the exam (subject s on date t , year y). X_{jsty} is a school- and date-specific vector of weather and air quality controls, which include precipitation, dewpoint, and ozone. $Time_{sty}$ represents a dummy for time of day (morning versus afternoon, $Time=1$ denotes an afternoon exam), and DOW_{sty} represents a vector of fixed effects for each day of the week in which exams were taken.

Student-by-year fixed effects ensure that we are comparing the performance of the same student across different exam sittings within the same testing window, some of which may be taken on hot days, others not, leveraging the fact that the average student takes 7 June Regents exams over the course of their high school career (between 3 and 4 per year). Subject fixed effects control for persistent differences in average difficulty across subjects. Year fixed effects control for possible spurious correlation between secular performance improvements and likelihood of hotter exam days due, for instance, to climate change.¹⁶

Table 5b presents the results from running variations of equation (4) for the subset of

¹⁶To the extent that temperature variation within student-month-year cells are uncorrelated with unobserved factors influencing test performance, one would expect the coefficient β_1 to provide an unbiased estimate of the causal impact of temperature on exam performance, subject to attenuation bias due to measurement error in weather variables as well as downward bias from positive grade manipulation. It is worth noting that measurement error could in principle be non-classical in a way that biases the estimates upward. For instance, if the average classroom has more students in lower performing schools, experienced classroom temperature scales non-linearly with outdoor temperature, and students in lower performing schools are more susceptible to heat stress, then β_1 may actually be biased upwards. Given relative homogeneity in average class sizes across the city, this seems second-order.

students who take at least 2 exams in any given year.¹⁷ As suggested by the first row of columns (1)-(4), exam-time heat stress exerts a significant causal impact on student performance. The estimates are robust to allowing for arbitrary autocorrelation of error terms within boroughs and test dates, which is the level of exogenous temperature shock recorded in the data.

Taking an exam on a hot day reduces performance by approximately -0.008 standard deviations ($se=0.002$) per °F. This amounts to -5.2 percent of a standard deviation in performance per standard deviation increase in temperature, or -13.5 percent of a standard deviation if a student takes an exam on a 90°F day as opposed to a more optimal 72°F day.¹⁸

The effect of a 90°F day is thus comparable in magnitude to roughly 1/4 of the Black-White score gap, or 3/4 of the within-school Black-White score gap. This effect is similar to the impacts on mathematical reasoning found by Graff Zivin et al. (2017), who find a 90°F day to reduce NLSY math scores by approximately -0.12 standard deviations.

A series of robustness checks, including models that replace student-by-year fixed effects with student- or school-by-year fixed effects, are presented in the online appendix. The point estimates using the school-by-year fixed effects specification are slightly larger (more negative) on average, and remain statistically significant. Also presented in the appendix are heterogeneity analyses by gender and ethnicity. I find relatively little evidence of heterogeneity by demographic groups, though it is possible that adaptive responses by teachers are offsetting impacts disproportionately for certain subgroups.

These results provide strong evidence that hot temperature affects contemporaneous student performance, either by reducing raw cognitive ability or by increasing the disutility of effort which in turn affects students' willingness to maintain focus during a three-hour exam. They suggest that temperature in the learning environment plays an important role in determining student outcomes, and that whatever compensatory effort is exerted by students due to the high stakes nature of some exams may not be enough to offset the physiological impacts of temperature documented in the lab.

¹⁷Results using un-standardized scale scores as the dependent variable are presented in the appendix. In models where student fixed effects are replaced by school fixed effects, a vector of demographic control variables are included to control for possible selection into subjects by student type.

¹⁸Precipitation has a slightly positive effect, and ozone has a negative but insignificant effect, with a 1 standard deviation increase in ozone corresponding to a point estimate roughly 1/5th the size of a 1 standard deviation temperature effect. Despite previous literature documenting adverse impacts of pm2.5 in Israel (Ebenstein et al., 2016), I find little evidence for that here, perhaps because average concentrations of pm2.5 are much lower in NYC than in Israel, as well as the fact that the performance impacts documented by Ebenstein et al. (2016) are highly non-linear, driven mostly by heavily polluted days with pm2.5 above 100 micrograms per cubic meter.

4.2 Persistent Impacts on Educational Attainment

Heat exposure during an exam, while reducing cognitive ability or concentration temporarily, presumably does not reduce the stock of knowledge or human capital – at least not immediately through the physiological impact of heat stress itself.

In a friction-less world with fully flexible educational institutions, students who perform poorly on a subject due to a hot exam sitting could immediately and costlessly retake the exam until she believes her “true ability” has been reflected in the exam score (i.e. until $s_{it} = h_{it}$). In this world, random heat exposure during exams should not affect the final amount of schooling achieved. However, in the presence of institutional rigidities that limit the effective number of possible retakes or impose time and effort costs to retaking an exam, it is possible that even short-run heat exposure can have ripple effects on long-run educational attainment. For example, requiring students to attend remedial courses may lead some to drop out early. Employers may treat students who graduated from high school in five or six years differently from those who graduated “on-time” for a variety of reasons, reducing the pecuniary return to education once a subject has been failed the first time. Exam scores may serve as important signals within the education system – to the student herself, to her peers, or to her parents and teachers – leading to dynamic complementarities in human capital investment (Cunha and Heckman, 2007).¹⁹

Effects of Heat Exposure on Pass Rates and College Proficiency Status

In NYC, students must score a 65 or above to pass a given subject exam and thus have it count toward receiving a high school diploma.²⁰ This cutoff does not change based on the realized distribution of performance in any given year. Students are also assigned “proficient” (i.e. college ready) status on each of the subjects in which they receive a grade of 75 or higher and “mastery” status for scores of 85 or higher. Beyond any personal motivational or within-school signalling value, these designations carry real weight externally in the sense that many

¹⁹Recent evidence suggests persistent effects of temporary score shocks in the context of teacher manipulation (Dee et al., 2016; Diamond and Persson, 2016) and air pollution (Ebenstein et al., 2016), with as yet inconclusive evidence regarding the specific mechanisms by which they occur. Dee et al. (2016) find substantial impacts of upward score manipulations on graduation status, especially for students who scored in the manipulable zone. Using administrative records from Swedish middle schools, Diamond and Persson (2016) also find substantial effects of upward score manipulations on subsequent performance, graduation likelihood, and later life income. Ebenstein et al. (2016) find that Israeli high school students who receive lower scores on their Bagrut (high school exit) exams due to air pollution are less likely to receive Bagrut certificates (comparable to high school diplomas) and receive lower wages later in life.

²⁰As mentioned previously, there are some exceptions to this rule. Until 2005, low-performing students were allowed the option of applying to receive a “local diploma” which required scores of 55 and above for exams to count toward the diploma. In the following regressions, I use the more stringent and universally accepted standard of “Regents Diploma” as the definition of passing score. Results of running the regression analyses below using the “Local Diploma” cutoff feature similar (slightly more negative) point estimates.

local colleges and universities such as City University of New York (CUNY) use strict score cutoffs in their admissions decisions.

To estimate the impact of contemporaneous heat stress on the likelihood that a student scores at or above the passing and proficiency thresholds, I run variations of the following models:

$$p_{ijsty} = \gamma_{iy} + \eta_s + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \beta_3 Time_{sty} + DOW_{sty} \beta_4 + \epsilon_{ijsty} \quad (2)$$

$$c_{ijsty} = \gamma_{iy} + \eta_s + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \beta_3 Time_{sty} + DOW_{sty} \beta_4 + \epsilon_{ijsty} \quad (3)$$

where p_{ijsty} is a dummy variable indicating whether student i passed – that is, scored a 65 or above on – subject s on date t , year y , and c_{ijsty} is a dummy variable indicating college proficiency status: i.e., a dummy for scores at or above 75 points.

The results suggest that acute heat exposure can have significant short term impacts on student performance, with potentially lasting consequences. Exam-time heat stress reduces the likelihood of passing by 0.31 (se=0.12) percentage points per °F, or -0.54% per °F from a mean likelihood of 0.57. These results are presented in tabular form in the appendix. Impacts on the likelihood of achieving proficiency status are slightly larger in aggregate, with a magnitude of -0.96% per °F hotter exam-time temperatures relative to a mean likelihood of 0.32. Unless higher-ability students are more sensitive to heat stress, this discrepancy seems to be driven in part by grade manipulation around the passing threshold.

Taken together, these estimates suggest that experiencing hot ambient temperatures during a high stakes exam can have non-trivial consequences for student performance, with a 90° day leading to approximately 9.7% lower chance of passing a given exam, and a 17.4% lower probability of achieving proficiency status for the average student.

Effects of Heat Exposure on Graduation Status

These short-run heat exposures appear to reduce final educational attainment by affecting the likelihood that students graduate from high school. We might expect this to occur either through mechanical links between pass rates and graduation requirements, or through dynamic complementarities given the option to retake failed exams in the following fall or spring. Figure 7a plots variation in 4-year graduation status against average exam-time temperature, and provides suggestive evidence of such persistent impacts.

