Temperature and Test Scores

Jisung Park
Harvard University Economics Department

March 22 2016

Abstract

Can temperature stress affect student achievement? Using administrative data from 2.6 million high stakes exams in New York City public high schools (1998-2008), I find evidence that heat exposure exerts a negative impact on contemporaneous exam performance and may also reduce cumulative learning attainment. For the average student, taking high school exit exams on a 90F day relative to a 75F day is associated with approximately 0.19 standard deviations lower exam scores and a 12.3% higher likelihood of failing that exam. Looking at the impact of hot days during the school year, I find that cumulative heat stress reduces subsequent test performance by approximately 1.3% of a standard deviation per day above 90F. The average short-run effect masks substantial heterogeneity, with low-performing students and schools experiencing disproportionate impacts. Availability of air conditioning at school seems to have a mild though incomplete protective effect, highlighting the importance of adaptation in estimating the welfare impacts of climate change (Deschenes et al, 2015; Kahn, 2016). These findings support the notion that climate change may affect underlying economic growth rates (Pindyck, 2013; Dell, Jones, Olken, 2014) by reducing human capital accumulation, and recent work documenting persistent impacts of neighborhoods on individuals' economic mobility (Chetty, Hendren, Katz, 2015).

Keywords: Temperature, Human capital, Climate Change
JEL codes: J24, O13, Q54

Acknowledgments

Special thanks to Larry Katz, Raj Chetty, Andrei Shleifer, Robert Stavins, Edward Glaeser, Claudia Goldin, David Cutler, Greg Mankiw, Jim Stock, Geoffrey Heal, Jeff Miron, Jeff Liebman, Mandy Pallais, Joe Aldy, Thomas Matte, Josh Goodman, Sean Reardon, Daniel Sullivan, Sam Stolper, Patrick Behrer, Jing Li, Richard Sweeney, and numerous seminar participants at Harvard, Columbia, UC Berkeley, and the NYC Department of Health and Mental Hygiene for valuable comments and feedback. Thanks also to the NYC Department of Education for data access, and to Nicolas Cerkez, Rodrigo Leal, Kevin Eskici, and William Xiao for excellent research assistance. Funding from the Harvard Environmental Economics Program, the National Science Foundation, and the Harvard Climate Change Solutions Fund is gratefully acknowledged.

1 Introduction

What is the relationship between temperature and educational attainment? Given the centrality of human capital to economic growth (Goldin and Katz, 2009), the well-documented correlation between income and climate across and within countries (Acemoglu and Dell, 2010; Dell, Jones, Olken, 2014), and the imminent prospect of human-caused climate change, this seems a question of deep and immediate policy relevance.

From the perspective of climate policy, ongoing uncertainty regarding the social cost of carbon is due in part to our poor understanding of whether its economic impacts will manifest primarily as shocks to the level of output or as permanent reductions to growth rates (Pindyck, 2013)[1]. From the standpoint of the economics of human capital, it is unclear whether schools should invest scarce resources in protecting students from ambient noise, air pollution, or environmental shocks, in part because there is little empirical evidence that temperature should factor into the human capital production function. More generally, the role that temperature — as opposed to institutional factors which happen to be correlated with temperature — plays in determining long-run economic trajectories both within and across countries remains unsettled (Acemoglu, Johnson, Robinson, 1999; Sachs, Gallup, Mellinger, 2001; Dell, Jones, Olken, 2014).

Recent empirical work suggests that heat stress can have adverse causal impacts on human welfare across a range of contexts. These include effects of hot days on human health (Deschenes et al, 2015), labor supply (Zivin and Neidell, 2014), and labor productivity (Sudarshan et al, 2014; Deryugina and Hsiang, 2014; Zivin et al, 2015), among others. In most cases, the impacts of heat stress are non-trivially large - for instance, labor supply and productivity reductions are found to be on the order of multiple percentage points per degree Fahrenheit above room temperature - suggesting that even mild heat stress can affect not only raw cognitive or physical ability, but also induce behavioral responses in measurable ways.

And yet, apart from evidence from low-stakes laboratory or home settings (Seppanen et al, 2006; Zivin et al, 2015), the role that temperature plays in human capital development remains poorly understood. This study seeks to provide a first step toward understanding whether and to what extent

[1]There is also considerable uncertainty over the distributional consequences of climate change, and how equity-weighting may affect the implied social cost of carbon (Anthoff et al, 2009).

temperature exerts a meaningful influence in the human capital production process, using evidence from high stakes exams in public schools.

Given existing evidence, one might expect extreme heat to affect student learning outcomes through at least two separate but related channels. First, heat stress might reduce students' cognitive ability or effort levels during high stakes exams, resulting in lower scores than would have been the case under optimal conditions. Second, heat exposure during class time over the school year may reduce pedagogically engaged hours[2] resulting in lower learning attainment. From a distributional perspective, if either the overall level of exposure or students' behavioral responses to heat stress depend on income or socioeconomic status, we might expect the same climatic event to have regressive welfare impacts.

To study the impact of temperature on educational outcomes, I use administrative records from New York City public schools, which provide student-level data for high school exit exams from 985 high schools between 1998 and 2009 - exams that are taken in late June of each year. My empirical strategy exploits the randomness of day-to-day and year-to-year temperature shocks to identify the causal impact of heat stress on short-run test performance and long-run learning attainment. Since students are assigned to test dates and locations without prior knowledge of temperature stress or the ability to reschedule, temperature on the day of an exam is unlikely to be correlated with student quality. Similarly, year-to-year fluctuations in the incidence of hot days are unlikely to be systematically correlated with student performance, especially when comparing the same schools over time.

I test three main hypotheses in this paper: first, whether heat exposure during the time of the NY State Regents exam leads to reduced test performance (I will refer to this channel as the "contemporaneous performance impact" of heat exposure); second, whether cumulative temperature stress over the course of a school year reduces learning, measured in terms of performance on these and other standardized assessments taken by students throughout their primary and secondary schooling ("cumulative learning impact"); and third, whether the impacts of temperature exposure vary by student ability or school quality.

The findings reveal that hot temperature exerts a statistically significant

[2]This may occur due to reduced cognitive capacity of students and/or teachers during class, increased likelihood of disruptive incidents, increased absenteeism, or reduced cognitive capacity of students and/or parents during homework time at home.

and economically meaningful impact on short run test performance, both in terms of exam scores and pass rates. Taking a high school exit exam on a 1 standard deviation hotter day (+7.8F) results in a 0.09 standard deviation (1.7 point) reduction in test performance. Put differently, taking an exam on a hot 90F day compared to an average 75F day results in a 0.22 standard deviation (3.4 point) reduction in test scores for the median student. The likelihood of failing any given subject exam increases by 12.3% (5.2 percentage points) per standard deviation increase in test day temperature. These effects are robust to a wide range of specifications, including models with school, year, and student fixed effects, as well as controls for precipitation, snow, time of day, and day of year.

These average treatment effects mask significant heterogeneity across students. High-performing students (measured by elementary and middle school standardized English and Math exam scores) seem to be relatively unaffected by heat stress, while low-performing students suffer disproportionately larger negative impacts: up to five times as large for bottom versus top performers. This is true both in absolute terms — percentage declines in scores and pass probabilities per degree Fahrenheit — as well as relative to their decile-specific average variabilities: percentage declines divided by the subpopulation standard deviations. Similar patterns persist when looking across schools of varying average performance levels as well.

Finally, the evidence suggests that cumulative heat stress may reduce learning attainment due to disrupted class time, though the estimates are less precisely estimated. An additional day above 85F during the school year reduces subsequent test performance by approximately 1.2% of a standard deviation (0.22 points), controlling for the contemporaneous performance impact of temperature on exam days.

Previous research examining students from the same school district suggests that the wage impacts of even minor learning impacts during primary and secondary school can be substantial — on the order of 1.3% lower wages at age 28 for a 0.10 standard deviation reduction in test scores (Chetty, Rockoff, Friedman, 2014). While more evidence is needed to substantiate this link, the existing evidence is consistent with emerging research on the long-run impacts of environmental shocks during childhood and in-utero (Currie et al, 2015; Isen et al, 2015; Lavy et al, 2015).

Using air conditioning installation data from a series of independent school building audits commissioned by the NYC School Construction Authority, I find evidence that air conditioning may mitigate some of the perfor-

mance impacts of temperature stress, though the protective effect is incomplete. This echoes recent findings that estimate the role of air conditioning in reducing mortality responses to heat stress (Deschenes et al, 2015), and underscores the importance (and difficulty) of taking adaptive measures into accurate account when estimating potential climate change impacts.

