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

Jisung Park*

November 14, 2017

Abstract

Understanding the link between temperature and educational outcomes is important in assessing the returns to 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 exams and subsequent educational attainment. Hot days reduce performance by up to 14% and lead to lasting impacts on high school graduation status. An analysis of teacher grade manipulation provides the first available evidence for ex post avoidance behavior to hot temperature.

Keywords: temperature, human capital, climate change, adaptation

JEL Codes: I21, O18, Q54, Q56

The realized welfare impacts of a hotter world will depend in large part on the persistence of the direct effects (e.g. health) and the effectiveness of avoidance behaviors aimed at mitigating them. Despite a growing empirical literature which estimates the impacts of temperature stress on human economic activity, it remains unclear whether human capital should be included as part of the climate damage function.[1] Moreover, adaptive responses – also referred to as avoidance behaviors or defensive investments – are an important yet understudied component of the welfare equation (Graff Zivin and Neidell, 2013, 2009; Graff-Zivin et al., 2011; Deschenes

*Park: UCLA Luskin School of Public Policy, jisungpark@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, and the Harvard Climate Change Solutions Fund is gratefully acknowledged.

[1]Current social cost of carbon estimates are based predominantly on integrated assessment models that do not include direct human capital impacts, and often assume contemporaneous levels effects on economic activity as opposed to persistent growth rate effects (Pindyck, 2013; Heal and Park, 2016).

et al., 2017). This is particularly true in the context of schooling, where data on defensive investments has historically been lacking, and where market imperfections are more likely to drive a wedge between socially optimal and privately realized investments.

This paper uses administrative data from the largest public school district in the United States to assess the relationship between hot temperature and educational attainment. Using detailed student-level data from 4.5 million high school exit exams, I investigate the relationship between hot temperature, high stakes exam performance, and subsequent high school graduation rates, as well as defensive investments by teachers and school administrators. High school graduation is likely to be an important determinant of lifetime earnings and individual welfare due to the persistent impact of formal schooling on employment, wages, and further educational investment (Angrist and Lavy, 1999; Cunha and Heckman, 2007; Chetty et al., 2011). This paper extends the existing laboratory and survey-based literature by assessing responses in high stakes environments, and is the first to document persistent impacts of temperature on educational attainment. It is also the first to document ex post compensatory investments in response to heat exposure: in this case, selective upward grade manipulation by teachers in response to hot exam sittings and incomplete air conditioning.

The empirical approach exploits quasi-experimental variation in outdoor temperature which, in conjunction with institutional features, allows for the identification of causal effects net of selection issues present in much of the previous literature. 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. 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. This allows for within-student estimation, something that has not been possible in the one existing study of temperature on high stakes exam performance (Cho, 2017), and minimizes bias from student selection and selective reporting – which may be a concern in the case of studies based on voluntary assessments such as the NLSY PIAT (Graff Zivin et al., 2017).[2] The richness of the data allows for an analysis of both ex ante and ex post compensatory behaviors, including school air conditioning and teacher grade adjustments, and sheds light on the potential interaction between institutional quality and adaptation to environmental change.[3]

The analysis is organized around three empirical research questions. First, does acute heat exposure meaningfully affect performance on high-stakes exams? That is, do the findings from survey and experimental contexts – wherein performance

[2]In addition to the potential for selective sorting based on unobservable student characteristics, survey-based studies face an additional challenge due to the fact that hot temperature may lead to systematic under-reporting of data by administrators. For instance, a substantial proportion of NLSY surveys are missing 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).

[3]This builds on recent work suggesting that the realized impact of natural disasters and other environmental shocks may depend on the quality of institutions (Kahn, 2005; Dell et al., 2012), the adoption of air conditioning (Barreca et al., 2016; Behrer and Park, 2017), and the possibility of moral hazard arising from government insurance schemes (Annan and Schlenker, 2015).

declines with elevated temperatures – extend to settings where there are real economic consequences? Second, can short-run heat exposures, which presumably do not reduce the stock of human capital *per se*, nevertheless affect longer-run outcomes? 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 have lasting economic consequences.[4] Third, how have relevant economic agents responded? Emerging evidence suggests that avoidance behaviors may comprise a large share of the total welfare impacts of environmental stress – e.g. Deschenes et al. (2017) – but also that existing policies or market imperfections may drive a wedge between ex ante optimal and ex post realized avoidance behaviors (Kahn, 2005; Annan and Schlenker, 2015; Behrer and Park, 2017).

The first main result is that heat exposure during high-stakes exams exerts a causal and economically meaningful impact on student performance, even when controlling for individual ability. 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 13 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, and at least 18% of the students in the study sample experience an exam with outdoor temperatures exceeding 90°F. The fact that adverse impacts of heat exposure on student performance persist in such high stakes settings suggests that reduced effort – due to increased disutility of mental exertion under heat stress – is not the primary channel driving earlier results on low-stakes, voluntary assessments (Seppanen et al., 2006; Graff Zivin et al., 2017).

Second, I find that heat exposure during exams subsequently affects a student's chances of graduating from high school. 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). This means that a one standard deviation increase in average exam-time temperature reduces a student's likelihood of graduating on time by roughly 2.5 percentage points.[5] This is despite the fact that students are able and often encouraged to retake failed exams in the following school year. These results 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 investment process whereby idiosyncratic shocks discourage students or lead to subsequent tracking by teachers and peers (Cunha and Heckman, 2007; Diamond and Persson, 2016). While, in the context of climate policy, understanding the potential cumulative effect of long-run heat exposure on human capital attainment may potentially be of first-order welfare relevance, these

[4]Persistent impacts have been found in the context of air pollution (Ebenstein et al., 2016) and 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.

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

results suggest nevertheless that hotter temperatures can have persistent impacts, especially given well-documented links between high school graduation and lifetime earnings.[6]

Third, I find evidence of ex post compensatory behavior by teachers which may indicate constraints to ex ante defensive investments such as school air conditioning. Using a bunching estimator that controls for the temperature-test score relationship noted above, I find that teachers engaged in selective upward grade manipulation in a way that compensated for perceived adverse testing conditions.[7] Specifically, I estimate the relationship between grade manipulation and exogenous variation in exam-time temperature using a school, subject, and date-specific bunching estimator at passing cutoffs. I find that teachers boosted grades more frequently for exams that were taken under hotter conditions.