Whereas short-run impacts of hot temperature could be identified *within* student cells, long-run impacts on graduation status cannot because the outcome variable is no longer

date-specific. This poses some additional challenges to precise estimation of causal impacts. Computing a measure of average heat exposure across multiple exam sittings by student results in mechanical correlation between average experienced temperature and the number of exams such that students who take more exams are more likely to be assigned average temperature values closer to the climatic mean in that month, and students who take fewer exams are more likely to be assigned extreme values.²¹

The comparison of interest is therefore the difference in graduation likelihood between students who, conditional on the number of draws from the climate distribution, experience different amounts of heat stress. One way to accomplish this is to compare within exam-count and year-count cells, controlling for observable factors. Collapsing the data to the student level, I estimate variations of the following model:

$$g_{ijcn} = \alpha_0 + \alpha_1 \overline{T}_{ij} + X_{ij}\alpha_2 + \chi_j + \theta_c + Z_i\alpha_3 + exams_n\alpha_4 + \epsilon_{ijc} \quad (4)$$

Here, g_{ijcn} is a dummy denoting whether student i in school j and entering cohort c who takes n June Regents exams over the course of her high school career has graduated after 4 years in high school. \overline{T}_{ij} denotes the average temperature experienced by student i while taking June Regents exams in school j , up through her senior year. X_{ij} is a vector of weather controls averaged at the student-by-school level. χ_j denotes school fixed effects. θ_c denotes cohort fixed effects. Z_i is a vector of student-level controls including race, gender, federally subsidized school lunch eligibility, and where applicable scores from previous standardized exams. $exams_n$ denotes a vector of fixed effects for the number of June exam takes.

The parameter of interest is α_1 , which captures the impact of an additional degree of heat exposure during exams on the likelihood of graduating on time. School fixed effects account for potential omitted variable bias due to unobserved determinants of graduation rates being correlated with average temperature in the cross-section (e.g. if urban heat island effects are stronger in poorer neighborhoods). Cohort fixed effects in graduation rates allow for the possibility that heat exposure and graduation rates are correlated due to secular trends in both variables – though warming trends and average improvements in NYC schools would suggest this effect to lead to downward rather than upward bias in the estimate of α_1 .

Table 7a presents the results from running variations of equation 4 with and without school and cohort fixed effects, as well as flexible controls for the number of exams. Standard

²¹That is, assuming that the average June climate in New York City can be represented by a distribution of daily temperature realizations, the average temperature across multiple days will exhibit a form of mean-reversion as one increases the number of draws from the underlying climate distribution (see online appendix for a visual representation).

errors are clustered at the borough by date and time level, based on the intuition that this conservatively approximates the level of quasi-random temperature variation, though the results are once again robust to alternative levels of clustering.

Columns (1)-(3) suggest that a 1 degree F increase in average exam-time temperatures is associated with a 0.71 (se=0.17) to 0.76 (se=0.22) percentage point decline in the likelihood of graduating on time. A one standard deviation in average exam-time temperature (+4.4°F) leads to a 3.12 to 3.34 percentage point decline in the likelihood of on-time graduation, or 4.59% to 4.91% decline relative to a mean on-time graduation rate of 68 percent.

These effects are large. Even without correcting for adaptive grading by teachers, I estimate that, over the period 1998 to 2011, upwards of 510,000 exams that otherwise would have passed received failing grades, affecting the on-time graduation prospects of at least 90,000 students. These estimates are described in greater detail in the appendix.

4.3 Avoidance Behaviors and Defensive Investments

Avoidance behaviors and defensive investments are a key component of the welfare impact arising from any environmental shock, which makes understanding the mechanisms and magnitude of such responses important for policy design (Courant and Porter, 1981; Graff-Zivin et al., 2011; Graff Zivin and Neidell, 2013). While laboratory studies often force exposure onto subjects to estimate a pure biological effect, in real-world settings, individuals can take actions to limit their exposure ex ante or mitigate the damage done ex post.

If hot temperature affects student outcomes, theory would predict avoidance behaviors and defensive investments along the most cost-effective margins. However, the choice set faced by students may be constrained by a variety of market and non-market factors, including the aggregation of local families' preferences regarding local public school investments. Students themselves are often constrained additionally by the fact that they are usually minors and thus depend on legal adults to make various decisions on their behalf, or on effective coordination across other students, teachers, and administrators. A typical secondary school student cannot skip classes or exams on a hot day without costly repercussions. Similarly, she most often cannot install an air conditioner in her classroom, even if she could afford such private provision of public goods financially. These factors, in addition to health and safety regulations in many US public schools, suggest that the set of options available to students may be constrained relative to other market contexts.²²

²²For instance, in the case of New York City public schools, air conditioners must meet efficiency standards and be obtained from and installed by a specific vendor chosen by the city, in addition to receiving city approval with regard to a variety of safety regulations, contractual obligations and energy considerations. In some cases, school "sustainability" policies prohibit administrators from investing in new infrastructure unless it can be demonstrated that it has a net neutral impact on carbon emissions, a bar that new air

Ex Post Compensatory Behavior: Teacher Responses

Using a similar student-level dataset from NYC public schools, Dee et al. (2016) document systematic grade manipulation by teachers on State Regents exams. They find that most of the manipulating behavior occurred at or around the passing margin of 65 and that, while varied in magnitude across schools and student types, such manipulation was a near-universal phenomenon within the NYCPS system.

Upon careful analysis of competing explanations, the authors suggest the most likely explanation to be the goodwill of teachers who seek to offset the impact of “a bad test day”:

In sum, these estimates suggest that manipulation was unrelated to the incentives created by school accountability systems, formal teacher incentive pay programs, or concerns about high school graduation. Instead, it seems that the manipulation of test scores may have simply been a widespread “cultural norm” among New York high schools, in which students were often spared any sanctions involved with failing exams, including retaking the test or being ineligible for a more advanced high school diploma (pg 27).

A hot test day may be viewed as a bad test day, particularly if air conditioning is inadequately provided. In that case, it seems possible for discretionary grade manipulation to have been a response to perceived performance impacts of heat stress. Teachers may be able to observe or at least intuit the disruptive impacts of elevated classroom temperatures on test day, especially since exams are taken in students’ home schools and graded by a committee of teachers from that school. If benevolently motivated, they may be inclined to engage in more grade manipulation precisely for those exams that took place under unusually hot conditions. Even if teachers do not actively intend to offset heat-related performance impacts, it is possible that such manipulation may, in effect, blunt some of the idiosyncratic effects of weather on student performance.

Estimating Compensatory Grading: Bunching at Score Thresholds

Figure 8 provides a histogram of Regents scale scores in all core subjects prior to 2011. As is clearly visible in the graph, there is substantial bunching at the passing kinks, especially at scores of 65 and 55, suggesting upward grade manipulation.

We would expect any form of grade manipulation for students who initially score just below the passing cutoff, even indiscriminate grade manipulation uncorrelated with exam-time temperature, to downward attenuate the estimates of heat-related performance impacts uncovered above. Indeed, running equation 1 on the subset of grades that fall within the manipulable zone, I find that the point estimate for the impact of temperature is substantially

conditioning cannot clear unless electricity is obtained completely from renewable sources.

reduced and no longer significant: β_T equals -0.0007 (se=0.0024) as opposed to -0.0082 (se=0.0021) in the full sample.

To assess the presence and magnitude of “compensatory grading”, I estimate a bunching estimator by school, subject, month, and year: in effect, the level of exam-time temperature variation. Starting with the student-exam level data, I calculate the fraction of observations in each 1 point score bin from 0 to 100 by core Regents subject. I then fit a polynomial to these fractions by subject, excluding data near the proficiency cutoffs with a set of indicator variables, using the following regression:

$$F_{ks} = \sum_{i=0}^q \psi_{ismyj} (Score)^i + \sum_{i \in -M_{cs}, +M_{cs}} \lambda_{ismyj} \mathbb{1}[Score = i] + \epsilon_{ksmyj} \quad (5)$$

where F_{ks} denotes the fraction of observations with score k for subject s (e.g. ELA), q is the order of the polynomial, and $-M_{cs}, +M_{cs}$ represent manipulable ranges below and above the passing thresholds. The subscripts m, y and j denote, month, year, and school respectively.

Following Dee et al, (2016), I define a score as manipulable to the left of each cutoff if it is between 50 - 54 and 60 - 64, and manipulable to the right if it is between 55 - 57 and 65 - 67 as a conservative approximation of their subject-and-year-specific scale score-based rubric. In practice, I use a fourth-order polynomial ($q=4$) interacted with exam subject s , but constant across years for the same exam subject. Realized bunching estimates are not sensitive to changes in the polynomial order or whether one allows the polynomial to vary by year or subject.²³

This generates a set of predicted fractions by score and subject. The average amount of bunching observed in my data is similar to that documented by Dee et al. (2016), who find that approximately 6% of Regents exams between 2003 and 2011 exhibited upward grade manipulation. I then calculate observed fractions for each score from 0 to 100 by school, month, year, and subject, and generate a measure of bunching that integrates the differences between observed and predicted fractions: that is, summing the excess mass of test results that are located to the right of the cutoff (above the predicted curve) and the gaps between predicted and observed fractions of test results to the left of the cutoff (below the predicted curve). The bunching estimator can be written as:

$$\zeta_{smyj} = \frac{1}{2} \sum_{i \in +M_{ck}} (F_{ks} - \hat{F}_{ksmyj}) + \frac{1}{2} \left| \sum_{i \in -M_{ck}} (F_{ks} - \hat{F}_{ksmyj}) \right| \quad (6)$$

²³As a robustness check, I also estimate a linear approximation of the above estimator by generating predicted fractions using a linear spline between boundary points along the distribution that are known to be outside the manipulable range by subject. I then generate an estimate of the extent of bunching by school-subject-month-year cell, taking the absolute value of the distance between observed and predicted fractions by Regents scale score. The results are similar using this simplified measure of bunching.

where ζ_{smj} denotes the degree of bunching at the passing cutoff for subject s , month m , year y , and school j .