This article contributes to the literature on the economic consequences of climate change by providing evidence on the human capital impacts of heat stress, which are not currently included as part of the climate damage functions used to assess the optimal price of carbon (Tol, 2009; Pindyck, 2013; Greenstone et al, 2013). It also sheds light on possible mechanisms through which increased temperatures might reduce not only the level of output but also its underlying growth rate, as documented, for instance, by Dell, Jones and Olken (2013), by depressing the effective rate of human capital accumulation over time, and supports the notion that the impacts of climate change may, on net, have highly regressive welfare impacts.

This article also contributes to the education literature by providing evidence of measurable impacts of environmental shocks on student attainment. The results reported here suggest that, especially in poorer school districts, heat-related learning disruption may be an important modifier to other well-documented determinants of the human capital production function such as teaching quality (Chetty et al, 2014) or non-cognitive skills (Cunha and Heckman, 2007). It also suggests that part of the racial achievement gap might be explained by geography, given the higher likelihood of heat exposure for Black and Hispanic minorities due to latitude and urban heat islands[3].

Finally, the substantial heterogeneity in heat-related performance impacts across student types, which tend to be highly geographically segregated, contributes to a growing literature on the importance of local neighborhoods in determining long run economic trajectories for families and individuals (Chetty, Hendren, and Katz, 2015).

The rest of the paper proceeds as follows. Section II discusses the data and institutional context. Sections III and IV present the econometric specifications and results for short-run test performance and long-run learning attainment respectively. Section V interprets the results, and Section VII concludes.

[3]Goodman, Park, and Smith (2016) explore this hypothesis using SAT and AP scores in the United States.

2 Background, Context, and Data

2.1 Temperature and Cognitive Performance

Previous research has documented a negative relationship between temperature stress and human task productivity on cognitive and physical tasks (Seppanen et al, 2006). Numerous lab experiments have shown that task performance falls with heat stress, even for highly trained and well-acclimated individuals such as British naval officers (Mackworth, 1945; Grether, 1973). At extreme levels, heat stress can be deadly, as the body becomes dehydrated and hyperthermia begins to cause dizziness, muscle cramps, and fever, eventually leading to acute cardiovascular, respiratory, and cerebrovascular reactions[4].

A recent economic literature has documented causal relationships between heat stress and productivity in a wide range of economically relevant contexts. These include impacts of hot days on labor supply decisions (Zivin and Neidell, 2014), labor productivity (Cachone et al, 2013; Sudarshan et al, 2014), violent crime (Anderson, 1987; Hsiang et al, 2013) and even local economic output (Deryugina and Hsiang, 2015; Park, 2016)[5]. There is also a suggestion that temperature stress on the day of the exam may affect student performance on cognitive assessments taken at home, though no evidence for cumulative learning impacts have yet been found (NLSY, Zivin et al, 2015)[6].

Few studies to date have explored the impact of temperature stress on students in school environments, or during high stakes exams. Given the increasing importance of human capital accumulation for individual as well as national economic trajectories (Goldin and Katz, 2009), public schools seem to be an important context in which to study the relationship between

[4]Exposure to heat is associated with increases in blood viscosity and blood cholesterol levels (Deschenes and Moretti, 2009), which can eventually cause increased morbidity in the form of heat exhaustion and stroke (Graff Zivin and Schrader, 2015). Combined with humidity and intense physical exertion such as exercise or manual labor, heat can lead to acute cardiovascular or respiratory failure. Deschenes et al (2015) estimate that an additional day with temperatures above 90F cause a 0.11 percent increase in county-level mortality rates in the United States.

[5]Heat stress in-utero has been shown to cause adverse economic outcomes later in life as well (Isen et al, 2015). For a review of this rapidly evolving literature, see Dell, Jones, and Olken (2014) or Heal and Park (2016).

[6]Studying the impact of pollution on cognitive ability, Lavy et al (2015) find a small but significant impact of temperature on Israeli Bagrut exams, which are mandated to be taken in air-conditioned rooms.

temperature and educational outcomes — both in terms of the contemporaneous performance impact of heat stress on the day of the exam, as well as the cumulative learning impact of heat stress over the course of a school year.

2.2 Institutional Context: New York City Public Schools

2.2.1 The New York City Public Schools System

The New York City public schools system is the largest school district in the United States, with over 1,500 schools and 1 million students across the five boroughs. The district features many of the qualities characteristic of large, urban public school districts in the US, including considerable heterogeneity in race and socioeconomic status, generally low student performance, and relatively old building stock.

Approximately 40% of the student body is Latino, 31% are African American, and 14% are Asian. Non-Hispanic whites comprise less than 15% of the public school student population. The vast majority of students are poor, with 78% qualifying for federally subsidized school lunch as of 2014. Many high-income students, especially on the Upper East Side of Manhattan, attend private schools and are thus not included in our sample. However, the NYC public school system includes several prestigious magnet schools, 'Specialized Schools' such as Stuyvesant Academy and Bronx Science, which consistently rank among the nation's best, resulting in a wide spectrum of student ability and achievement even within the public school district.

The 4-year graduation rate, at 70%, is below the national average of 81% but comparable to other large urban public school districts. Once again, system-wide averages mask remarkable 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 1). In contrast, schools in higher income neighborhoods of Manhattan or Staten Island have average graduation rates above 75%.

The typical NYC public school building was constructed in 1932, meaning that the vast majority were built before centralized air conditioning systems began to be included as the default for new construction. Though many have been significantly renovated since inception, approximately 38% report having no air-conditioning equipment as of 2012. Of the 62% that have air conditioning, 42% (302 out of 719) have defective components, according to

building conditions assessments conducted by third-party auditors. The lack of adequate air conditioning in NYC public schools has been suggested by parents, teachers, and school administrators[7] as a potential impediment to learning achievement, though formal empirical assessments have not been conducted to date.

2.2.2 New York State Regents Exams

The primary measure of performance used in this study is the NY State Regents exam. These exams comprise a series of standardized subject assessments which are administered by the New York State Education Department (NYSED), and are more commonly known as high school exit exams. The purpose of these examinations is to determine students' eligibility to receive a high school (Regents) diploma, in addition to providing students with baseline assessments of college readiness by core subject (e.g. Algebra, Physics).

These are high stakes exams for the average NYC student. Students must receive a passing score on at least five core Regents subjects to obtain a Regents diploma, and many public universities (e.g. CUNY) use Regents score cutoffs in college admissions decisions[8] [9]. Students often take several Regents exams per year, usually in the spring of 10^{th}, 11^{th}, and 12^{th} grade.

Because the exams in these core subjects are administered over the course of 1 to 2 weeks in late June[10] — when the average daytime high temperatures

[7] `http://www.nytimes.com/2015/06/24/nyregion/new-yorks-public-school-students-sweat-out-the-e`
`html?_r=0`

[8] Prior to 2012, the passing score for a Regents Diploma was 65, but low-performing schools were able to offer 'Local Diplomas' with a less stringent passing requirement of 55 or above on the five core exams. As of 2012 (the cohort of students who were 9th graders in 2008), the Local Diploma option was no longer available, and the passing threshold became 65 or above for all students except those with known disabilities.

[9] The core subjects are Math (Integrated Algebra, Geometry, and/or Trigonometry), English, Science (Physics, Earth Science, Biology, or Chemistry), US History & Government, and Global History & Geography.

[10] Regents exams are also offered during two additional waves each year — once in August, at the conclusion of summer school, and again in January, prior to the start of spring term. These scores are included as part of the NYC DOE data set, however I focus only on the subset of June exams, for two reasons. First, both August and January waves likely feature substantial selection on unobservables. Summer school is primarily remedial for public school students in NYC, meaning that August exams are taken only by a small subset of under-performing pupils. Similarly, January exams are taken by students who

are 79 degrees Fahrenheit and frequently exceed 90F (32C) — it is not uncommon for students to be exposed to disruptive temperatures on exam day. It is also possible for students to be exposed to heat stress during class time in the semesters leading up to the exams, since the school year commences in early September — when average daytime highs are around 76F, and frequently veer above 90F — and ends during the last week of June[11]. Moreover, 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. This means that low ability students are more likely to be exposed to potentially disruptive heat exposures throughout their careers.