While it is impossible to definitively attribute teachers' intentions, this behavior is consistent with a setting in which market imperfections such as informational asymmetries between teachers and school administrators result in sub-optimal ex ante defensive investments – which could have taken the form of more flexible testing arrangements or better school air conditioning. Compiling school AC data from publicly available engineering assessments, I find that fewer than 60% of NYC public schools had any form of AC equipment during this period, and that at least 40% of the schools that did had defective units in need of repair.[8]

This paper is broadly related to a growing literature exploring the causal impact of climate on economic outcomes, including impacts of temperature shocks on human health (Barreca et al., 2016), labor supply (Graff Zivin and Neidell, 2014), violent crime (Kenrick and MacFarlane, 1986; Hsiang et al., 2013) and local economic output (Dell et al., 2012; Heal and Park, 2013; Deryugina and Hsiang, 2014), as well as the literature on avoidance behaviors and climate adaptation (Graff Zivin and Neidell, 2009; Graff-Zivin et al., 2011; Burke et al., 2015; Barreca et al., 2016; Deschenes et al., 2017). In particular, this paper shares with a smaller set of recent papers – including Graff Zivin et al. (2017), Garg et al. (2017) and Cho (2017) – an emphasis on understanding the mechanisms through which hot temperature may affect student performance. In comparison to the existing literature, this paper establishes a causal connection between temperature and longer-run educational attainment (as opposed to cognition in the short run), and also provides an analysis of ex ante and ex post avoidance behaviors.

The findings suggest that classroom temperature may be an important factor for policymakers to consider when allocating public resources, especially in contexts where heat exposure is frequent, high-stakes exams pose hurdles to further schooling,

[6] In this sense, these findings add to a nascent literature which documents later-life labor market impacts of early environmental shocks, including air pollution and heat exposure (Currie, 2009; Peet, 2014; Isen et al., 2015).

[7] Previous studies have used bunching estimators to document strategic behavior, including grade manipulation by teachers (Chetty et al., 2011; Dee et al., 2016; Diamond and Persson, 2016). I adapt this approach to account for potential mechanical correlation between temperature and the test-score distribution, estimating the relationship between bunching and temperature as a fraction of scores in the manipulable zone.

[8] This is compared to residential and commercial AC penetration rates of 90% or higher for the city on average. Comparing schools that do and do not have some form of AC, I find limited evidence for protective effects of having AC equipment present, consistent with inadequate investment.

and where market failures in infrastructure investment are possible. This paper also provides empirical support for the view that climate and human capital may interact in a way that contributes to the long-debated relationship between hotter climates and slower growth, though more careful research is needed to verify whether repeated heat exposure may reduce the rate of human capital accumulation. This is especially true given competing mechanisms in the existing literature, and the potential for endogenous innovation in adaptation technologies.

The rest of this paper is organized as follows. Section 1. provides a brief overview of the relevant literature on heat and human welfare and describes the underlying conceptual framework. 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.

1 The Welfare Economics of Hot Temperature

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, or 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.[9]

1.1 Heat Exposure and Human Capital

In seminal work, Graff Zivin et al. (2017) provide evidence of the adverse impacts of temperature stress on cognitive performance in US households. They use NLSY survey data which includes short (10 minute), voluntary assessments that were administered to a sample students at home, and find evidence of contemporaneous impacts of hot days on math performance but not verbal performance, and little evidence of cumulative or persistent impacts over time.

Empirical work in school settings is limited mostly to similary short cognitive assessments, where it is unclear that the incentives faced by students and teachers are strong enough to induce adaptive responses relevant in more high stakes settings or in the presence of persistently hotter temperatures.[10] Schoer and Shaffran (1973)

[9] See Heal and Park (2016) for a review 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 (2014) for impacts of temperature on labor and task productivity.

[10] In 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. In the case of pollution during high-stakes exams in Israel, Ebenstein et al. (2016) find evidence for persistent and economically meaningful impacts that extend well beyond

compare the performance of students in a pair of classrooms set up as a temporary laboratory, with one classroom cooled and one not. They find higher performance in cooled environments relative to hot ones, but the assessments have no bearing on actual schooling outcomes.

Garg et al. (2017) and Cho (2017) study the influence of hot temperature in Indian and Korean schools respectively, but study low-stakes exams and are unable to provide direct evidence for avoidance behaviors. Garg et al. (2017) study the impact of temperature on Indian students and find that heat exposure reduces cognitive performance in the short run, and that years with more days above 29°C during the growing season reduce subsequent performance on cognitive assessments, which they suggest is driven primarily by reduced crop yield and subsequent indirect impacts of heat on nutrition and health. Cho (2017) uses variation in summer-time temperature to assess the relationship between summer heat and subsequent exam performance of Korean high school students, and finds evidence that cumulative heat exposure reduces subsequent performance. The primary identification in Cho (2017) is driven by school fixed effects (student-level identifiers are not present in the data), making the results vulnerable to selection bias arising from changes in student composition. Due to data limitations, Cho (2017) is unable to provide evidence of avoidance behaviors such as air conditioning or health investments to prevent food poisoning or other heat-related illnesses. While these exams are high stakes, they are taken in November, which rules out the possibility of exploring the impact of hot temperature on contemporaneous exam performance.

This study assesses the impact of temperature on human capital production by using administrative data from New York City public schools. In contrast to the existing literature, it assesses the impact of hot temperature on high stakes exams and subsequent educational attainment, and explores the role of avoidance behaviors in greater detail. The richness of the dataset permits causal identification using within-student variation in temperature, and allows linking of heat exposure to longer-run outcomes such as high school graduation rates and college eligibility. School-level air conditioning data and estimates of strategic teacher behavior permits an analysis of the potential extent of – and possible constraints to – efficient adaptation responses.

1.2 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 $h_T(\Sigma_t^T T_t)$.

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

formal schooling.

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 $\frac{\Delta h_{it+\tau}}{\Sigma_t^{t+\tau} \Delta T_{it}} \leq 0$, if the schooling environment features high costs of retaking exams, or in the presence of dynamic complementarities due to policies such as tracking.

3. Repeated heat exposure may reduce human capital accumulation and educational attainment over time: $\frac{\Delta h_{it}}{\Sigma_0^{t-1} \Delta T_{it}} \leq 0$. This may be due to reduced student cognition or teacher performance during study or class time. They may also arise from reduced attendance, heat-related illnesses (e.g. heat stroke, food poisoning), or increased violent behavior at school or at home.

4. Any combination of (a) information asymmetries between students and school administrators or teachers and school administrators, (b) market failures in the provision of local public goods (e.g. collective action problems in providing adequate electrification), or (c) liquidity constraints in the context of air conditioning or other infrastructure investment, can lead to socially sub-optimal adaptive responses to repeated heat exposure in educational settings.

2 Institutional Context, Data, and Summary Statistics

2.1 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. The median student is relatively low-performing and low-income, though a substantial minority attend high-achieving magnet schools including Stuyvesant Academy and Bronx Science, which consistently rank among the nation's best. Approximately one fifth of NYC students attend private schools which are not included in our sample.

The average 4-year graduation rate, at 68%, is below the national average but comparable to other large urban public school districts (e.g. Chicago). 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).

2.2 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 used to determine high school diploma eligibility as well as college admissions.

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.[11] 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.