I then examine the relationship between ζ_{smj} and exam-time temperature in that cell, which corresponds to the temperature experienced by students taking subject s in school j in June of year y , with controls for precipitation and humidity.

To assess the magnitude of this relationship controlling for school-, subject-, and/or year-level differences in the degree of manipulation that are unrelated to temperature, I run a series of regressions with ζ_{smj} as the dependent variable:

$$\zeta_{smj} = \delta_0 + \delta_1 T_{smj} + \delta_2 X_{smj} + \chi_j + \eta_s + f(\text{Year}_y) + \epsilon_{smj} \quad (7)$$

where T_{smj} denotes temperature, X_{smj} denotes precipitation and humidity, χ_j , η_s , and θ_y denote school, subject, and year fixed effects respectively, and $f(\text{Year}_y)$ denotes a cubic time trend in scores. The parameter of interest is δ_1 , which represents the increase in grade manipulation due to exam-time temperature.

The amount of bunching increases by approximately 0.10 to 0.16 percentage points per degree F, or 1.7% to 2.8% per degree F hotter exam-time temperature. Coefficients are positive and significant in specifications with and without school and year fixed effects. This relationship is depicted graphically in figure 9a. The short-run performance impacts documented above will, in this sense, be net of compensatory grading.²⁴ The implied magnitudes are non-trivial. The difference in overall share of exams manipulated between a 90°F day and a 72°F day can be as much as 50%, suggesting temperature fluctuations may represent a large component of the variation in extent of grade manipulation throughout the period.

It is possible that, due to the distributional properties of most Regents exams, heat-related performance impacts may lead to a mechanical increase in the number of grades that fall within the manipulable zone. This could in principle lead to a correlation between bunching behavior and exam-time temperature. To account for this potential mechanical correlation, I run the above analysis replacing the dependent variable with the fraction of manipulable scores actually manipulated.

²⁴This likely results in a smaller point estimate than otherwise would have been the case. The only case in which the bias may be upward is if teachers grade differentially and punitively according to the temperatures *they* experience while grading, and temperature during the exam is correlated with temperature during grading. If hot temperatures make teachers less productive and causes more errors, this will simply add noise to the score variable. If hot temperature makes teachers irritable and more punitive in grading, then we might expect the beta coefficient to be picking up some of the correlation between test day temp and grading punitive-ness, although the most striking feature of the score distribution as described below is that the majority of grade manipulation seems to be positive in direction, making this unlikely in practice.

Figure 9b presents a binned scatterplot of the bunching estimator and exam-time temperature by subject-month-year-school cell, adding school fixed effects to allow for arbitrary differences in the average amount of grade manipulation across schools, and expressing the bunching estimate as a proportion of scores within the manipulable zone (50-54, 60-64). It suggests a clear positive relationship between the degree of grade manipulation and the ambient temperature during the exam being graded. It is consistent with teachers boosting students' grades more often when students experienced hot testing environments.²⁵

While it is not possible to infer teachers' intentions based on these results alone, these results are highly suggestive of adaptive grading as a compensatory response to hot temperature. Teachers may have an intuitive sense of whether a particular student scored below his or her "true ability" – regardless of whether or not they perceive this to be due to temperature or other exam-time conditions – and respond by manipulating grades in the case of students on the passing margin. Irrespective of whether teachers' explicit intentions are to compensate for heat-related impacts, however, the realized effect has been for this behavior to mitigate the adverse welfare impacts associated with exam-time heat exposure.

Potential Constraints to Ex Ante Adaptive Investments

Consistent with market and non-market constraints to efficient ex ante investments, public school air conditioning seems to have been far from complete. According to NY Building Construction Assessment Surveys (BCAS), 62% of the NYC public school buildings for which data are available were reported as having some kind of air conditioning equipment on its premises as of 2012. This means that at least 38% of these buildings did not have any form of air conditioning equipment available (Figure 10). Moreover, of the 62% that were reported as having air conditioning, 42% were cited as having defective components, according to the third-party auditors conducting the BCAS assessments.

Such air conditioning penetration and maintenance levels are lower than what might be expected given private demand in the region. Air conditioning penetration for commercial and residential buildings in New York City are above 90%, and even in the poorest neighborhoods average 70% or higher. While summer vacation means that occupancy may be lower for schools during hotter months, a large fraction of public school students are required to attend summer school, suggesting substantial building occupancy throughout the year. Public

²⁵The temperature in the grading environment may in principle be a potential confounding factor, since it is not observed. According to teacher reports, most exams seem to have been graded during the evening of that exam. It is therefore possible that temperature during grading is correlated with temperature during exams, if teacher's lounges and offices do not have functioning AC. Unless hotter temperature makes teachers more likely to exert such additional effort to identify threshold students and manipulate grades however, it seems unlikely that these patterns are a result of temperature directly affecting teacher mood or cognition, though I cannot rule out this possibility.

records and local media reports suggest that air conditioning is raised repeatedly as a matter of contention during teachers union strikes. Parents and teachers in New York City have signed petitions as recently as 2013 asking the school district to invest in air conditioning infrastructure.²⁶

Because school level AC data is only available as a cross-section in 2012 and does not provide information regarding utilization rates, a detailed comparison of how temperature affects students in schools with and without AC was not feasible. However, using available data as a coarse proxy for AC utilization between 1998 and 2011, I find some noisy but suggestive evidence that school air conditioning can be effective when utilized. These results are presented in the Appendix. Because cross-sectional variation in AC status is not experimental, and likely correlated with many potential confounding factors, I am hesitant to interpret these associations as more than suggestions consistent with existing work (Barreca et al., 2016). The most striking aspect of the data on air conditioning is just how much lower AC penetration in public schools seems to be relative to what might be expected given local climate and income levels.

5 Discussion and Conclusion

This paper explores the impact of hot temperature on high stakes exams, subsequent educational attainment, and avoidance behaviors. Using administrative data from the largest public school district in the United States, I find that hot temperatures exert a causal and economically meaningful impact on student outcomes by reducing performance on high-stakes exams. The research design exploits quasi-random, within-student temperature variation to identify the impact of hot days on performance. These short-run impacts, which presumably do not reduce the stock of human capital, nevertheless result in persistent impacts on educational attainment as measured by high school graduation status. This is despite a pattern of what appears to be ex post compensatory behavior by teachers who upward manipulate borderline scores for exams taken under hot conditions.

Taking an exam on a 90°F day results in a 14% of a standard deviation reduction in exam performance relative to a more optimal 72°F day, controlling for student ability. These short-run performance impacts lead to persistent impacts on educational attainment. A 90°F day

²⁶Such phenomena are not limited to New York City schools. During a major teacher union strike in Chicago in 2012, “Timetable for air conditioning” was listed as one of four major contract demands (an agreement to provide universal air conditioning in Chicago public schools was not reached until 2016). Media reports of inadequate AC in New York City and elsewhere abound. See: <http://www.denverpost.com/2011/09/08/heat-related-illnesses-spur-petition-for-sept-school-start-in-denver/>; and New York Times: <https://mobile.nytimes.com/2015/06/24/nyregion/new-yorks-public-school-students-sweat-out-the-end-of-the-semester.html>.

results in a 10.9% lower probability of passing a subject, and, for the average New York City student, a 2.5% lower likelihood of graduating on time, despite the ability of students to retake failed exams. I estimate that, over the period 1998 to 2011, upwards of 510,000 exams that otherwise would have passed received failing grades due to hot temperature, affecting at least 90,000 students, possibly many more.

Teachers seem to have selectively boosted grades of students who experienced hot exam sittings. Using a variant of bunching estimators developed in previous work (Dee et al., 2016), I find a pattern of grade manipulation that is systematically related to temperature during the exam, even when controlling for potential mechanical correlation between temperature and the fraction of manipulable scores. One interpretation is that teachers may have tried to offset a portion of the long-term consequences of idiosyncratic environmental shocks such as hot test days.²⁷

Such responses are consistent with relatively low levels of school air conditioning, which likely constitute *ex ante* defensive investments that can require overcoming substantial coordination problems. Based on a web-scrape of building level engineering surveys, I find that fewer than 62% of public schools in NYC had any form of air conditioning as of 2012, and that 40% of those that do are reported as having defective components. This is compared to average residential and commercial sector air conditioning penetration rates of well over 90%, suggesting possible institutional constraints to *ex ante* defensive investments.