Test dates and times are determined over a year in advance, and synchronized across schools in the NYC public school system to prevent cheating. Each exam is approximately 3 hours long, with a morning session that begins at 9:15am and an afternoon session beginning at 1:15pm. The exact dates and ordering of subjects within testing period vary from year to year, allowing for additional identification of possible temperature impacts using the interaction between daily temperature and afternoon/morning status, as reported below.

Exams are centrally administered by the NYSED and, with a few exceptions, follow a consistent format over the study period, as tests in all subjects and years in our sample are written by the same state-administered entity. Exams are scored on a 0-100 scale, with scaling conducted according to subject-specific rubrics provided by the NYSED each year[12]. All scores are

---

are enrolled in classes outside the usual order or cycle, either due to advanced standing or remediation. If, for instance, August exams are makeup exams for students who failed their first attempt (in June), including these entries in the broader sample may bias the results of this analysis. Second, unlike the main June wave, which is spread out over the course of 5 to 10 days, August and January exams are both administered over two days each year, reducing the amount of daily temperature variation available for identification. Thus, I focus only on June exams in the analysis.

[11]The record daytime high temperatures for June and September in Manhattan are 101 and 102 degrees F respectively.

[12]These scores are in principle comparable across years as well, though, as is often the case with unstandardized test results, it is difficult to gauge the nature and extent of year-to-year changes in scoring stringency arising from policy-induced changes at the State level. A simple solution to this problem would be to standardize scores by subject and year. However, if changes in average scores at the school or district level are driven in part by year-to-year differences in temperature, as our model suggests, standardizing scores would net out such aggregate differences, allowing only the identification of differential

therefore comparable across schools and students within years. In principle, they are comparable across years as well, though there are many reasons to believe this may not necessarily be the case in practice, leading us to prefer a model that incorporates year fixed effects when possible.

In summary, using Regents scores from the NYC DOE offers several distinct advantages. First, they are high stakes exams used to determine diploma eligibility and possibly affecting college enrollment. Second, they are offered at a time of year when temperatures are likely to be hot but not always so, due to the considerable variability in day-to-day temperatures in June. And because they occur at the end of the school year, they are also more likely than periodic assessments to reflect cumulative impacts of heat stress that may have accrued over the course of the school year. Third, they are taken by a diverse mix of students, which allows for rich heterogeneity analysis[13]. Fourth, 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 or by selective grading criteria[14]. Fifth, location and timing information for each student can be matched to ambient temperature data for that day, as well as to air conditioning data for the school in which the students were enrolled.

2.3 New York City Department of Education Data

I obtain student-level administrative data from the New York City Department of Education (NYC DOE). The data is de-identified to protect students'

impacts across schools or student groups within subject-year groupings. As such, I use scaled (unstandardized) Regents scores as the primary dependent variable in the analysis, including controls for district-wide trends in average scores as well as, in the case of identifying short-run performance impacts, school-by-year-specific fixed effects.

[13]It also makes it more likely to be able to infer impacts of temperature stress in other urban contexts where the demographic makeup may differ.

[14]One possible drawback of using the Regents scores may come from their high stakes nature. While they are centrally administered, they are locally graded (albeit with a series of controls provided by the State), which means that teachers may be tempted to manipulate scores to help students clear the passing threshold. Indeed, Rockoff et al (2013) document evidence for manipulation at the failing cutoffs for a small proportion of the exams: approximately 3-5% of all Regents exams between 1999 and 2011. It is worth noting, however, that any manipulation of scores would likely attenuate estimates of temperature-induced performance reductions downward toward zero.

privacy, and includes the universe of all public school students who took one or more Regents exams over the period 1999 to 2014. This corresponds to all students in the 1996 to 2014 cohorts, with cohorts defined as year of high school entry. Each year includes approximately 75,000 students.

The data includes exam scores recorded at the student-, subject-, year-level, with students represented by unique 9-digit ID numbers. Each cell includes information about the school in which the student was enrolled during that school year. I assume that students took their exams at the school in which they were enrolled, which is standard protocol.

While the administrative dataset is incredibly rich, exam dates are not provided in the student-level data. As such, I obtain exam dates and times from publicly available exam schedules, which provide the date and time of each NY Regents exam taken by NYC public school students over the past two decades. However, due to inconsistencies in the way exam subjects and terms are coded in later years, I drop years 2010 to 2014, and use only the records of exams taken in the years 1999 to 2009. The exact matching process, in addition to the rationale for limiting the sample to these years, is described in more detail in the Appendix. Moreover, demographic control including ethnicity, gender, disability status, and whether or not the student was eligible for reduced price lunch are only available for cohorts entering after 2006, limiting the student-level heterogeneity analysis.

2.3.1 NYC Building Condition Assessment Surveys

Information on building-level air conditioning equipment comes from the New York City School Construction Authority (SCA), which administers detailed, building-level surveys for NYC public schools[15].

While the universe of survey results were not publicly available, building-specific 2012 BCAS reports were available in pdf format on the school's website for a large subset of schools (94%). I obtained survey records for 1,159 school buildings by downloading each pdf, scraping information pertaining

[15]In a 1989 legislative mandate, the New York State Education Department set specific inspection and reporting requirements for school buildings, largely in response to reports of bureaucratic bloat and corruption in the contract bidding and procurement process for school infrastructure projects. The SCA was charged with the task of conducting a series of Building Condition Assessment Surveys (BCAS) for every school building every 5 years. These surveys were carried out by a team of engineers — employed by independent contractors — who recorded detailed information about each building's mechanical and electrical systems according a pre-specified rubric.

to air conditioning equipment, all of which are recorded in the same format. The 1,159 records constitute 94.5% of the 1,226 unique school addresses in the New York City public school system as of 2012[16].

The BCAS surveys include detailed information on air conditioning equipment, which can be used to identify schools in which — and potentially school-specific-years during which — air conditioners may have been operational and thus able to offset some if not all of the heat stress induced by high outdoor temperatures[17]. I use this information to construct an indicator variable for whether or not a school building had non-defective central AC installed in any given year, using the earliest year of installation as the first year in which any AC was available for each school.

This is an imperfect proxy for the true variable of interest, which is the effective cooling achieved by school by day and year, since current installations may have replaced existing (older) ones. Furthermore, we do not know whether, on any given day or year, schools that had AC equipment actually used them, or whether all students benefited from them equally. However, I take existence of air conditioning equipment as a necessary condition for effective cooling on hot days, and use this information to test whether the impact of heat stress is mitigated in schools that have AC equipment relative to ones that don't.

2.4 Weather Data

Temperature and precipitation data are taken from Schlenker et al (2011), who provide daily min, max, and mean temperatures, precipitation (in inches)

[16]Many NYC public schools are housed in buildings that contain multiple schools, especially in the case of middle and high schools. As such there are approximately 1,550 schools located at 1,226 unique addresses in our study sample.

[17]Information available included an indicator for whether or not there was an air conditioning unit present, as well as information regarding the type of unit present, which allows us to provide some differentiation between schools with 'central AC' systems and those with window units, though the classification scheme is not completely straightforward. Types of units include 'packaged rooftop system', 'DX split system', 'chilled water system', and 'packaged terminal system'. Assuming that 'packaged terminal systems' refer to window units, and 'chilled water systems' I conservatively classify schools with 'packaged rooftop' and 'DX split' systems as having 'central AC'. In addition, there is information on unit capacity (e.g. 50 tons), overall unit quality on a 1 to 5 scale as assessed by the engineer for each unit present, whether or not any system parts were found to be defective, as well as the year in which the unit was installed. If multiple units are present, information on each type of unit is recorded separately.

and dew point information from a national network of weather stations, area-weighted to the Borough (county) level over the period 1950-2011. These variables are matched geographically using street addresses for each school, assigning daily temperatures by Borough[18]. Annual snowfall totals for New York City are taken from weather station readings in Central Park and assigned uniformly across all schools in the city.

While we would in an ideal world have a measure of temperature inside the test room, temperature inside a non-air-conditioned building is presumed to be highly correlated with the ambient temperature outdoors, and still somewhat correlated in the case of buildings that have air conditioning equipment but comprise old building stock, which often feature poor ventilation and leaky thermal envelopes.

Given urban heat island effects, and the occupational density of many school buildings (which often results in multiple floors), we would expect indoor air temperatures to be significantly warmer than outdoor temperatures in the case of schools without adequate AC. Moreover, the presence of AC heat exhaust spillovers would mean that a non-air conditioned building would bear the local externality of surrounding buildings that do have AC, a likely phenomenon given the nearly universal rates of commercial air conditioning in New York City (NYC Department of Health and Mental Hygeine, 2015).