The vast majority of students take their Regents exams during a pre-specified two-week window in mid-to-late June each year.[12] The test dates, times, and locations for each of these Regents exams are determined over a year in advance by the NY State education authority (NYSED), and synchronized across schools in the NYCPS 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.[13] 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 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.[14] 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 until grading reforms were implemented in 2011 in response to a series of reports suggesting grade manipulation (Dee et al., 2016).

In summary, using scores from Regents exams offers several distinct advantages. 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, allowing for a wide range of quasi-experimental variation within a relatively small geographic locale. Because they occur at the end of the school year, they are also

[11] The 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.

[12] For 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.

[13] Throughout 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.

[14] In 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.

more likely than periodic assessments or college entrance exams to reflect cumulative impacts of hot temperature that may have accrued over the course of the school year. 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.

2.3 Student Outcome Data

I obtain student-level information from the New York City Department of Education (NYC DOE). This includes the universe of all public school students who took one or more Regents exams over the period 1999 to 2011.[15] Information on exam dates comes from archived Regents exam schedules from the NYC DOE database. A detailed description of the matching procedure and subsequent sample restrictions are provided in the online appendix.

2.4 Weather Data

Weather data comes from NOAA, 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 were available 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, Manhatten, 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.[16] 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.[17]

[15] I 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. For students who are missing these records, I assign imputed average z-scores by decile of the realized Regents score distribution. These exams are generally taken in January and March.

[16] Further 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.

[17] The air quality controls used here are admittedly 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 cannot separately identify the precise effects of pm, ozone, and temperature simultaneously.

2.5 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 or usage status by year, nor does it provide information regarding where *within* a school AC was available. As such, I take AC installation status as a coarse proxy for the true variable of interest, which is effective AC utilization at the classroom level.

2.6 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 4a 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.

Cumulative heat exposure during the school year can vary considerably. On average, NYC students experience between 19 and 39 days above 80°F per school year, with a mean value of 26.7 and a standard deviation of 5.6. In addition, there are on average 2.5 days during the school year above 90°F. Most of these days occur during the months of September, October, and June. Summer school students, which comprise roughly the bottom third of the student body but are not included in this analysis due to likely selection bias, are subject to an additional 9 days above 90°F on average.

In 2012, 62% of the NYC public school buildings for which I have building assessment data were reported as having some kind of air conditioning equipment

on its premises, including window units, which means that 38% of these buildings did not have any form of air conditioning equipment available (Figure 4b). Of the 62% that were reported as having air conditioning, 42% (259 out of 644) were cited as having defective components, according to the third-party auditors conducting the BCAS assessments.

3 Empirical Strategy and Primary Results

To organize the empirical findings, I report the results by research question.

1. Does Exam-Time Heat Exposure Affect 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. It suggests that exams taken on hot days 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 (on average 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.[18]

Table 5b presents the results from running variations of equation (4) for the subset of students who take at least 2 exams in any given year.[19] 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, as well as to using the full sample of scores: i.e. retaining all students with 1 or more exam records and replacing student fixed effects with school fixed effects and a vector of demographic controls.

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.[20]

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 comparable in raw magnitude to the impacts on mathematical reasoning found by Graff Zivin et al. (2017), who find a 90°F day to reduces 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 in short-run impacts, though it is possible that adaptive responses by teachers are offsetting impacts disproportionately for

[18] To 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 it is possible for measurement error to 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.

[19] Results 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.

[20] Precipitation 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. The maximum recorded value of pm2.5 in my data is 38.8 micrograms per cubic meter.

certain subgroups (as discussed in greater detail below).

These results provide strong evidence that heat stress affects student performance, either by reducing raw cognitive ability or by increasing the disutility of effort which in turn affects students' willingness or ability 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. However, these results are not in themselves directly informative of welfare impacts or effects on the stock of human capital or longer-run educational attainment, which is explored in the next section.

2. Does Short-run Heat Exposure Affect Longer-run Educational Outcomes?

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

In a stylized, friction-less world with fully flexible educational institutions, unlucky students who fail a subject due to a hot exam sitting would immediately retake the exam until she believes her "true ability" has been reflected in the exam score: $s_{it} = h_{it}$. 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 (by, for instance, requiring students to attend remedial courses), it is possible that even short-run heat exposure can lead to ripple effects on long-run educational attainment. In addition, 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. Similarly, 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; Diamond and Persson, 2016).

Recent 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.[21]

[21] 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.

Pass Rates and College Proficiency

If heat exposure during an exam pushes some students below important (cardinal) score thresholds that affect access to further educational opportunities, one might expect even short "doses" of heat exposure to give rise to lasting impacts on educational attainment.

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. This cutoff does not change based on the realized distribution of performance in any given year.[22] 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 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.

[22] 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, as do Dee et al. (2016). Results of running the regression analyses below using the "Local Diploma" cutoff feature similar (slightly more negative) point estimates.

Graduation Status

I find evidence that these short-run heat exposures reduce final educational attainment by affecting the likelihood that students graduate from high school. Figure 7a presents a binned scatterplot of 4-year graduation status on average exam-time temperature by student during June Regents exams up through the student's senior year. It plots residual variation in a dummy variable denoting graduation status, controlling for school-level averages, student-level demographic characteristics, and the number of June Regents exams taken by student. It suggests that students who experience greater exam-time heat stress are less likely to graduate on time.

Whereas short-run impacts of heat stress could be identified *within* student cells, long-run impacts on graduation status cannot because the outcome variable is no longer date-specific. This poses additional challenges to causal identification. 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.[23]

The comparison of interest is 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:

$$g_{ijcn} = \alpha_0 + \alpha_1 \overline{T_{ij}} + X_{ij}\alpha_2 + \chi_j + \theta_c + Z_i\alpha_3 + exams_n\alpha_4 + \epsilon_{ijc} \tag{4}$$

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 previous ability (combined ELA and math z-scores); and $exams_n$ 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.[24]

[23] That 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).

[24] The intuition is that variation in experienced temperature among students in the same school and cohort will be plausibly uncorrelated with residual variation in graduation status within school and cohort cells. Suppose there are two students, Jill and Karen, who entered high school in 2000. In 2001, because of differences in the sequence of subjects that Jill and Karen took, Jill takes Regents exams on Monday, Wednesday, and Thursday, and Karen takes Regents exams on Monday,

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 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 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.

Cumulative Heat Exposure and Learning

While lack of geographic variation and a relatively short panel do not allow for a definitive analysis of potential cumulative impacts, I find evidence suggestive of hot days during the school year having an additional adverse impact on learning, controlling for the short impacts documented above, consistent with emerging evidence (Cho, 2017; Garg et al., 2017). As described in the appendix, cumulative heat exposure over the course of the preceding school year, measured by the number of days where temperatures exceed 80°F, is associated with reductions in end of year exam performance, controlling for the exam-day effects of heat stress noted above. A year with five additional 80°F+ days is associated with 2.1% of a standard deviation reduction in learning on average, effects that are similar in magnitude with previous findings. However, the impact of hotter days (90°F and above) is positive, suggesting that the relevant measure of cumulative heat exposure is imprecisely measured.