The findings presented in this paper have several policy implications. First, they suggest that ambient temperature may be an important variable to consider when designing education policy: for instance, in determining how much weight to put on high stakes exams, either for student advancement, teacher promotion, or school funding decisions (Chetty et al., 2014; Jacob and Rothstein, 2016). Similarly, in prioritizing various policy options aimed at reducing achievement gaps, environmental conditions such as climate – or school infrastructure investments that may mediate the relationship between climate and realized temperature in the learning environment – might play a larger role than previously estimated, given the interactive nature of the payoffs (Cellini et al., 2010; Jackson et al., 2015; Lafortune et al., 2016).

From an equity standpoint, it may be important to consider whether students taking high stakes standardized exams across varying geographies and built environments are on a level climatic playing field. Such fairness concerns may be especially important for nationally and sometimes internationally harmonized examinations such as the SAT, GRE, and LSAT in

²⁷A possible unintended consequence of eliminating teacher discretion in New York City public schools in 2011 may have been to expose more low-performing students to climate-related human capital impacts, eliminating a protection that applied predominantly to low-achieving Black and Hispanic students.

the United States, as well as analogous exams in other countries. Unless schools and homes in hotter regions have access to perfectly offsetting amounts of ex ante defensive investments or adaptation capital, individuals taking nationally or internationally standardized exams such as the SAT or ACT in a hotter region may be placed at a disadvantage relative to their peers in cooler regions.

Moreover, as the span of geographies covered by a standardized exam widens, the potential for materially different climatic conditions – and associated differences in test-day temperature – increases. The SAT, for instance, is taken more or less simultaneously not only across the fifty United States but also across countries as diverse as Bangladesh, China, Swaziland, Ukraine, and Venezuela. The affordability of air conditioning may be a binding constraint in many developing economies considering the well documented relationship between income and air conditioning ownership at the household level (Davis and Gertler, 2015) and well-documented liquidity constraints in the context of energy-intensive appliance demand (Gertler et al., 2016). Such factors may also be relevant in thinking about the persistence of racial achievement gaps in the United States, given correlations between race, income, and average climate across neighborhoods within the U.S.

These findings also suggest that climatic factors may interact with existing educational systems in a way that reduces allocative efficiency. Hot exam days may add noise to the signal-extraction process of high-stakes testing, leading to inefficiencies in labor and higher education markets, as documented by Ebenstein et al. (2016). Whether there are cost-effective ways to mitigate such persistent impacts, and whether the social benefits of high stakes standardized testing outweigh these and other costs remains to be seen.

From the perspective of climate policy, this study suggests that current social cost of carbon estimates may omit important elements of the climate damage function: especially those mechanisms, including human capital accumulation, that may affect the rate of growth as opposed to the level of economic activity (Pindyck, 2013; Heal and Park, 2016), though more careful research on the impact of cumulative heat exposure on the pace of learning is needed. To the extent that future climate change may result in a disproportionate increase in realized heat exposure for the poor within and across countries, these findings lend further support to the notion that climate change may have unequal impacts across the income distribution (Hallegatte et al., 2018).

Finally, this paper raises new questions in an old debate regarding geography and economic prosperity (Acemoglu et al., 2001; Rodrik et al., 2004). How much of the variation in student achievement across and within countries (Figures 11a and 11b) can be explained by the cumulative influence of temperature stress on learning? Is it possible that hotter, poorer countries are subject to more challenging baseline learning conditions due to a combination

of hot climate, lack of protective capital, and inflexible institutions? More careful research is needed to answer these questions, particularly regarding the links between cumulative heat exposure, school infrastructure, and human capital accumulation. It is worth noting that while the average New Yorker is exposed to approximately 14 days above 90°F per year, the average resident of New Delhi experiences over 80 such days annually, with climate models projecting up to 150 such days per year in New Delhi by 2050.

References

- Acemoglu, Daron, Simon Johnson, and James A Robinson (2001), “The colonial origins of comparative development: An empirical investigation.” *The American Economic Review*, 91, 1369–1369.
- Anderson, G Brooke, Francesca Dominici, Yun Wang, Meredith C McCormack, Michelle L Bell, and Roger D Peng (2013), “Heat-related emergency hospitalizations for respiratory diseases in the medicare population.” *American journal of respiratory and critical care medicine*, 187, 1098–1103.
- Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S Shapiro (2016), “Adapting to climate change: The remarkable decline in the us temperature-mortality relationship over the twentieth century.” *Journal of Political Economy*, 124, 105–159.
- Behrer, A Patrick and Jisung Park (2017), “Will we adapt? temperature, labor and adaptation to climate change.”
- Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein (2010), “The value of school facility investments: Evidence from a dynamic regression discontinuity design.” *The Quarterly Journal of Economics*, 125, 215–261.
- Census-Bureau, U.S. (2017), “Public education finances: 2015.”
- Chetty, Raj, John N Friedman, and Jonah E Rockoff (2014), “Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood.” *The American Economic Review*, 104, 2633–2679.
- Cho, H (2017), “Effect of summer heat on test scores: A cohort analysis.” *Journal of Environmental Economics and Management*, 83, 185–196.
- Courant, Paul N and Richard C Porter (1981), “Averting expenditure and the cost of pollution.” *Journal of Environmental Economics and Management*, 8, 321–329.
- Cunha, Flavio and James Heckman (2007), “The technology of skill formation.” *American Economic Review*, 97, 31–47.
- Currie, Janet, Eric A Hanushek, E Megan Kahn, Matthew Neidell, and Steven G Rivkin (2009), “Does pollution increase school absences?” *The Review of Economics and Statistics*, 91, 682–694.

- Davis, Lucas W and Paul J Gertler (2015), “Contribution of air conditioning adoption to future energy use under global warming.” *Proceedings of the National Academy of Sciences*, 112, 5962–5967.
- Dee, Thomas S, Will Dobbie, Brian A Jacob, and Jonah Rockoff (2016), “The causes and consequences of test score manipulation: Evidence from the new york regents examinations.” *NBER Working Paper*.
- Dell, Melissa, Benjamin F Jones, and Benjamin A Olken (2012), “Temperature shocks and economic growth: Evidence from the last half century.” *American Economic Journal: Macroeconomics*, 4, 66–95.
- Dell, Melissa, Benjamin F Jones, and Benjamin A Olken (2014), “What do we learn from the weather? the new climate–economy literature.” *Journal of Economic Literature*, 52, 740–798.
- Deryugina, Tatyana and Solomon M Hsiang (2014), “Does the environment still matter? daily temperature and income in the united states.” *NBER Working Paper*.
- Deschênes, Olivier and Michael Greenstone (2011), “Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us.” *American Economic Journal: Applied Economics*, 3, 152–185.
- Diamond, Rebecca and Petra Persson (2016), “The long-term consequences of teacher discretion in grading of high-stakes tests.” *NBER Working Paper*.
- Ebenstein, Avraham, Victor Lavy, and Sefi Roth (2016), “The long-run economic consequences of high-stakes examinations: evidence from transitory variation in pollution.” *American Economic Journal: Applied Economics*, 8, 36–65.
- Fahle, Erin M, Benjamin R Shear, Demetra Kalogrides, Sean F Reardon, Richard DiSalvo, and Andrew D Ho (2017), “Stanford education data archive.”
- Garg, Teevrat, Maulik Jagnani, and Vis Taraz (2017), “Human capital costs of climate change: Evidence from test scores in india.”
- Gertler, Paul J, Ori Shelef, Catherine D Wolfram, Alan Fuchs, et al. (2016), “The demand for energy-using assets among the world’s rising middle classes.” *American Economic Review*, 106, 1366–1401.
- Goodman, Joshua (2014), “Flaking out: Student absences and snow days as disruptions of instructional time.” *NBER Working Paper*.
- Graff Zivin, Joshua, Solomon M Hsiang, and Matthew Neidell (2017), “Temperature and human capital in the short-and long-run.” *Journal of the Association of Environmental and Resource Economists*.
- Graff-Zivin, Joshua and Matthew Neidell (2012), “The impact of pollution on worker productivity.” *The American Economic Review*, 102, 3652–3673.
- Graff Zivin, Joshua and Matthew Neidell (2013), “Environment, health, and human capital.” *Journal of Economic Literature*, 51, 689–730.