2.5 Summary Statistics

2.5.1 Test Scores

Table 1 presents summary statistics for the 730,000 students and 2.7 million exam scores over the period 1998-2009 that form the basis for our analysis.

The median student takes 5 exams in the core subjects represented in this data set. On average, students in NYC public schools score below college proficiency, though there is considerable heterogeneity by borough and district. The mean score across all students and years in the sample is 64.7, with a standard deviation of 18.

[18]This is an admittedly crude assignment methodology, given the potential for spatial variation in temperatures within Boroughs on any given day or in a given year. However, using a secondary weather dataset from the PRISM model, which provides gridded reanalysis data at the daily level by 800m by 800m resolution, I verify that cross-sectional variation is relatively small compared to day-to-day variation in temperatures, or year-to-year variation in the number of hot days.

Less 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.

2.5.2 Heat Stress

Figure 2 presents a histogram of daily temperature readings as reported during the exam dates relevant to our sample. Exam-day temperatures range from a low of 64F to a high of 92F. Figure 3 shows how temperature changes from day to day during a typical exam month (June), illustrating graphically the source of short-run identifying variation.

To measure the cumulative learning impacts of temperature stress, I measure the number of hot days during the school year, using start and end dates for each of the school years in our sample[19]. These weather variables are constructed by summing extreme heat day indicator variables over the course of a school year leading up to the first day of exams. The indicator variables denote whether a particular day had maximum temperatures in a given temperature bin, and are divided into 5F increments over the school year (e.g. 80F-85F, 85F-90F).

Figure 4 presents the incidence of extreme heat stress per school year over the sample period by borough. There is significant year-to-year variation in the number of days with maximum temperatures above 80F. There also seems to be non-trivial cross-sectional variation in cumulative heat stress across schools in different boroughs.

On average, NYC students experience between 19 and 39 days above 80F per school year, with a mean value of 26.7 and a standard deviation of 5.6. The average number of extremely hot days with temperatures above 90F is 2.5, with a standard deviation of 1.3. Most of these days occur during the months of September, October, and June. Summer school students are, on average, subject to an additional 9 days above 90F on average. Despite documented warming for the US and the world as a whole over the past several decades, temperatures in the NYC area seem to have remained relatively stable over the study period (tests for stationarity and trend-stationarity do not suggest time trends in these extreme heat day variables).

[19]School year start and end dates, which vary each year but tend to be from early September to late June, are taken from the NYC DOE website.

3 Research Design and Results: Short-Run Performance Impacts

The following sections outline the empirical strategy and main results. I start with the estimation of short-run performance impacts — i.e. the contemporaneous impact of hot temperatures on exam performance. I then move to identifying the cumulative impact of heat stress and temperature-related disruptions over the course of the school year on human capital accumulation, controlling for the impact of these short run impacts of exam-day heat exposure.

3.1 Short-Run Performance Impacts: Empirical Framework

In estimating the short-run cognitive impact of temperature on test scores, the identification strategy exploits the fact that NYC public school students take their exit exams over the course of several days each year. This enables one to examine the relationship between temperature exposure and test performance across the same student's exams. Specifically, I compare scores for exams taken on a hot day with scores from the same school on a cooler day during the same year, controlling for average differences in overall performance or difficulty of exams across subjects, by fitting variations of following econometric model:

$$y_{ijtTs} = \beta_0 + \beta_1 Temp_{jtTs} + \beta_2 X_{jtTs} + \gamma_s + \eta_j + \theta_T + \eta_j \times \theta_T + \epsilon_{ijtTs} \quad (1)$$

Here, y_{ijtTs} represents the score of student i, school j, test date t, year T, and exam s; $Temp_{jtTs}$ denotes daily maximum temperature experienced by school j during that particular exam; X_{jtTs} is a vector of control variables including daily precipitation, dew point, and the time of day of the exam (AM vs PM); γ_s denotes subject fixed effects to account for average differences in difficulty across subjects (Algebra vs US History); η_j denotes school fixed effects; θ_T denotes year fixed effects; and ϵ_{ijtTs} is a student, school, date, subject, and year-specific error term.

The parameter of interest is β_1, which captures the impact of contemporaneous heat stress on exam performance, controlling for the fixed effects noted above. Given that the model includes school-by-year fixed effects, β_1 is identified by day-to-day fluctuations in temperature over the course of any given

exam wave (5-10 days each June). This means that there are approximately 47 different exam-level temperature treatments for the city as a whole, and 235 treatments by Borough, which is the level at which official weather stations are available[20]. To the extent that the temperature data used here is a noisy estimate of the true experienced temperature in each school, we expect the estimates of β_1 to be attenuated toward zero (Maddala, 1992; Stock and Watson, 2007).

The main identification assumption is that variation in temperature across exams is not correlated with student ability or other unobserved factors affecting student performance. This seems a plausible assumption given that (1) tests are effectively compulsory, in the sense that students must take (and pass) them in order to graduate with a Regents diploma, and without any opportunity for rescheduling; (2) test dates are set by the NY State Education Department over a year in advance; and (3) all students must take their exam at their local high school[21].

I also run a specification that includes student fixed effects in order to verify that unobserved changes in student composition at the school-year-school-date level are spuriously driving the results.

$$y_{ijtTs} = \beta_0 + \beta_1 Temp_{jtTs} + \beta_2 X_{jtTs} + \rho_i + \epsilon_{ijtTs} \tag{2}$$

Here, ρ_i denotes individual i's ability, as approximated by her average score across all core Regents exams she takes over her high school career (median=5; 90th percentile = 11). The richer specification adds approximately 730,000 fixed effects, one for each student id, and assumes that the average scaled score across all 5+ subjects provides useful information regarding a student's underlying ability. I also examine variants of Equation (2) which

[20]Localized urban heat islands may be especially relevant in this context, as differences in tree cover, local air currents, and surface reflectivity can account for differences in summer surface temperatures of more than 20F between neighborhoods within New York City (Gaffin et al, 2008). Assuming that school-to-school differences in temperature on any given day are significant, this suggests up to 112,142 (47 days x 2 morning/afternoon sessions x 1,193 school locations) true underlying treatments if one allows for spatial variability in temperature due to local atmospheric patterns and urban heat island effects. In principle, using a more spatially refined measure of temperature should enable the econometrician to leverage these treatments.

[21]Note that Equation (1) also includes school-by-year fixed effects, to control for annual shocks that are correlated at the school level. This may correspond to the firing of a principal or a string of violent outbreaks at a particular school.

include year-, subject-, and school-fixed effects, as above, as well as a specification that includes a students' average previous standardized ELA and Math scores from elementary and middle school as an alternative measure of previous ability[22].

Without any fixed effects to absorb unobserved variation in schools or students, one might expect the above identification assumption to be violated due, for instance, to residential sorting on climate amenities. For example, if poorer schools are located in more densely urbanized environments within the city, which are more likely to experience heat stress due to urban heat island effects (less tree cover, lower surface albedo), pooled OLS may overstate the causal relationship between heat stress and test scores. Conversely, if housing in more densely populated areas tend to be more expensive, the opposite may be true.

3.2 Short-Run Performance Impacts: Main Results

Figure 5 presents a binned scatterplot that motivates the regression analyses that follow. This graph is derived from the estimation of the version of equation (1) that includes year, school, subject, and school-by-year fixed effects, and plots residual test scores on outdoor temperatures during the time of the exam. Figure 6 presents a similar binned scatterplot derived from a model with student fixed effects. The figures demonstrate that, on average, students perform worse than their school- and individual-specific averages when they face hotter than average temperatures.

Table 2 reports baseline results for the short-run performance impact of heat stress on exam performance across a variety of specifications. In columns (1), (2), and (3), I report the relationship between test scores and daily max temperatures on the day of the exam. In columns (4), (5), and (6), I report the results of running the baseline specifications with temperature values that account for diurnal differences within a given test day (morning versus afternoon).

In both sets of regressions — with and without corrections for afternoon temperature — I report coefficients for models that include school- and school-by-year effects (columns 1 and 4), school-specific trends (columns 2 and 5), and student fixed effects (columns 3 and 6). All regressions in Table 2

[22]The construction of the average previous standardized scores indices is described in more detail in the heterogeneity analysis section.

include subject and year fixed effects in addition to controls for precipitation and time of day.