Tuesday, and Friday. Suppose a similar phenomenon occurs during their sophomore, junior, and senior years, such that they take the same overall number of June exams. The variation in overall experienced temperature between Karen and Jill in 2001 will likely be exogenous to any unobserved differences in Jill and Karen's likelihood of graduating from high school.

3. How Do Agents Adaptively Respond?

A growing economic literature explores the role of optimizing responses in the face of environmental shocks. In the realm of pollution impacts, these responses are often referred to as "avoidance behaviors" or "defensive investments"; in the climate-economics literature, the term more often used is "adaptation". In the analysis that follows, I use these terms interchangeably.

Avoidance behaviors 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 (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. Such avoidance behaviors can drive a wedge between potential biological exposure and realized economic impact, and ignoring such behavior can lead to a mischaracterization of social welfare, since such defensive investments presumably entail some cost (Courant and Porter, 1981; Graff Zivin and Neidell, 2013).

If environmental stress affects health and productivity in economically meaningful ways, standard economic theory would predict defensive investments along the most cost-effective margins – at least in the absence of market failures. However, in many institutional settings, a variety of factors can constrain individual responses, driving an additional wedge between potential and realized impacts. In other words, the adaptation choice set faced by individual agents will depend crucially on the extent of market failures, pre-existing policies, or institutional quality.[25]

Studying such institutional factors is arguably more important in assessing the impact of temperature on human capital, since students are often minors and thus depend in large part on legal adults to make various decisions on their behalf.[26] 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 can afford it financially. Nor, in most cases, can her parents or teachers, even if they divine a clear preference or need on part of their students. Overlapping health and safety regulations as well as typical capital budgeting cycles in US public schools suggest that the set of options available to students and teachers

[25] For instance, outdoor laborers facing high levels of air pollution may or may not be able to flexibly adjust working hours, depending on the contractual setting (Graff-Zivin and Neidell, 2012). Even individuals who work primarily indoors may not be able to invest in air purifiers if doing so requires overcoming informational asymmetries and transaction costs (e.g. collective bargaining). While empirical work on these interactions and the potential for market failures in avoidance behavior is still nascent, the quality of existing institutions has been found to affect the severity of mortality responses to environmental disasters such as earthquakes (Kahn, 2005), and the adoption of heat-resistant crops by farmers in the US (Annan and Schlenker, 2015). In the context of energy infrastructure, market failures arising from imperfect information, liquidity constraints, or principal agent problems may lead to sub-optimal investment. Gertler et al. (2016) find evidence for liquidity constraints in the adoption of energy-intensive appliances, and a long literature has noted the potential for principal agent problems in the context of renter-occupied homes (Allcott and Greenstone, 2012).

[26] It is possible to represent the student's investment in schooling in terms of the canonical principal-agent problem, where the student is the principal and teachers/school administrators are the agents.

may be constrained relative to other market contexts.[27]

Ex Ante Defensive Investments: Air Conditioning

Air conditioning is a potential defensive investment which, at least in the context of health and labor productivity, has proven to be effective at mitigating the direct effects of hot temperature (Barreca et al., 2016; **?**). The fact that many US school districts and teachers unions have clashed over installation of air conditioning, however, suggests possible information problems associated with the efficacy of various adaptation options.[28] Parents and teachers in a number of major public school districts (e.g. New York, Los Angeles, Denver) have signed petitions asking school districts to upgrade air conditioning equipment. Press reports of inadequate AC in New York City schools abound.[29]

Average air conditioning penetration for commercial and residential buildings in New York City is around 90%, suggesting that historical rates of heat exposure may be frequent enough to merit investment in AC in the region. As noted above, however, fewer than 60% of NYC public schools had any form of air conditioning as of 2012, and only 36% are reported as being free from mechanical defects.[30] While it is possible that other unobserved factors (e.g. building age, occupancy rates) give rise to this discrepancy, accounts from teachers, parents, and principals suggest that a variety of institutional factors inhibit potentially welfare-enhancing investment.

Comparing schools that do and do not have air conditioning as of 2012, I find evidence consistent with inadequate air conditioning provision. The impact of hot temperature in schools with AC equipment is somewhat lower than those without AC, but still negative. Given the low quality of air conditioning data, and the lack of temporal variation, these analyses should not be taken as definitive statements regarding the potential efficacy of school air conditioning, but rather as descriptive evidence that air conditioning penetration and utilization are likely incomplete. These analyses are presented in greater detail in the Appendix.

[27] For 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 conditioning cannot clear unless electricity is obtained completely from renewable sources.

[28] 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).

[29] 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.

[30] Summer vacation means that occupancy may be lower for schools during hotter months, making this comparison somewhat less informative. However, a substantial fraction (often 25% or more) of public school students are required to attend summer school due to low achievement status.

Ex Post Compensatory Behavior: Teacher Responses

When environmental insults occur in a setting where sufficient ex ante investments have not been made, ex post compensatory behaviors may play an important role in mitigating the realized welfare impacts. For instance Deschenes et al. (2017) find asthma medication purchases, including ex post medication, to comprise a large part of the total WTP for air quality. In the context of heat and human capital, Graff Zivin et al. (2017) suggest the possibility of ex post parental investments leading to a discrepancy between short-term impacts on cognition and long-term effects on human capital attainment. So far, ex post responses to heat exposure have not been documented empirically. Given the potential constraints noted above, it seems possible that teachers or parents may engage in ex post compensatory behaviors.

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. One might call this selective response by teachers "adaptive grading", an ex post avoidance behavior undertaken in the presence of institutional barriers to ex ante investments. 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. The short-run performance impacts documented above will, in this sense, be net of adaptive grading [31]

[31] This 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

Estimating Adaptive 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 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 "adaptive grading", I estimate a version of this 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:

$$F_{ks} = \Sigma_{i=0}^{q} \psi_{ismyj} \dot{(Score)}^{i} + \Sigma_{i \in -M_{cs}, +M_{cs}} \lambda_{ismyj} \dot{\imath}[Score = i] + \epsilon_{ksmyj} \qquad (5)$$

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.[32]

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

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.

[32] As 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.

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:

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

where ζ_{smyj} 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_{smyj} 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_{smyj} as the dependent variable:

$$\zeta_{smyj} = \delta_0 + \delta_1 T_{smyj} + \delta_2 X_{smyj} + \chi_j + \eta_s + f(Year_y) + \epsilon_{smyj} \tag{7}$$

where T_{smyj} denotes temperature, X_{smyj} denotes precipitation and humidity, χ_j, η_s, and θ_y denote school, subject, and year fixed effects respectively, and $f(Year_y)$ 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.