- Graff Zivin, Joshua and Matthew Neidell (2014), “Temperature and the allocation of time: Implications for climate change.” *Journal of Labor Economics*, 32, 1–26.
- Graff-Zivin, Joshua, Matthew Neidell, and Wolfram Schlenker (2011), “Water quality violations and avoidance behavior: Evidence from bottled water consumption.” *The American Economic Review: Papers and Proceedings*, 101, 448–453.
- Grether, WF (1973), “Human performance at elevated environmental temperatures.” *Aerospace Medicine*, 44, 747–755.
- Hallegatte, Stephane, Jisung Park, Mook Bangalore, and Evan Sandhoefner (2018), “Households and heat stress: estimating the distributional consequences of climate change.” *Environment and Development Economics*.
- Heal, Geoffrey and Jisung Park (2013), “Feeling the heat: Temperature, physiology & the wealth of nations.” *NBER Working Paper*.
- Heal, Geoffrey and Jisung Park (2016), “Temperature stress and the direct impact of climate change: A review of an emerging literature.” *Review of Environmental Economics and Policy*, 10, 347–362.
- Hsiang, Solomon M (2010), “Temperatures and cyclones strongly associated with economic production in the caribbean and central america.” *Proceedings of the National Academy of sciences*, 107, 15367–15372.
- Isen, Adam, Maya Rossin-Slater, and Reed Walker (2017), “Relationship between season of birth, temperature exposure, and later life wellbeing.” *Proceedings of the National Academy of Sciences*, 201702436.
- Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico (2015), “The effects of school spending on educational and economic outcomes: Evidence from school finance reforms.” *The Quarterly Journal of Economics*, 131, 157–218.
- Jacob, Brian and Jesse Rothstein (2016), “The measurement of student ability in modern assessment systems.” *Journal of Economic Perspectives*, 30, 85–108.
- Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach (2016), “School finance reform and the distribution of student achievement.” Technical report, National Bureau of Economic Research.
- Mackworth, Norman H (1946), “Effects of heat on wireless operators.” *British journal of industrial medicine*, 3, 143.
- Pindyck, Robert S (2013), “Climate change policy: What do the models tell us?” *Journal of Economic Literature*, 51, 860–872.
- Roback, Jennifer (1982), “Wages, rents, and the quality of life.” *Journal of political Economy*, 90, 1257–1278.
- Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi (2004), “Institutions rule: the primacy of institutions over geography and integration in economic development.” *Journal of economic growth*, 9, 131–165.

- Roth, Sefi (2016), “The contemporaneous effect of indoor air pollution on cognitive performance: Evidence from the uk.” *Unpublished Manuscript*.
- Schoer, L and J Shaffran (1973), “A combined evaluation of three separate research projects on the effects of thermal environment on learning and performance.” *ASHRAE Transactions*, 79, 97–108.
- Seppanen, Olli, William J Fisk, and QH Lei (2006), “Effect of temperature on task performance in office environment.” *Lawrence Berkeley National Laboratory*.
- Shah, Manisha and Bryce Millett Steinberg (2017), “Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital.” *Journal of Political Economy*, 125, 527–561.
- Sievertsen, Hans Henrik, Francesca Gino, and Marco Piovesan (2016), “Cognitive fatigue influences students performance on standardized tests.” *Proceedings of the National Academy of Sciences*, 113, 2621–2624.
- Sinha, Paramita, Maureen L Cropper, and Martha Caulkins (2015), “Household location decisions and the value of climate amenities.” *NBER Working Paper*.
- Sudarshan, Anant and Meenu Tewari (2013), “The economic impacts of temperature on industrial productivity: Evidence from indian manufacturing.”
- Tiebout, Charles M (1956), “A pure theory of local expenditures.” *Journal of political economy*, 64, 416–424.

Figures and Tables

EXAMINATION SCHEDULE: JUNE 2016

Students must verify with their schools the exact times that they are to report for their State examinations.

June 1 WEDNESDAY	June 14 TUESDAY	June 15 WEDNESDAY	June 16 THURSDAY	June 17 FRIDAY	June 20 ^a MONDAY	June 21 TUESDAY	June 22 WEDNESDAY	June 23 THURSDAY
9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	RATING DAY
Algebra II (Common Core) *	RE in Global History & Geography	Living Environment	Algebra I (Common Core)	Physical Setting/ Earth Science Algebra 2/ Trigonometry	RCT in Mathematics* <small>* Suggested date for administering locally developed tests aligned to the Checkpoint A and Checkpoint B learning standards for languages other than English (LOTE).</small>	Physical Setting/ Chemistry RCT in Global Studies*	RCT in Writing	
1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	Uniform Admission Deadlines Morning Examinations: 10:00 a.m. Afternoon Examinations: 2:00 p.m.
SPECIAL ADMINISTRATION: Integrated Algebra	RE in English Language Arts (Common Core)	RE in U.S. History & Government	Comprehensive English	Geometry (Common Core)	RCT in U.S. History & Government*	Physical Setting/ Physics RCT in Reading*	RCT in Science*	

* Available in Restricted Form only. Each copy of a restricted test is numbered and sealed in its own envelope and must be returned, whether used or unused, to the Department at the end of the examination period.

(a) Sample Regents Exam Schedule

The University of the State of New York
 REGENTS HIGH SCHOOL EXAMINATION

ALGEBRA 2/TRIGONOMETRY

Friday, June 19, 2015 — 9:15 a.m. to 12:15 p.m., only

Student Name: _____

School Name: _____

The possession or use of any communications device is strictly prohibited when taking this examination. If you have or use any communications device, no matter how briefly, your examination will be invalidated and no score will be calculated for you.

1 Which list of ordered pairs does *not* represent a one-to-one function?

- (1) (1, -1), (2, 0), (3, 1), (4, 2)
- (2) (1, 2), (2, 3), (3, 4), (4, 6)
- (3) (1, 3), (2, 4), (3, 3), (4, 1)
- (4) (1, 5), (2, 4), (3, 1), (4, 0)

Use this space for computations.