In column (1), I estimate that a 1 degree F increase in test day temperature is associated with a 0.173 (se = 0.019) percentage point reduction in the student's exam score, significant at the 1% level. This means that a student taking an exam on a 90F day as opposed to a more typical 75F day ends up receiving a -0.154 standard deviation lower score (-2.65 points). It is worth noting that most studies of temperature stress find 65F to be the optimal temperature zone for thermal comfort and cognitive function. This suggests that the total impact of 90F day is on the order of -0.25 standard deviations (-4.37 points).

Column (2) suggests similar impacts when modeling school-specific linear time trends as opposed to school-by-year fixed effects. Column (3) suggests that the negative impacts of temperature stress persist when measured relative to a student's individual ability, which I approximate using her average score across all subjects taken (often over the course of 3 years). The magnitude of the impact is approximately $1/12^{\text{th}}$ to $1/10^{\text{th}}$ of a standard deviation reduction in score per standard deviation increase in temperature.

The impacts documented here are similar in magnitude to the effect of increasing teacher value-added by 1 standard deviation (Chetty et al, 2014), or reducing class size from 31 to 25 students (Angrist and Lavy, 1999). However, as discussed in greater detail below, the welfare implications of these raw average treatment effects are presumably very different (much smaller), because the static, first-order effect is to introduce noise in the signal extraction process for a given level of human capital, rather than reduce the amount of human capital attained[23].

Table 3 presents the results from running equation 2 — the model with student fixed effects — with and without subject and year fixed effects. As columns (1)-(3) show, the impact of temperature stress remains consistently negative and significant at the 1% level across specifications.

[23]Potential dynamic welfare consequences may arise if lower-than-expected exam scores lead to discouragement and reduced future educational effort, or if there are significant sheepskin effects for those who are pushed below the diploma-receiving threshold due to temperature stress.

3.2.1 Robustness Checks

In each of the specifications reported in columns (1) to (3), there seems to be a negative impact of taking an exam in the afternoon as opposed to the morning: -0.798 (se = 0 .164) percentage points lower scores controlling for subject, year, school, and school-by-year fixed effects.

This could be due to fatigue as the day progresses, as has been documented in the context of Danish standardized exams (Sievertsen et al, 2016). It could, however, also be a product of correlation between time of day and temperature, given diurnal variations in temperature.

Using nighttime minimum temperatures from the previous day, I construct a measure of experienced temperature during the time of the exam (9:15am-12:15pm for morning sessions, 1:15pm-4:15pm for afternoon sessions)[24]. Running equation 1 using this time-of-day corrected temperature measure, I find a stronger negative relationship, reported in column (4): β_1 equals -0.221 (se = 0.025)[25]. Columns (3)-(6) of Table 2 and Table 3 report results of running variants of equations 1 and 2 described above using this time-of-day corrected temperature metric, including a dummy for Time of exam. All feature slightly larger estimates of β_1, suggesting that diurnal temperature differences are important in understanding the impact of temperature stress on exam performance.

It is also possible that students are simply more fatigued toward the latter dates of the testing period, which may lead to spurious correlation between scores and temperatures since it is more likely to be hot later in the month than earlier. Adding controls for day of month reduces the amount of residual temperature variation considerably, since the main specification already includes subject and year fixed effects. However, running equation (1) with

[24]Specifically, I take the mean value of the previous day's minimum (nighttime low) temperature and the current day's daily maximum temperature as the 9am morning temperature. Afternoon exams are assigned the daily max temperature, which usually occurs around 1pm.

[25]In this specification, the coefficient on the Time of Day dummy is smaller, though still significant, suggesting that students taking exams during the afternoon suffer adverse impacts due to both greater heat stress and greater fatigue (or that our simple AM/PM temperature correction model is biased). This is consistent with Sievertsen et al (2016) who find that Danish students taking standardized exams later in the day do significantly worse, and that students taking exams after recess do better, though the magnitudes of time of day are much smaller than the impact of temperature documented here; they find a 1.5% of a SD decline for students taking exams at 1pm as opposed to 9am.

controls for day of month yields a negative and significant estimate of -0.127 (se =0.026) per degree F, suggesting that spurious correlations between heat and fatigue are not driving the results. Adding controls for up to three days' lagged temperatures, I find that the negative impact of contemporaneous temperature persists, and, if anything, becomes more negative, though this is likely due to correlation between temperatures across days[26].

While the evidence presented here strongly suggests that temperature has a causal impact on exam performance, one cannot rule out the possibility that atmospheric factors such as ozone, which has been shown to be positively correlated with temperature and potentially harmful to health and cognitive function, are driving the relationship uncovered here. Though not empirically verified in the context studied here, existing epidemiological research suggests that, building canyons in highly urbanized contexts such as New York City typically have a mitigating effect on ozone, as O_3 molecules are broken up as they descend from atmosphere to street-level (Hogrefe et al, 2004)[27]. The strongest documented impacts of air pollution on cognition and health seem to be from particulate matter such as PM 2.5 (Lavy and Ebenstein, 2015), which is typically not systematically correlated with temperature but rather wind speed and direction. SO_2 emissions typically occur during winter months, so are likely not a factor during June Regents exams.

3.2.2 Pass Rates

To assess whether temperature affects pass rates, I construct a linear probability model where the left hand side variable is a dummy indicating whether or not the student scored above the passing threshold on any given exam, and run variants of equations 1 and 2 from above (school-fixed effects and

[26]That is, during the month of June across the years in our sample, temperature on day t is strongly positively correlated with temperature on day t-1, but weakly negatively correlated with temperature on day t-2, suggesting normal periodicity in atmospheric patterns that cause hot or cool spells. Given the relatively short panel length, it is difficult to credibly estimate lagged versus contemporaneous impacts separately, though it would be interesting to know whether heat during the night before an exam is more or less damaging than contemporaneously, and also whether students adapt or suffer greater impacts during prolonged heat waves.

[27]This is in contrast to agricultural settings, where flat, open air environments mean that realized ozone exposure by workers is often much higher and can lead to adverse impacts on productivity and health (Graff Zivin and Neidell, 2015).

student fixed effects models):

$$p_{ijtTs} = \beta_0 + \beta_1 Temp_{jtTs} + \beta_2 X_{jtTs} + \gamma_s + \eta_j + \theta_T + \eta_j \times \theta_T + \epsilon_{ijtTs} \quad (3)$$

$$p_{ijtTs} = \beta_0 + \beta_1 Temp_{jtTs} + \beta_2 X_{jtTs} + \rho_i + \epsilon_{ijtTs} \quad (4)$$

Here, p_{ijtTs} represents the probability that individual i in school j scores above the passing threshold on the exam administered on date t in subject s during year T, and is estimated using an indicator variable for whether or not the score was above the 65 point Regents Diploma cutoff.

Table 4 presents the results from running the main specifications above using this passing indicator as the dependent variable.

Column (1) shows that, for the median NYC public school student, taking an exam on a 1 degree F hotter day results in a 0.54 (se = 0.057) percentage point lower chance of passing that exam, significant at the 1 percent level[28]. This suggests that a day with temperatures around 90F results in a 13.7% lower pass probability (19.3% increase in a chance of failing), compared to a 75F day, given a mean pass rate of 58.7%. Columns (2) through (6) show that this effect is robust to a wide range of specifications, again including models with or without student fixed effects and year fixed effects.

4 Short-Run Performance Impacts: Heterogeneity Analysis

In this section, I examine heterogeneity in the treatment effect of test-day temperature.

The motivation is to examine whether certain students are more sensitive to the performance impacts of temperature stress, due to behavioral or other factors which may differ across income lines.

There are several reasons one might expect the impact of heat stress to be different across students of differing socioeconomic status, including different levels of executive function and non-cognitive skills as well as home or school environment.

[28]This translates to approximately $1/12^{\text{th}}$ of a SD reduction in pass probability per SD increase in test day temperature.

4.1 Executive function, Non-Cognitive skills and resilience to distraction

Children who grow up in poverty have been shown to exhibit reduced executive function, which is critical in maintaining focus in presence of distractions (Fitzpatrick et al, 2014). It is possible that low ability students are more likely to be thrown off course due to the distractive influence of environmental factors such as temperature, noise, or pollution. For instance, Lavy et al (2015) find that the test score impact of elevated particulate pollution is up to four times higher for low ability and low SES high school students in Israel. Similarly, students with low non-cognitive skills may be more easily discouraged by adverse test-taking conditions, suffering more negative impacts, while students of similar cognitive ability but higher emotional resilience may be able to maintain concentration despite the heat.