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.

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.

While these results are highly suggestive of adaptive grading, it is not possible to infer teachers' intentions based on these results alone. It could be the case that teachers 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 that they 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.

4 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 (1) reducing performance on high-stakes exams and (2) reducing high school graduation rates and affecting college enrollment, despite (3) ex post compensatory behaviors by teachers who upward manipulate borderline scores for exams taken under hot conditions. The research design exploits quasi-random, within-student temperature variation to identify the impact of hot days on performance. The breadth and depth of the data set allows for credible causal estimation as well as an assessment of avoidance behaviors. Linking exam data to administrative records on graduation status allows for an analysis of the potential for persistent effects on educational attainment.

Taking an exam on a 90°F day results in 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 can lead to substantial reductions in longer-run educational attainment. A 90°F day 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, consistent with the presence of dynamic complementarities in human capital production Cunha and Heckman (2007). 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.

A series of additional analyses provide a window into the important interactions between institutional setting and avoidance behaviors. The evidence suggests that, as has been shown in the context of human health (Kahn, 2005), the quality of institutions may be an important factor that determines the realized welfare impacts of any given environmental shock. At less than 62% as of 2012, the AC penetration rate in New York City public schools seems to be far below residential and commercial sector averages for the region, which is over 90%. Of those schools that have AC equipment, 40% are reported as having defective components, suggesting possible institutional constraints to ex ante defensive investments.

Perhaps in response to insufficient ex ante defensive investments (or the perception of insufficient investment), 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 control-

ling 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, which presumably affect students' scores but do not reduce human capital *per se*.[33]

These results have several implications. First, they suggest that temperature should be included among the long list of relevant inputs to schooling and education policy. For instance, the timing of high-stakes exams – and the characteristics of the built environment in which they are administered – may affect social welfare, either from the standpoint of allocative efficiency or distributional equity. 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). They may also affect distributional equity if individuals taking nationally or internationally standardized exams such as the SAT or ACT in a hotter region (or in an older school building) are placed at a disadvantage relative to their peers in cooler regions (or climate-controlled buildings). The latter dimension may be of particular importance in developing countries considering the well documented relationship between income and air conditioning ownership at the household level (Gertler et al., 2016). Such factors may also be relevant in thinking about the persistence of racial achievement gaps in the United States, given cross-county correlations between race, income, and average climate.

Second, this study adds to a growing literature on avoidance behaviors and their relevance to economic welfare and environmental policy (Graff-Zivin et al., 2011; Graff-Zivin and Neidell, 2012), and is the first to document ex post defensive investments in response to hot temperature. The findings underscore a need for more careful research regarding potential market failures that may drive a wedge between realized adaptation strategies and the efficient adaptation frontier (Kahn, 2016; Annan and Schlenker, 2015).

Third, 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 operate through direct heat-stress of the human body (Tol, 2009; Heal and Park, 2016).[34] 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.

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 10a and 10b) can be explained by the cumulative influence of temperature stress on

[33] A 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.

[34] These findings also support the notion that climate change may affect not only the level of economic activity but overall growth rates (Pindyck, 2013), though more research on the impact of cumulative heat exposure on learning is needed.

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. 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.

Allcott, Hunt and Michael Greenstone (2012), "Is there an energy efficiency gap?" *The Journal of Economic Perspectives*, 26, 3–28.

Angrist, Joshua D and Victor Lavy (1999), "Using maimonides' rule to estimate the effect of class size on scholastic achievement." *The Quarterly Journal of Economics*, 114, 533–575.

Annan, Francis and Wolfram Schlenker (2015), "Federal crop insurance and the disincentive to adapt to extreme heat." *The American Economic Review*, 105, 262–266.

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." *Harvard Project on Climate Agreements Working Paper*, 16–81.

Burke, Marshall, Solomon M Hsiang, and Edward Miguel (2015), "Global non-linear effect of temperature on economic production." *Nature*, 527, 235–239.

Chetty, Raj, John N Friedman, Tore Olsen, and Luigi Pistaferri (2011), "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records." *The Quarterly Journal of Economics*, 126, 749–804.

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 (2009), "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development." *Journal of Economic Literature*, 47, 87–122.

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.

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.

Deryugina, Tatyana and Solomon M Hsiang (2014), "Does the environment still matter? daily temperature and income in the united states." *NBER Working Paper*.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro (2017), "Defensive investments and the demand for air quality: Evidence from the nox budget program." *American Economic Review*, 107, 2958–2989.

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.

Garg, Teevrat, Maulik Jagnani, and Vis Taraz (2017), "Human capital costs of climate change: Evidence from test scores in india."

Gertler, Paul J, Orie 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.

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 (2009), "Days of haze: Environmental information disclosure and intertemporal avoidance behavior." *Journal of Environmental Economics and Management*, 58, 119–128.

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.

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, Marshall Burke, and Edward Miguel (2013), "Quantifying the influence of climate on human conflict." *Science*, 341, 1235367.

Isen, Adam, Maya Rossin-Slater, and Reed Walker (2015), "Heat and long-run human capital formation." *Unpublished Manuscript*.

Kahn, Matthew E (2005), "The death toll from natural disasters: the role of income, geography, and institutions." *The Review of Economics and Statistics*, 87, 271–284.

Kahn, Matthew E (2016), "The climate change adaptation literature." *Review of Environmental Economics and Policy*, 10, 166–178.

Kenrick, Douglas T and Steven W MacFarlane (1986), "Ambient temperature and horn honking: A field study of the heat/aggression relationship." *Environment and behavior*, 18, 179–191.

Mackworth, Norman H (1946), "Effects of heat on wireless operators." *British journal of industrial medicine*, 3, 143.

Peet, Evan D (2014), "Environment and human capital: The effects of early-life exposure to pollutants in the philippines."

Pindyck, Robert S (2013), "Climate change policy: What do the models tell us?" *Journal of Economic Literature*, 51, 860–872.

Reardon, Sean F., Demetra Kalogrides, Andrew Ho, Ben Shear, Kenneth Shores, and Fahle Erin. (2016), "Stanford education data archive."

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*.

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."

Tol, Richard SJ (2009), "The economic effects of climate change." *The Journal of Economic Perspectives*, 23, 29–51.

Figures and Tables

The University of the State of New York
THE STATE EDUCATION DEPARTMENT
Office of State Assessment
Albany, New York 12234

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 ◆ 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.**	
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* ◆ Suggested date for administering locally developed tests aligned to the Checkpoint A and Checkpoint B learning standards for languages other than English (LOTE).	Physical Setting/ Chemistry RCT in Global Studies*	RCT in Writing	**RATING DAY**
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.

Use this space for computations.