2 The terminal side of an angle measuring $\frac{4\pi}{5}$ radians lies in Quadrant

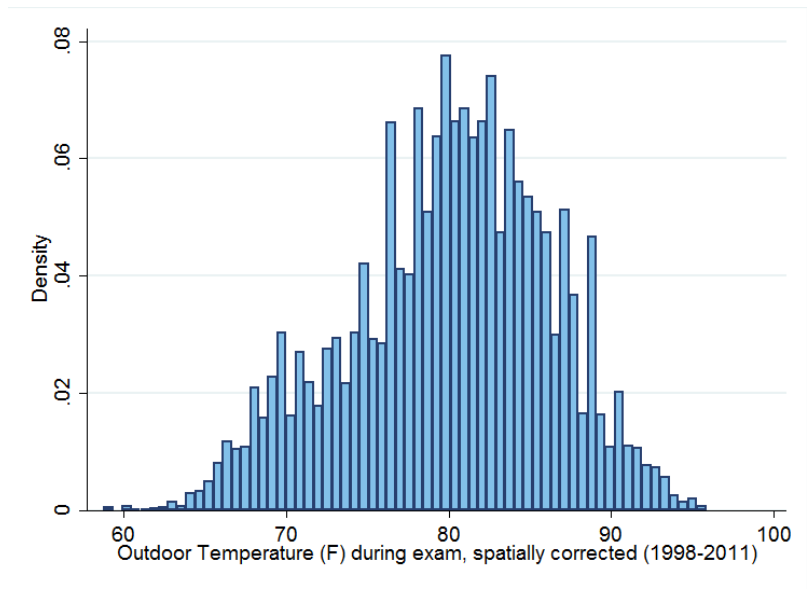
- (1) I
- (2) II
- (3) III
- (4) IV

3 If $f(x) = 2x^2 + 1$ and $g(x) = 3x - 2$, what is the value of $f(g(-2))$?

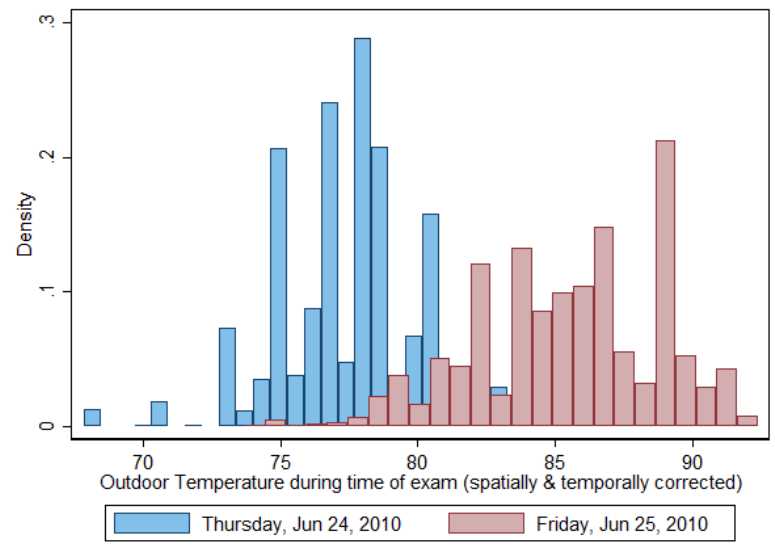
- (1) -127
- (2) -23
- (3) 25
- (4) 129

(b) Sample Regents subject exam cover sheet and questions

Figure 1: Sample Exam Schedule and Cover Page



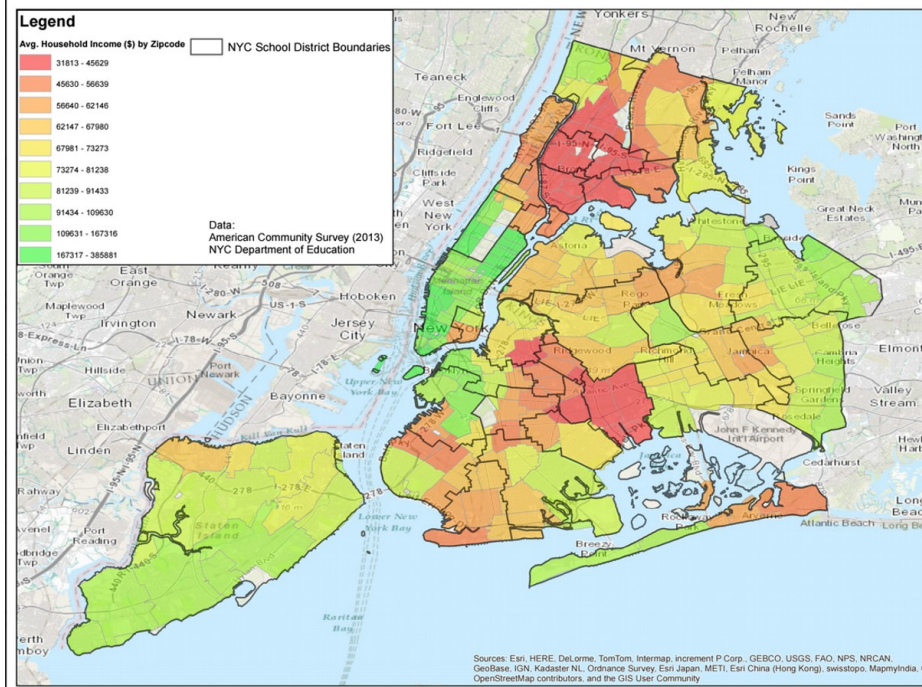
(a) Total variation in exam-time temperature



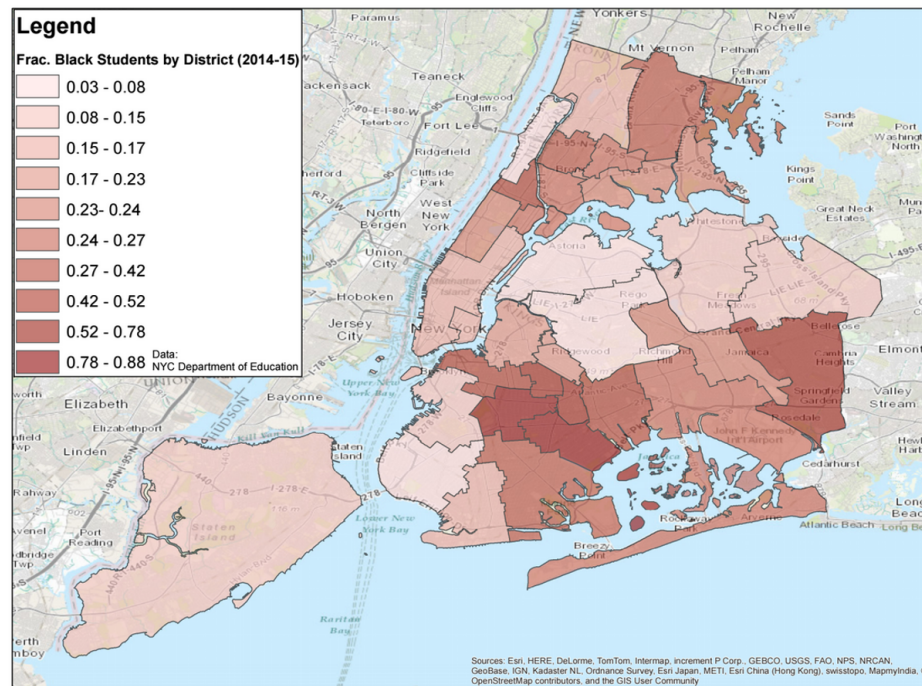
(b) Temperature on two consecutive test days

Figure 2: Short-Run Identifying Variation in Temperature

Notes: This figure illustrates the source of identifying variation for short-run performance impacts of heat exposure. It presents realized exam-time temperatures for (a) all June Regents exams (1999-2011) and (b) for two subsequent days within a Regents exam period – Thursday, June 24th, 2010, and Friday, June 25th, 2010 – inclusive of spatial and temporal temperature corrections. Temperatures are measured at the school level, weighted by number of exam observations by date and time.



(a) Average household income by zip code



(b) Black students by district

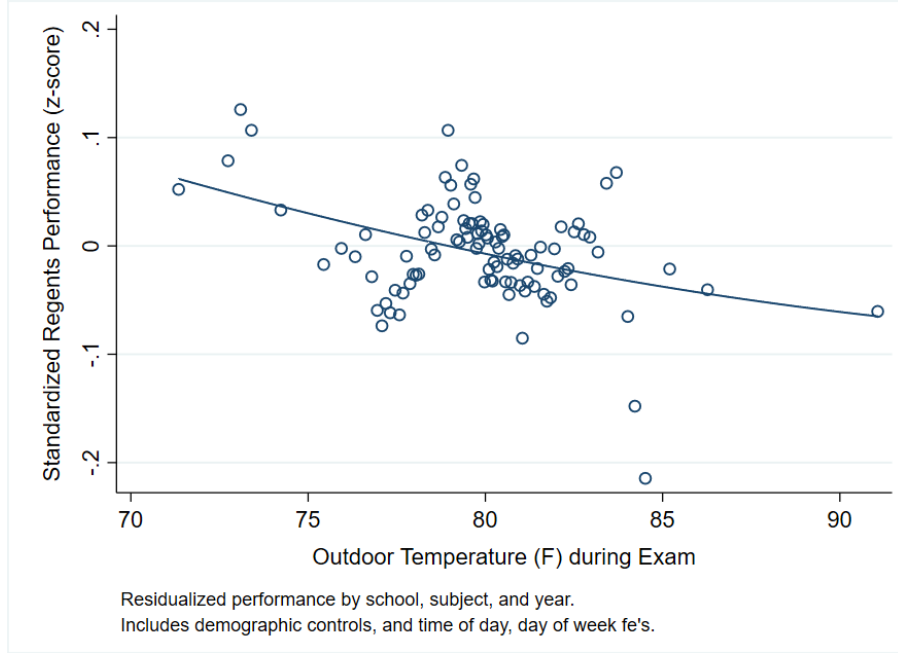
Figure 3: NYC public school districts by income and race

Notes: Panel (a) presents average household income in 2010 by zip code, with New York City Public School districts super-imposed. Panel (b) presents the average percentage of black students in 2014-2015 by sub-district within the New York City Public Schools system.

Ethnicity	Score	Pass	Proficiency	Previous Ability
Asian	74.73 (16.80)	0.78 (0.41)	0.57 (0.49)	0.98 (1.54)
Black	61.21 (17.05)	0.50 (0.50)	0.23 (0.42)	-0.18 (1.34)
Hispanic	61.49 (17.23)	0.51 (0.50)	0.24 (0.42)	-0.16 (1.32)
Multiracial	69.65 (17.44)	0.69 (0.46)	0.44 (0.50)	0.34 (1.26)
Native American	61.96 (18.08)	0.51 (0.50)	0.26 (0.44)	-0.22 (1.45)
White	72.92 (16.78)	0.75 (0.43)	0.52 (0.50)	1.02 (1.56)
Total	64.86 (17.92)	0.57 (0.49)	0.32 (0.47)	0.16 (1.42)

Figure 4: Summary Statistics by Ethnicity

Notes: Table (4) presents summary statistics for student performance variables. Standard deviations are in parentheses. ‘Pass’ and ‘Proficiency’ denote the fraction of scores above passing and college proficiency thresholds. Previous ability is measured as average z-scores from standardized math and verbal assessments in grades 3 through 8.



(a) Residualized variation in test performance

	(1)	(2)	(3)	(4)
	Z-score	Z-score	Z-score	Z-score
Temperature (F)	-0.00850*** (0.00231)	-0.00736*** (0.00207)	-0.0102*** (0.00233)	-0.0108*** (0.00226)
Afternoon	-0.0297* (0.0130)	-0.0334** (0.0119)	-0.0180 (0.0142)	-0.0156 (0.0127)
Fixed Effects				
Student by Year	X			
Subject	X	X	X	X
Time of Day, Day of week	X	X	X	X
Student		X		
Year		X	X	
School			X	
School by Year				X
N	3581933	3581933	3581933	3581933
r2	0.774	0.717	0.252	0.271

Robust standard errors in parentheses, clustered at the borough (station) by date-time level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Dependent variable is standardized performance by subject

Figure 5: Short-Run Impacts of Heat Exposure on Exam Performance

Notes: Panel (a) presents a binned scatterplot of residualized exam performance by percentile of the temperature distribution controlling for school, subject, and year fixed effects. Each dot represents approximately 220,000 exam observations. Panel (b) presents the main regression results. Fixed effects are suppressed in output, and 919,067 singleton observations are dropped. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.