4.2 Reduced Effort in the context of Uncertain Returns to Schooling

Low-ability or low-SES students may reduce effort due to greater uncertainty over the returns to human capital investment, which becomes magnified in the context of adverse test-taking conditions.

Suppose the impact of reduced cognitive function on the translation between existing human capital and achieved test score is multiplicative. That is, ones test score is a function of underlying human capital plus some random noise - noise which becomes magnified under hot conditions. Assume both low and high SES individuals suffer the same direct disutility of heat stress, but that low SES individuals face a more uncertain mapping between test score and future wages, due perhaps to fewer family connections. Then one may expect lower SES students to reduce effort by more in response to the same temperature exposure. Similarly, if low ability students have higher underlying score variance, and heat stress affects this variance multiplicatively, then low ability students might face a more substantially reduced expected score than high ability students. If expected wage returns are linear in score, the same heat shock will cause a greater reduction in expected wage returns for lower ability (higher variance) students, leading to a greater reduction in optimal effort exerted by low ability types relative to high ability types.

4.3 Home and School environment

In addition, it is possible that low SES students are more exposed to heat stress during the night or morning before the exam, given lower rates of air conditioning penetration in lower income neighborhoods. In New York, households in the South Bronx or Northern Brooklyn are significantly less likely to report having air conditioning at home than in more affluent neighborhoods such as Upper Manhattan (NYC Department of Health and Mental Hygeine, 2005).

Perhaps most directly, the availability and utilization of air conditioning in the test taking environment seems to be a crucial mediating factor which may depend on student characteristics such as parental income. I explore the direct impact of air conditioning availability separately below.

4.4 Effect Heterogeneity: Results

I measure effect heterogeneity along two dimensions: (1) individual ability, and (2) air conditioning status, which is one of many potentially relevant measures of school quality.

4.5 Student Ability

Individual pre-test ability is approximated using linked administrative records from elementary and middle school, which provide standardized math and English (ELA) exam scores for a large subset of the individuals in my sample (approximately 63% of students)[29]. Specifically, I group students into percentiles of the elementary-and-middle school ability distribution, averaged across the entire district over the sample period, using an arithmetic mean value of scaled scores from standardized English (ELA) and Math assessments which are administered annually between grades 3 through 8. I take the average z-score across all records that are available for each student, and rank them according to percentiles of this prior performance distribution.

As shown in Table 5, this measure of prior achievement seems to capture useful information regarding student ability, indicated by the tight correlation between prior rank and Regents exam performance. Students in the top 10%

[29]There is evidence suggesting that non-cognitive ability and cognitive ability are highly correlated, and would be expected to be correlated with average school performance. I cannot use Regents scores, however, as they are endogenous to the temperature regressors.

of the prior achievement distribution receive an average score of 85.5 on their Regents exams, compared to 44.6 for students in the bottom 10%. Similarly, students in the top 10% pass virtually all of their core Regents subjects (97% pass rate), compared to the bottom 10% students, who pass on average only 13% of the time. As shown by the average combined z-score over all students in the high school dataset, +0.38, students who stay in school long enough to be eligible to take Regents exams tend to be above average in elementary and middle school.

Running equation (1) by quantiles of individual achievement yields substantial differences in effect sizes by level of ability. The impact of a 1 degree F hotter day is -0.081 (se = 0.021) percentage points for students in the top decile and -0.439 (se = 0.021) for those in the bottom decile (Figure 7). This suggests that the raw score impact is approximately 5 times as large for students in the lowest decile of the prior ability distribution compared to those in the top decile.

One might be concerned that this is an artefact of top-truncation of scores for top-performing students. Even when expressed as proportions of the percentile-specific standard deviations, however, this discrepancy persists: an impact of -0.065 SD in test scores per SD in temperature for the top decile, -0.10 for the middle deciles, and -0.213 for the bottom decile.

Similarly, for students in the top 5% bin, the impact of an additional degree F outdoor temperature is -0.09 points, significant at the 1% level. In contrast, bottom 5% students do worse by -0.51 points per degree F, significant at the 1% level.

4.6 Air Conditioning

To assess the possibility of adaptive responses in the form of air conditioning, I examine a model that includes an interaction between temperature and AC status, where AC status is defined as the presence of non-defective central AC by school and year. Table 6 provides average AC status by Borough, and Figure 8 presents this information on a cartographically.

Contrary to economic intuition, there appears to be a negative relationship between school-level achievement (measured both in terms of average scores or average graduation rates) and air conditioning penetration. A simple OLS regression of score on AC status suggest that schools with non-defective AC equipment are lower-performing on average (-3.9 points, or -0.21 standard deviations).

This negative correlation, combined with the heterogeneity in the effect of temperature by average student ability and school-level performance noted above, suggests that it is important to examine the role of air conditioning in the context of student quality. This is especially true in light of the fact that we have potentially very noisy data on air conditioning; the data is on AC installation status as of 2012, as opposed the more ideal AC usage by date or year.

To assess the impact of air conditioning, controlling for effect heterogeneity by student ability, I run equation (1) on the subset of schools with graduation rates in the top decile, with an interaction term between exam-time temperature and AC status. Table 7 presents the results for top and bottom 10% of the student ability distribution, as well as top and bottom terciles. High-performing (top decile) schools without AC are significantly more sensitive to temperature stress than high-performing schools with AC: -0.102 (se = 0.022) versus -0.033 (se = 0.021) points per degree F. Interestingly, low-performing students (bottom decile) in schools without AC are similarly sensitive to temperature stress compared to low-performing students in schools with AC.

These gradients persist across the school performance distribution, and are robust to alternative characterizations of AC status, including ones that infer the school-years during which any given school did or did not have AC by using date of installation information.

5 Long-Run Learning Impacts

5.1 Long-Run Learning Impacts: Empirical Framework

The empirical strategy for measuring cumulative learning impacts is similar to the approach used for contemporaneous performance impacts. Our main focus is on comparing educational outcomes in years with more cumulative heat exposure during class time, controlling for the impact of temperature during the exam itself. I estimate models of the form:

$$y_{ijtTs} = \alpha_0 + \alpha_1 Temp_{jtTs} + \alpha_2 Days_{jT} + \alpha_3 X_{jtTs} + \alpha_4 Year + \gamma_s + \eta_j + \epsilon_{ijtTs} \quad (7)$$

where y_{ijtTs} denotes exam score for student i in school j for a Regents exam taken on date t of year T in subject s; $Temp_{jtTs}$ denotes outdoor temperature in the vicinity of school j on that date and time; $Days_{jT}$ is a vector of

temperature-day variables which denote the number of hot days per school year, described in greater detail below; and X_{jtTs} denotes a vector of controls for precipitation and humidity on the day of the exam, annual snowfall, annual precipitation, and a controls for the number of days in other — cooler — temperature bins.

As before, the specification includes school and subject fixed effects. However, because variation in long-run heat exposure is measured at the annual level, we cannot use year fixed effects while retaining meaningful variation, and instead explore versions that include global- as well as district-specific time trends in test scores.

The coefficient of interest, now denoted α_2, is the impact of hot days during the school year on test performance in subject exams taken at the conclusion of that school year. The incidence of hot days is measured by tallying the number of days between 80-85F, 85-90F, 90F and above during each school year in our sample, beginning with fall semester fall dates and ending the day before the first Regents exam. The control vector, X_{jtTs}, includes the number of days in all other temperature bins in 5 degree increments (0-5F, 5-10F. . ., 75-80F), omitting the 'optimal' bin, which I take to be 60-65F[30].

Because there are only 10 years and five boroughs in our sample statistical precision is greatly reduced relative to analysis of contemporaneous performance impacts. As before, standard errors are clustered at the school-by-year level.

5.2 Long-Run Learning Impacts: Results

Figure 9 presents a binned scatterplot that motivates the analysis of regression results that follow. It shows a negative relationship between end-of-year Regents exam scores and the number of days per school year with daily maximum temperatures above 80 degrees F, controlling for the variables in equation 7 (including the number of days in other temperature bins, binned

[30]This is based on running equation (7) for all possible optimal bins and choosing the specification for which coefficients on days falling on either side of the threshold exhibit the most consistent negative pattern. In this sense, we are imposing a parametric assumption of a single-peaked relationship between temperature days and 'productivity', based on previous work exploring the impacts of temperature days on county-level output (Deryugina and Hsiang, 2015; Park, 2016), mortality (Deschenes et al, 2015), and agricultural yield (Butler and Huybers, 2013).

in 10 degree intervals).