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)

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

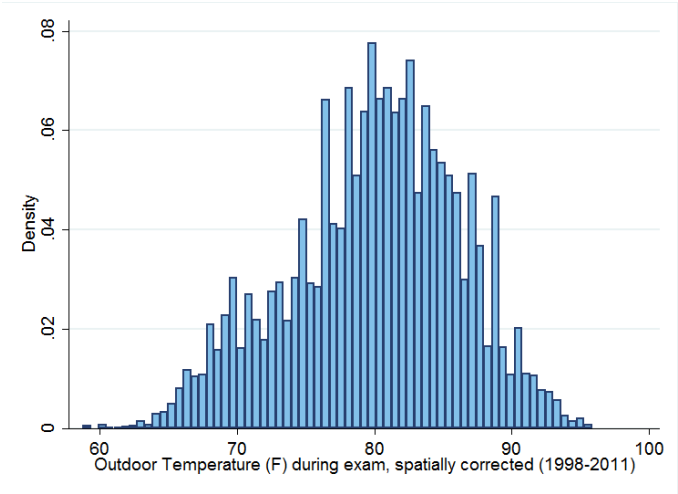
(1) I (3) III
(2) II (4) IV

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

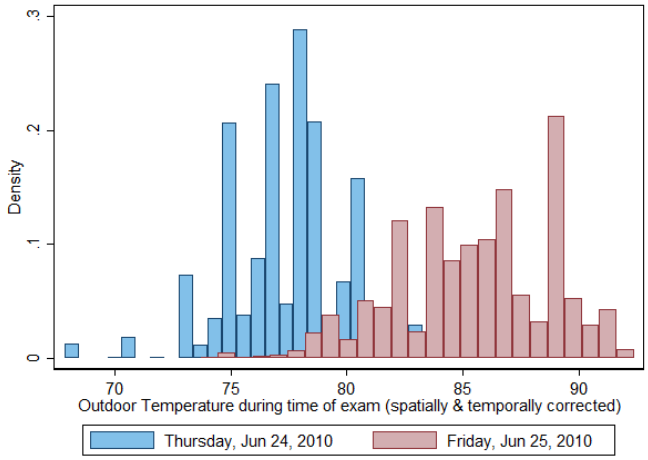
(1) −127 (3) 25
(2) −23 (4) 129

(b) Sample Regents subject exam cover sheet and questions

Figure 1: Sample Exam Schedule and Cover Page



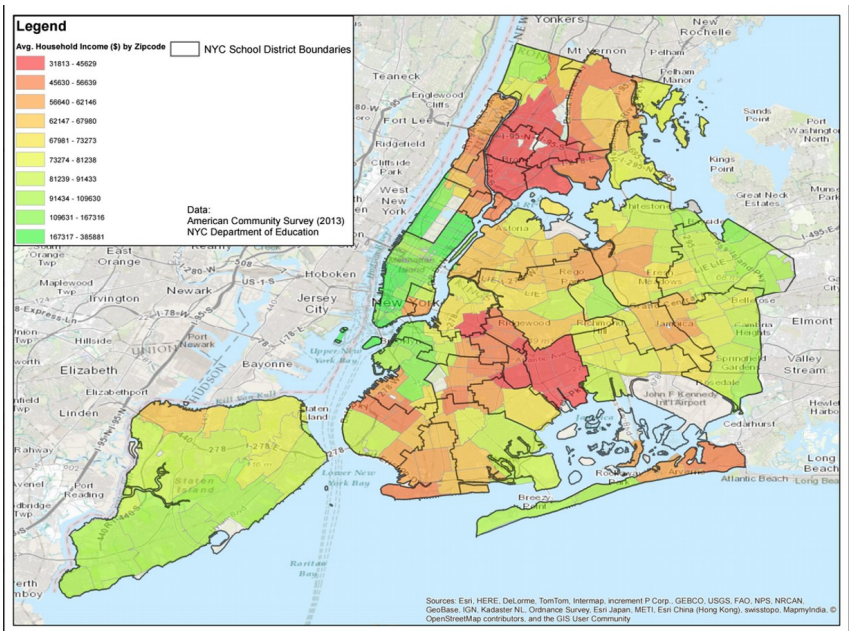
(a) Total variation in exam-time temperature



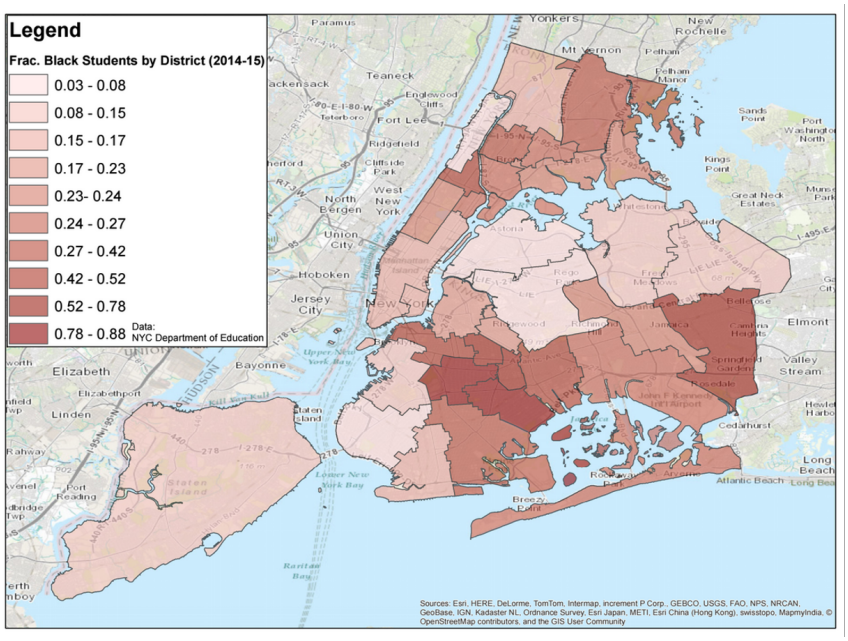
(b) Temperature on two consecutive test days

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



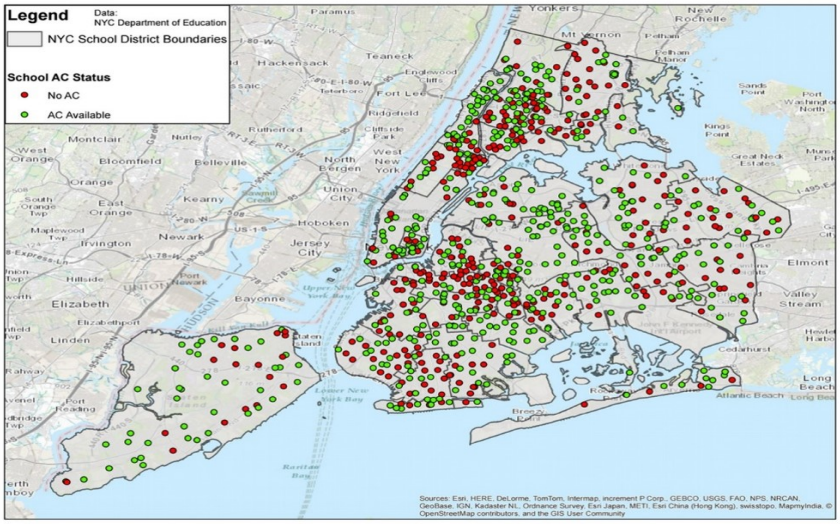
(b) Black students by district

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.