	(1) Pass	(2) Pass	(3) Pass	(4) Pass
Temperature (F)	-0.00371*** (0.00107)	-0.00335*** (0.000932)	-0.00496*** (0.00102)	-0.00522*** (0.000986)
Afternoon	-0.0133* (0.00576)	-0.0140** (0.00531)	-0.00680 (0.00575)	-0.00633 (0.00525)
Fixed Effects				
Student by Year	X			
Subject	X	X	X	X
Time of Day, Day of week	X	X	X	X
Student		X		
Year		X	X	
School			X	
School by Year				X
N	3581933	3581933	3581933	3581933
r2	0.647	0.557	0.151	0.168

Robust standard errors in parentheses, clustered at the borough (station) by date-time level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(a) Pass rates

	(1) Proficient	(2) Proficient	(3) Proficient	(4) Proficient
Temperature (F)	-0.00372*** (0.00101)	-0.00322*** (0.000874)	-0.00567*** (0.000977)	-0.00581*** (0.000987)
Afternoon	-0.0108* (0.00505)	-0.00997* (0.00471)	-0.00170 (0.00560)	-0.00145 (0.00524)
Fixed Effects				
Student by Year	X			
Subject	X	X	X	X
Time of Day, Day of week	X	X	X	X
Student		X		
Year		X	X	
School			X	
School by Year				X
N	3581933	3581933	3581933	3581933
r2	0.701	0.624	0.221	0.235

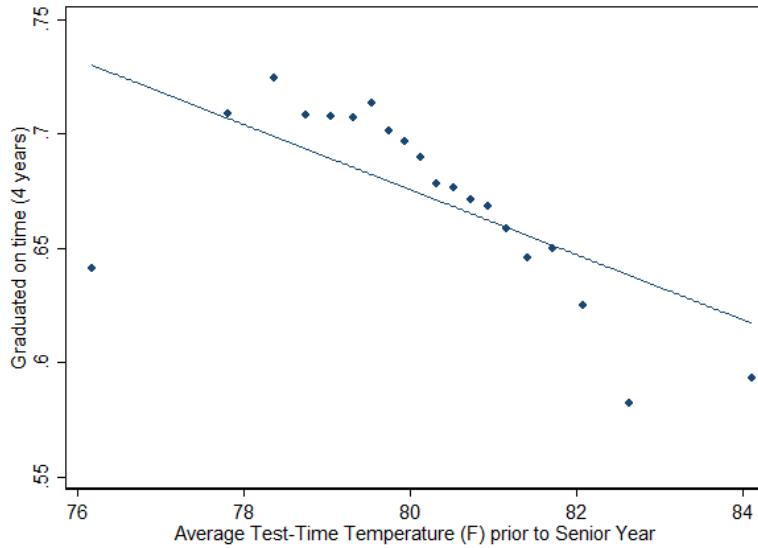
Robust standard errors in parentheses, clustered at the borough (station) by-date-time level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Proficiency status

Figure 6: Short-Run Impacts on Pass and Proficiency rates

Notes: Fixed effects are suppressed in output, and 919,067 singleton observations are dropped. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.



(a) 4-year graduation status and exam-time temperature

	(1)	(2)	(3)
	Graduated	Graduated	Graduated
Avg Exam-Time Temp (°F)	-0.00712*** (0.00173)	-0.00758*** (0.00223)	-0.00733** (0.00231)
Number of June exams			0.193*** (0.00688)
Number of June exams ²			-0.0151*** (0.000809)
Number of June exams ³			0.000312*** (0.0000225)
Fixed Effects			
School	X	X	X
Number of June exams	X	X	
Cohort		X	X
N	515192	515192	515192
r ²	0.232	0.238	0.236

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Impacts on graduation status by regression specification

Figure 7: Persistent Impacts of Short-Run Heat Exposure: Graduation Status

Notes: Panel (a) presents a binned scatterplot of 4-year graduation status by quantile of exam-time temperature distribution. Temperatures are averaged by student for June exam sessions up through senior year. Residual variation after controlling for school and number of exam fixed effects, student-level observables, and weather/air quality controls. In panel (b), the dependent variable is a dummy for whether or not student graduated in four years. All regressions include controls for daily precipitation, ozone, and dewpoint. Fixed effects are suppressed in output.

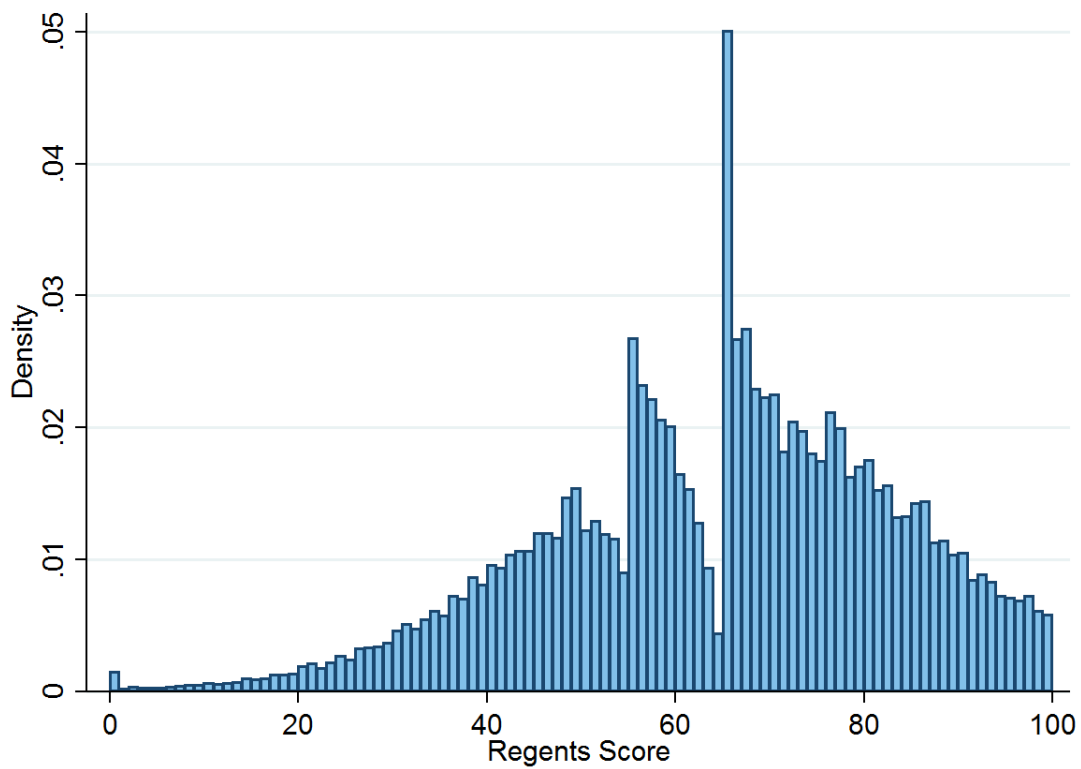
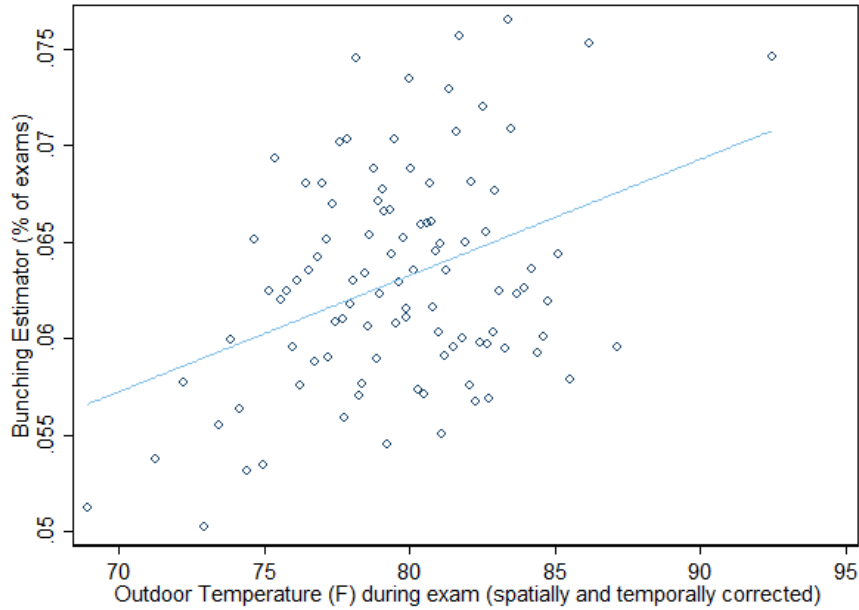
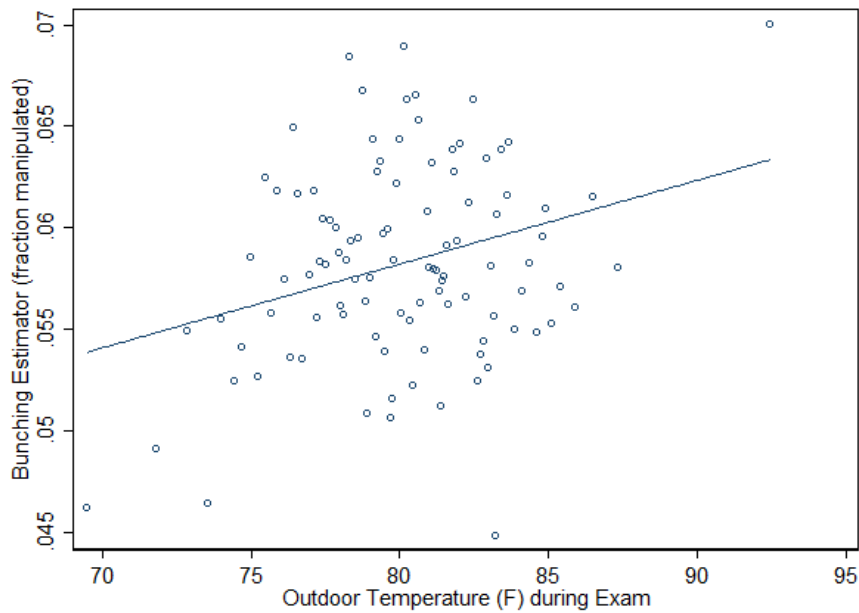


Figure 8: Exam scores exhibit bunching at pass/fail cutoffs, suggesting upward grade manipulation

Notes: This figure presents a histogram of Regents exam scores from June 1999 to June 2011. A large number of observations bunch at the pass/fail cutoffs, scores of 55 and 65 for local and Regents diploma requirements respectively.