Table 8 presents the results from running versions of equation (7) that include global (columns 1 and 3) and district-specific time trends (column 2), and degree days binned at 10 degree F (1 and 2) and 5 degree F (column 3) intervals. Columns 1 and 2 suggest that students who experience an additional day above 90F during the school year experiences a -0.21 to -0.23 point decline in Regents scores at the end of that school year, significant at the 5% level. Column 3, which controls for 12 as opposed to 6 other degree day bins (binning at the 5F as opposed to 10F level) also suggests that hot days have a negative impact on student achievement, but estimates are not statistically significant. All three columns suggest that the contemporaneous performance impact of temperature during the exam is robust to the inclusion of degree day controls.

Figure 10 presents the results graphically, plotting the coefficients for each degree day bin, this time omitting the 60-70F bin. While the estimates are noisy, the results suggest that years with more days on the hotter end of the degree day distribution are associated with depressed exam scores, controlling for school-specific factors and district-level trends.

Assuming that the incidence of hot days is not correlated with other determinants of test scores not captured in the model above, the point estimates suggest that a year with an additional extreme heat day (above 90F) reduces scores by approximately 1.2% of a standard deviation.

6 Discussion

The estimates in Tables 3 and 4 suggest that heat exposure during high stakes exams can reduce exam performance by 0.22 points per degree F, or approximately 0.10 standard deviations per standard deviation in daily temperature during the month in which exams are administered (June). These raw score effects are comparable in magnitude to the impacts of reducing teacher value-added by 1 standard deviation (Chetty, Friedman, Rockoff, 2014), or increasing class size from 25 to 31 students (Angrist and Lavy, 1999). For the average NYC public school student, taking an exam on a 15 degree hotter day can reduce the likelihood of passing a Regents exam by almost 14%. These performance effects are significantly larger for low-performing students, suggesting that the welfare burden of heat exposure may be regressive (Table 6). The estimates in Table 8 suggest that heat

exposure during the school year may reduce cumulative learning attainment, perhaps by reducing the amount or effectiveness of pedagogically engaged hours, though the results are less conclusive on the question of cumulative human capital impacts.

The estimates presented here may be biased for a number of reasons. They may be upward biased to the extent that temperature shocks are correlated with other adverse atmospheric conditions that are not controlled for explicitly, including air pollutants. On the other hand, they may be downward biased to the extent that temperature is measured imprecisely (attenuation bias due to classical measurement error), or to the extent that teachers may in some instances manipulate grades on the margin to allow students who perform poorly to scrape by above the passing threshold, as has been documented by Dobbie et al (2014).

It is important to bear in mind that the welfare impacts of the short run performance impact estimates depend crucially on contextual factors, including institutions. Random, short-run diminutions to test performance caused by weather fluctuations, while adding noise to the signal extraction process, likely do not reduce underlying human capital the way, for instance, having a good teacher during 6^{th} grade might increase human capital accumulation in the long run. There are, however, a few ways in which even such transient impacts of heat stress may have welfare-relevant consequences.

The effect of heat exposure on high stakes exams may lead to inefficient allocation of skills in the labor market, as suggested by Lavy and Ebenstein (2015). If high stakes exams are used as the primary method for ranking students and matching skill types to occupations, and temperature shocks introduce random noise into this measure of skill, aggregate welfare may be reduced due to inefficient allocation of labor.

It is also possible that students who receive lower scores than expected on high stakes exams may become discouraged and choose to invest less in subsequent human capital accumulation than would otherwise have been optimal given true underlying ability. This seems plausible given evidence suggesting that students are often unsure about their ability and are constantly updating their beliefs based on school performance (Zafar, 2013).

From a distributional equity standpoint, policymakers may worry about heat-related performance declines insofar as the realized impact falls disproportionately on low-income individuals. This may be due to heterogeneity in the sensitivity of low- vs high-ability individuals to the same heat shock, as documented here. It may also, at a societal level, be due to different rates of

temperature exposure along racial or socioeconomic lines, as has been documented in the case of many environmental pollutants (Currie, 2007; Currie et al, 2014; Chay and Greenstone, 2005). It is possible, for instance, that part of the persistent racial achievement gap in AP or SAT scores between Whites and Blacks or Whites and Hispanics may be due to differential likelihood of heat exposure during these exams, given nationally synchronized timing and the higher proportion of Blacks and Hispanics living in Southern States and in densely urbanized (and thus heat-island affected) neighborhoods.

Finally, if heat exposure has a cumulative impact on human capital attainment, as suggested by the analysis presented here, there may be important implications for climate policy. The link between temperature and human capital accumulation informs two outstanding puzzles in the social cost of carbon literature: first, whether human capital impacts, so far omitted entirely from integrated assessment models of climate damages, belong in the climate damage function (Tol, 2009); and second, whether the impacts of climate change on GDP will take the form of levels effects or growth rate effects (Dell, Jones, Olken, 2014; Pindyck, 2013).

These findings suggest that there may be growth rate effects from hot years, due to the negative impact of temperature stress on the rate of human capital accumulation. If human capital production involves significant dynamic complementarities (Cunha and Heckman, 2007), then the overall growth impacts might be magnified over time.

It is unclear how much adaptation can mitigate the impacts in the long run, especially given the potential for continued technological innovation. However, in the climate change context, these estimates are likely closer to the lower end of the spectrum. First, these estimates come from New York City, which features one of the highest per pupil educational expenditures in the world. Moreover, even though New Yorkers have lower air conditioning penetration and utilization rates than some warmer parts of the US, the average household AC penetration rate, at above 90%, is much higher than most other countries, including OECD countries. It is worth bearing in mind that, in many developing countries, school buildings are much more exposed to the elements, though it is possible that students are better acclimated to heat stress in tropical climates.

7 Conclusion

Periodic heat exposure in the classroom is a common phenomenon in many parts of the world, but the impact of temperature on human capital production process remains largely unexplored. The prospect of human-induced climate change makes understanding the role that temperature plays — and the possible adaptation strategies that may be adopted — an important policy priority.

The application of a research design based on 2.6 million high school exit exams over the period 1998-2008 yields three primary findings. First, on average, heat exposure exerts a causal and statistically meaningful impact on test performance, and possibly on cumulative learning attainment. Heat exposure on the day of the exam reduces performance, controlling for student ability, and hotter than average years tend to be associated with reduced learning, controlling for this short run exam-day impact.

Second, this study finds evidence of significant heterogeneity in heat sensitivity across students of higher and lower ability, suggesting that the same heat shock may cause greater disruption to disadvantaged students, either because they are less able to cope with distractions, or because they are less inclined to exert additional effort given greater uncertainty about the returns to investment in education.

Third, air conditioning seems to have a mild but incomplete mediating effect in the short run, though the imprecision of the results suggest that local factors (e.g. building age, equipment quality and utilization) may matter considerably.

The magnitude of the short run effects are large, comparable in magnitude to achievement gains associated with a one standard deviation improvement in teacher quality (teacher value added) or a reduction in class size from 31 to 25. For the average student, the difference between an average 75F day and a 90F day is approximately -0.20 standard deviations lower achievement, and a 12.3% higher chance of failing an exam. These impacts are substantially larger for lower performing students (measured using previous standardized exam scores), who suffer per degree impacts that are as much as five times as large as those for high performing students.

These results underscore opportunities for further research along several dimensions. From the perspective of environmental policy, this study provides clues regarding the mechanisms by which climate change may exert growth rate impacts — as some cross country estimates suggest (Dell, Jones,

Olken, 2012) — as opposed to impacts on the level of GDP. Measuring these impacts in developing country contexts and translating them into potential impacts on GDP, similarly to the existing literature linking health and human capital to growth rates (Hanushek and Woessman, 2013; Fogel, 2004; Weil, 2007), may provide meaningful contributions to our understanding of the true social cost of carbon.

This paper also raises broader questions around the determinants of educational attainment. At the national level, this study provides additional clues regarding the persistence of racial achievement gaps in education (Card and Rothstein, 2007). Given the well-documented correlation between income, race, and environmental amenities (Currie et al, 2015), Black and Hispanic students may be subjected to disproportionately high levels of heat stress during their educational careers, and more likely to suffer heat stress during timing synchronized high stakes exams such as the AP. It also seems possible that differential impacts of heat stress are driving part of summer learning loss, a phenomenon wherein low-income students' achievements decline over the summer faster than the achievement of high-income children, at least in some settings (Alexander et al, 2001; Fryer and Levitt, 2004).