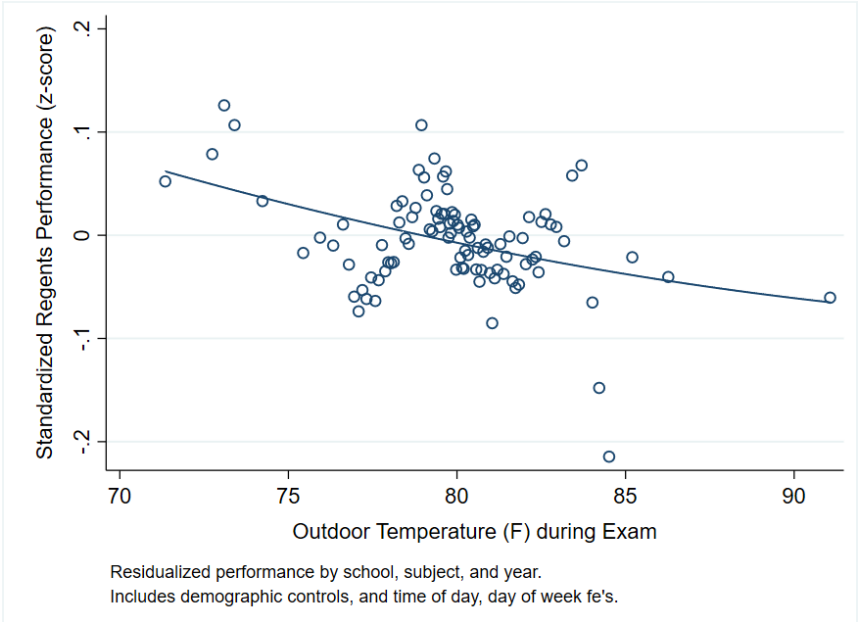
Ethnicity	Score	Pass	Proficiency	Previous Ability
Asian	74.73	0.78	0.57	0.98
	(16.80)	(0.41)	(0.49)	(1.54)
Black	61.21	0.50	0.23	-0.18
	(17.05)	(0.50)	(0.42)	(1.34)
Hispanic	61.49	0.51	0.24	-0.16
	(17.23)	(0.50)	(0.42)	(1.32)
Multiracial	69.65	0.69	0.44	0.34
	(17.44)	(0.46)	(0.50)	(1.26)
Native American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

(a) Summary Statistics by Ethnicity



(b) AC status by school

Notes: Panel (a) 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. Panel (b) 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.



(a) Residualized variation in test performance

	(1) Z-score	(2) Z-score	(3) Z-score	(4) Z-score
Temperature (F)	-0.00850***	-0.00736***	-0.0102***	-0.0108***
	(0.00231)	(0.00207)	(0.00233)	(0.00226)
Afternoon	-0.0297*	-0.0334**	-0.0180	-0.0156
	(0.0130)	(0.0119)	(0.0142)	(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***	-0.00335***	-0.00496***	-0.00522***
	(0.00107)	(0.000932)	(0.00102)	(0.000986)
Afternoon	-0.0133*	-0.0140**	-0.00680	-0.00633
	(0.00576)	(0.00531)	(0.00575)	(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

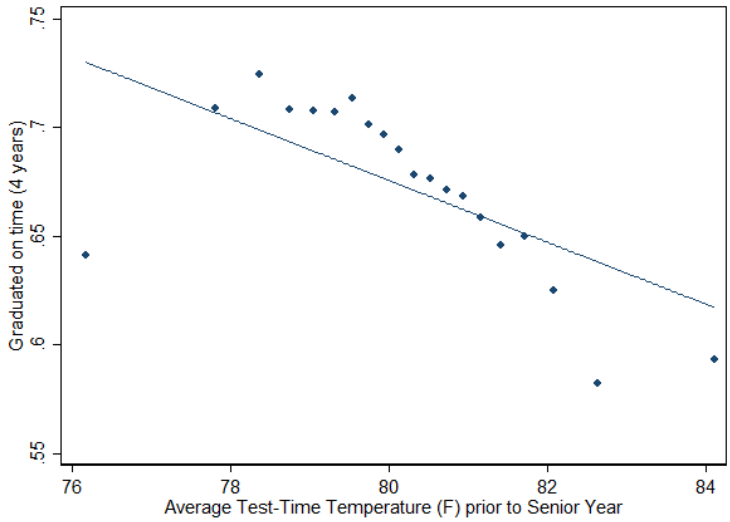
	(1) Proficient	(2) Proficient	(3) Proficient	(4) Proficient
Temperature (F)	-0.00372***	-0.00322***	-0.00567***	-0.00581***
	(0.00101)	(0.000874)	(0.000977)	(0.000987)
Afternoon	-0.0108*	-0.00997*	-0.00170	-0.00145
	(0.00505)	(0.00471)	(0.00560)	(0.00524)
Fixed Effects				
Student by Year	X			
Subject	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) Graduated	(2) Graduated	(3) Graduated
Avg Exam-Time Temp (°F)	-0.00712***	-0.00758***	-0.00733**
	(0.00173)	(0.00223)	(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
r2	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: Long-Run Consequences 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.