(a) Grade Manipulation varies with exam-time temperature by subject, school, and take.



(b) Grade Manipulation expressed as a fraction of scores in manipulable range.

Figure 9: Ex Post Compensation: Grade Manipulation by Teachers

Notes: Panel (a) presents a binned scatterplot of bunching at the school-subject-date level by quantile of the exam-time temperature distribution, controlling for subject and year fixed effects and daily weather and air quality controls. Panel (b) expresses bunching as a fraction of manipulable scores, to account for potential mechanical correlation between temperature and the number of scores falling in the manipulable zone. Included in the analysis are all June Regents exams in core subjects between 1999 and 2011.

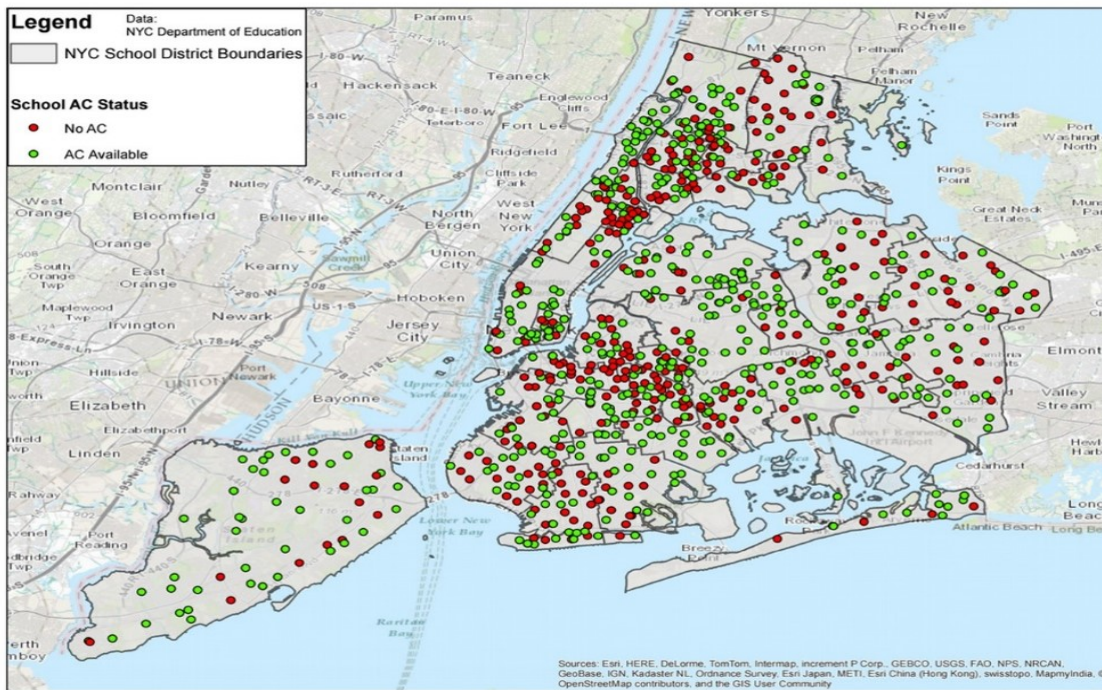
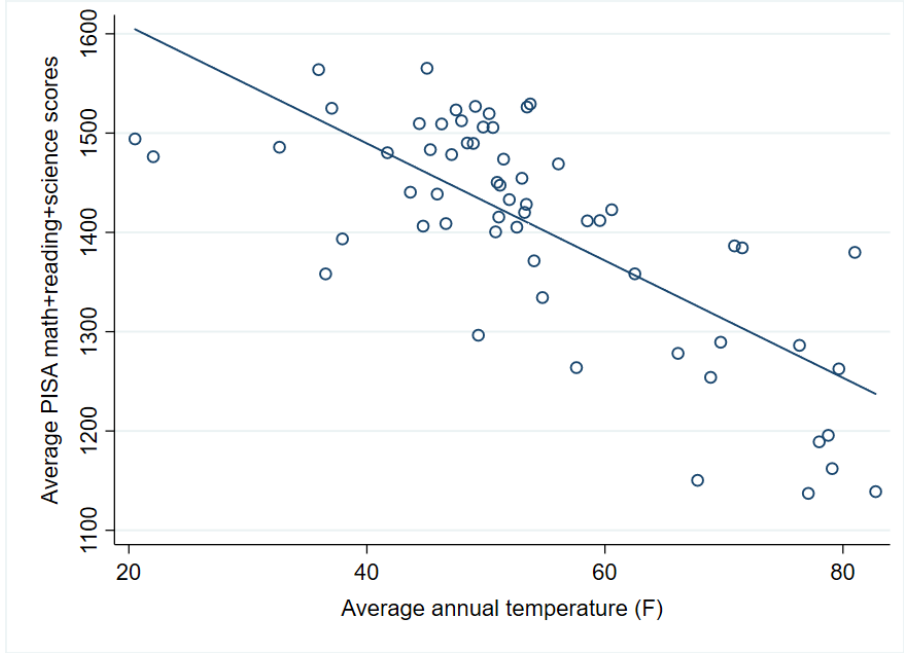
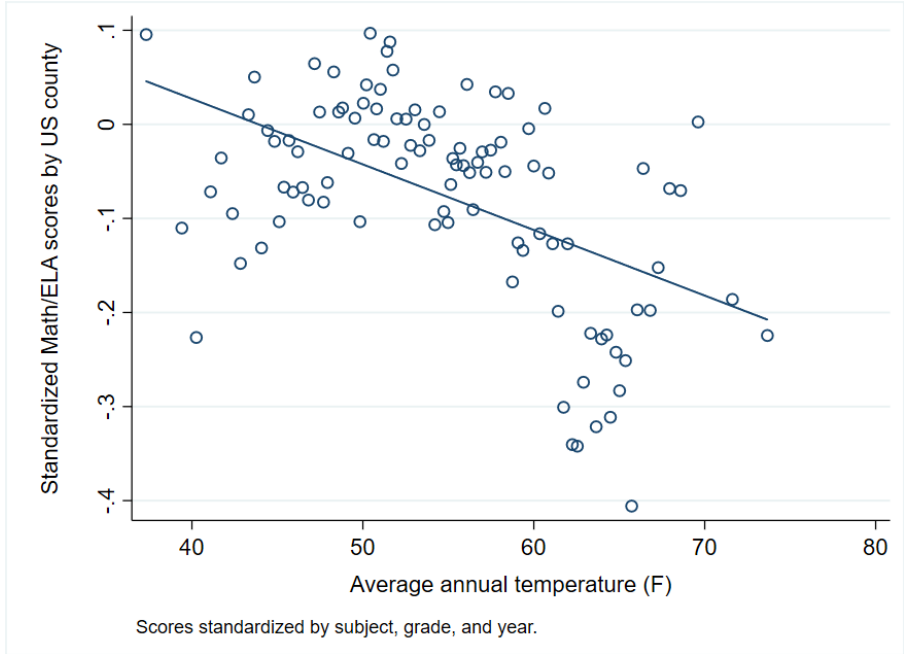


Figure 10: School Air conditioning status as of 2012

Figure 10) provides a map of New York City public schools, with green dots representing schools that had any air conditioning equipment as of 2012, and red dots representing schools that did not. Data comes from a scrape of publicly available Building Condition and Assessment Survey (BCAS) reports from each school website.



(a) Climate and PISA Math, Reading, and Science Achievement Across Countries



(b) Climate and Standardized Math and Reading scores within United States

Figure 11: Climate and Student Performance in Cross-Section

Notes: Panel (a) presents a scatterplot of mean PISA scores (math+reading+science) and average annual temperature by country. A standard deviation in PISA scores corresponds to approximately 300 points (100 per subject). Panel (b) presents a binned scatterplot of standardized 3-8th grade math and reading scores (2009-2013) by percentile of the county-level temperature distribution (scores standardized by subject-grade-year as in Fahle et al. (2017)). Average annual temperatures in both cases are measured over the period 1980-2011.