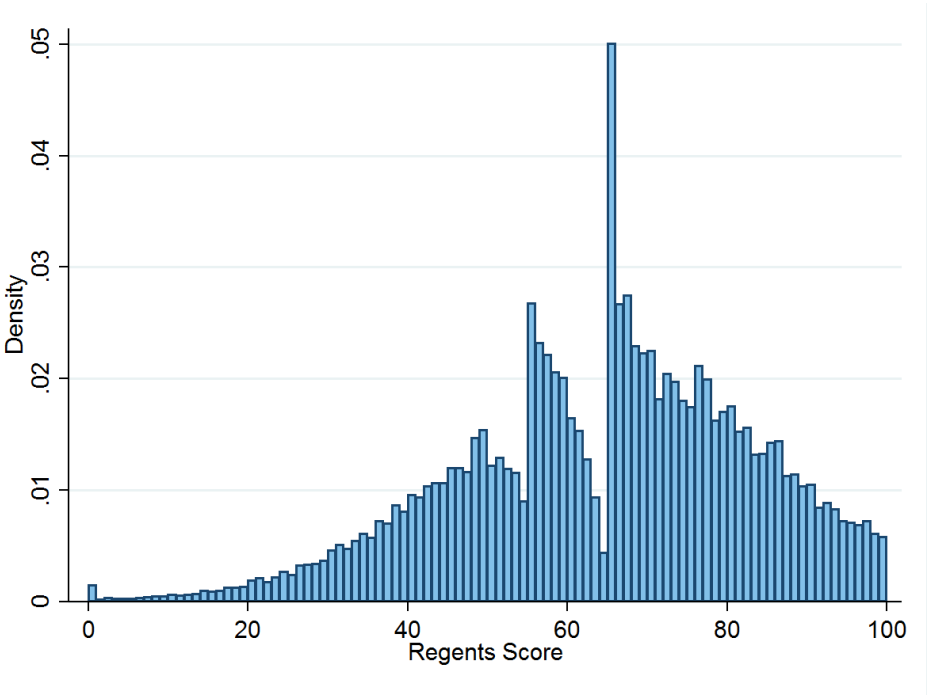
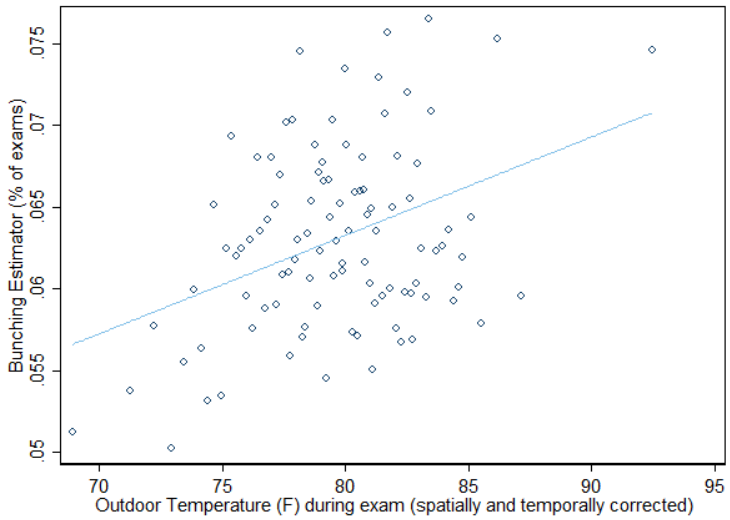
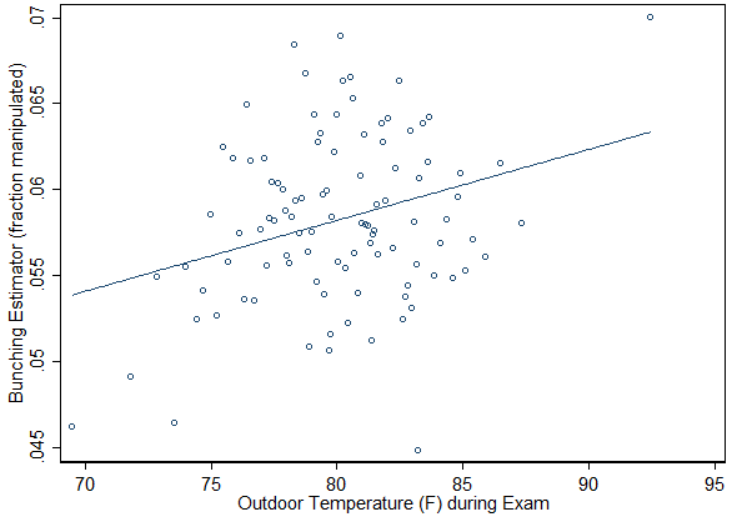


Figure 8: Exam scores exhibit bunching at pass/fail cutoffs, suggesting upward grade manipulation

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.



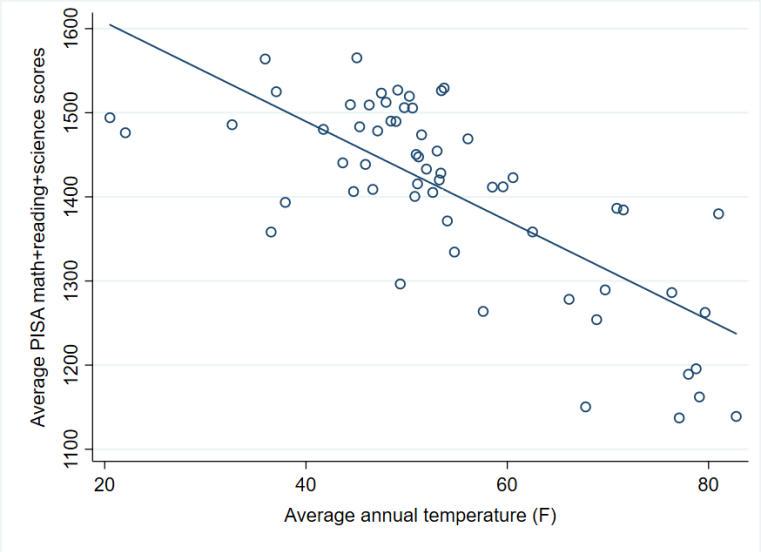
(a) Grade Manipulation varies with exam-time temperature by subject, school, and take.



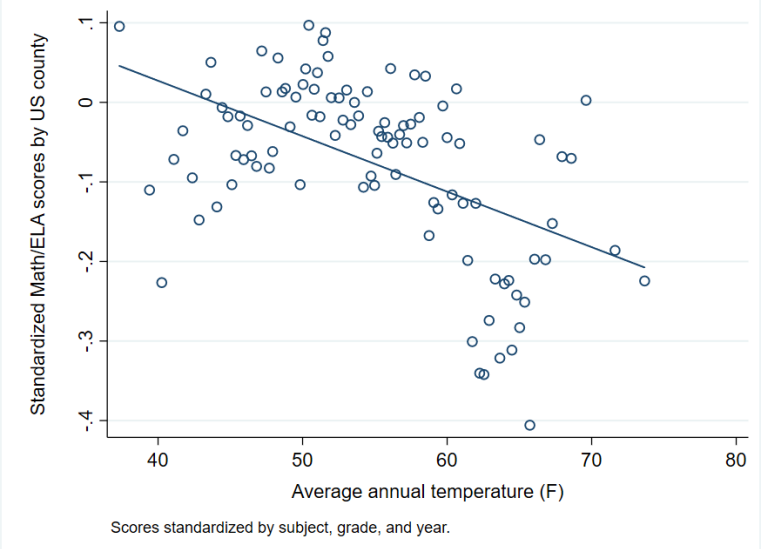
(b) Grade Manipulation expressed as a fraction of scores in manipulable range.

Figure 9: Ex Post Avoidance Behavior: Grade Manipulation

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.



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



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

Figure 10: 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 Reardon et al. (2016)). Average annual temperatures in both cases are measured over the period 1980-2011.