It remains to be seen whether heat stress during schooling can have long run impacts on welfare, as suggested, for instance, by recent work documenting persistent wage effects of heat stress in utero (Isen et al, 2015).

References

Alexander, Karl L., Doris R. Entwisle, and Linda S. Olson. "Schools, achievement, and inequality: A seasonal perspective." Educational Evaluation and Policy Analysis 23.2 (2001): 171-191.

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. "Lead Exposure and Racial Disparities in Test Scores." Brown Univ. Work. Pap (2015).

Anthoff, David, Cameron Hepburn, and Richard SJ Tol. "Equity weighting and the marginal damage costs of climate change." Ecological Economics 68.3 (2009): 836-849.

Anderson, Craig A. "Temperature and aggression: effects on quarterly, yearly, and city rates of violent and nonviolent crime." Journal of personality and social psychology 52.6 (1987): 1161.

Angrist, Joshua D., and Victor Lavy. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." The Quarterly Journal of Economics 114.2 (1999): 533-575.

Cachon, Gerard, Santiago Gallino, and Marcelo Olivares. "Severe weather and automobile assembly productivity." Columbia Business School Research Paper 12/37 (2012).

Card, David, and Jesse Rothstein. "Racial segregation and the black?white test score gap." Journal of Public Economics 91.11 (2007): 2158-2184.

Chay, Kenneth Y., and Michael Greenstone. "Does Air Quality Matter? Evidence from the Housing Market." Journal of Political Economy 113.2 (2005).

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." THE AMERICAN ECONOMIC REVIEW 104.9 (2014): 2593-2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." THE AMERICAN ECONOMIC REVIEW 104.9 (2014): 2633-2679.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. No. w21156. National Bureau of Economic Research, 2015.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., & Rivkin, S. G. (2009). Does pollution increase school absences?. The Review of Economics and Statistics, 91(4), 682-694.

Currie, Janet. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." Journal of Economic Literature (2009): 87-122.

Deryugina, Tatyana, and Solomon M. Hsiang. "Does the environment still matter? Daily temperature and income in the United States." (2014).

Fitzpatrick, C., McKinnon, R. D., Blair, C. B., & Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children?. Learning and Instruction, 30, 25-31.

Fogel, R. W. (2004). Health, Nutrition, and Economic Growth*. Economic Development and Cultural Change, 52(3), 643-658.

Gaffin, S. R., Rosenzweig, C., Khanbilvardi, R., Parshall, L., Mahani, S., Glickman, H., . . . & Hillel, D. (2008). Variations in New York City's urban heat island strength over time and space. Theoretical and applied climatology, 94(1-2), 1-11.

Goldin, Claudia Dale, and Lawrence F. Katz. The race between education and technology. Harvard University Press, 2009.

Goodman, Joshua. Flaking out: Student absences and snow days as disruptions of instructional time. No. w20221. National Bureau of Economic Research, 2014.

Grether, W. (1973). "Human performance at elevated environmental temperatures." Aerospace Medicine 44(7): 747-755.

Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton. "Developing a social cost of carbon for US regulatory analysis: A methodology and interpretation." Review of Environmental Economics and Policy 7.1 (2013): 23-46.

Hanushek, Eric A., and Ludger Woessmann. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." Journal of Economic Growth 17.4 (2012): 267-321.

Hogrefe, C., Biswas, J., Lynn, B., Civerolo, K., Ku, J. Y., Rosenthal, J., & Kinney, P. L. (2004). Simulating regional-scale ozone climatology over the

eastern United States: model evaluation results. Atmospheric Environment, 38(17), 2627-2638.

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. "Quantifying the influence of climate on human conflict." Science 341.6151 (2013): 1235367.

Isen, A., Rossin-Slater, M., & Walker, R. (2015). Heat and long-run human capital formation. Unpublished Manuscript.

Matthew E. Kahn The Climate Change Adaptation Literature. Rev Environ Econ Policy 2016.

Lavy, Victor, Avraham Ebenstein, and Sefi Roth. The impact of short term exposure to ambient air pollution on cognitive performance and human capital formation. No. w20648. National Bureau of Economic Research, 2014.

Lavy, Victor, Avraham Ebenstein, and Sefi Roth. The Long Run Human Capital and Economic Consequences of High-Stakes Examinations. No. w20647. National Bureau of Economic Research, 2014.

Lavy, Victor. "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics." The American economic review 99.5 (2009): 1979-2011.

Mackworth, Norman H. "Effects of heat on wireless operators." British journal of industrial medicine 3.3 (1946): 143.

Maddala, Gangadharrao Soundaryarao, and Kajal Lahiri. Introduction to econometrics. Vol. 2. New York: Macmillan, 1992.

Melillo, Jerry M., Terese TC Richmond, and G. W. Yohe. "Climate change impacts in the United States." Third National Climate Assessment (2014).

Seppanen, O., et al. (2006). "Effect of temperature on task performance in office environment." Lawrence Berkeley National Laboratory.

Sievertsen, Hans Henrik, Francesca Gino, and Marco Piovesan. "Cognitive fatigue influences students' performance on standardized tests." Proceedings of the National Academy of Sciences (2016): 201516947.

Stock, James H., and Mark W. Watson. "Introduction to econometrics." (2007).

Sudarshan, Anant, and Meenu Tewari. The economic impacts of temperature on industrial productivity: Evidence from indian manufacturing. Working Paper, 2013.

Weil, David N. "Accounting for the Effect of Health on Economic Growth." The Quarterly Journal of Economics (2007): 1265-1306.

Zafar, Basit. 'How Do College Students Form Expectations?'. Journal of Labor Economics 29.2 (2011): 301?348

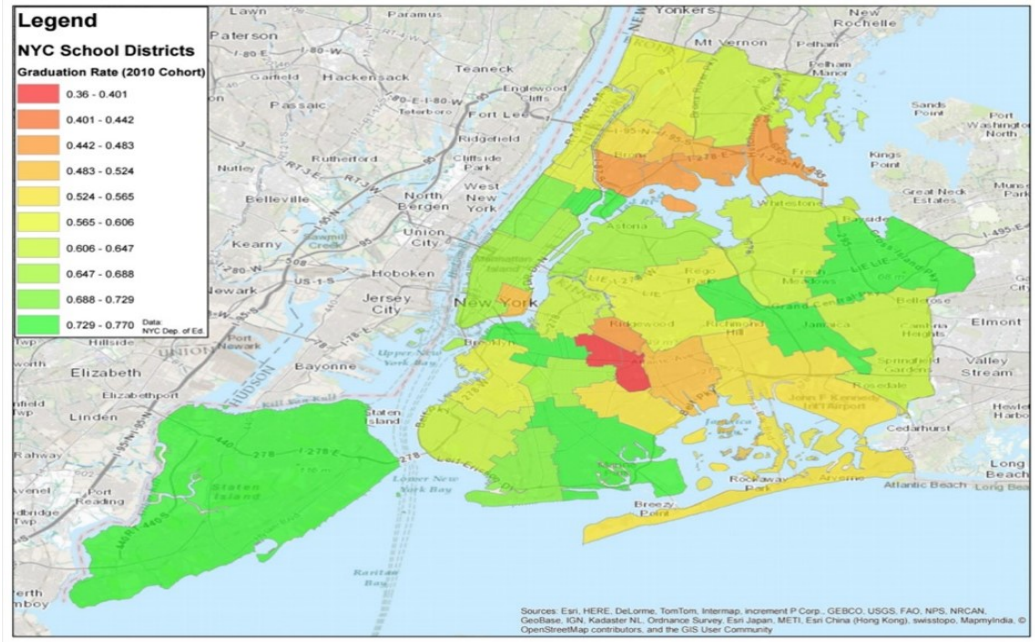
Zivin, Joshua Graff, and Matthew Neidell. "The Impact of Pollution on Worker Productivity." The American Economic Review 102.7 (2012): 3652-3673.

Zivin, Joshua Graff, and Matthew Neidell. "Temperature and the allocation of time: Implications for climate change." Journal of Labor Economics 32.1 (2014): 1-26.

Zivin, Joshua S. Graff, Solomon M. Hsiang, and Matthew J. Neidell. Temperature and Human Capital in the Short-and Long-run. No. w21157. National Bureau of Economic Research, 2015.

Figures and Tables

Figure 1



Notes: This figure presents average 4-year graduation rates by school district (2010) in New York City public high schools

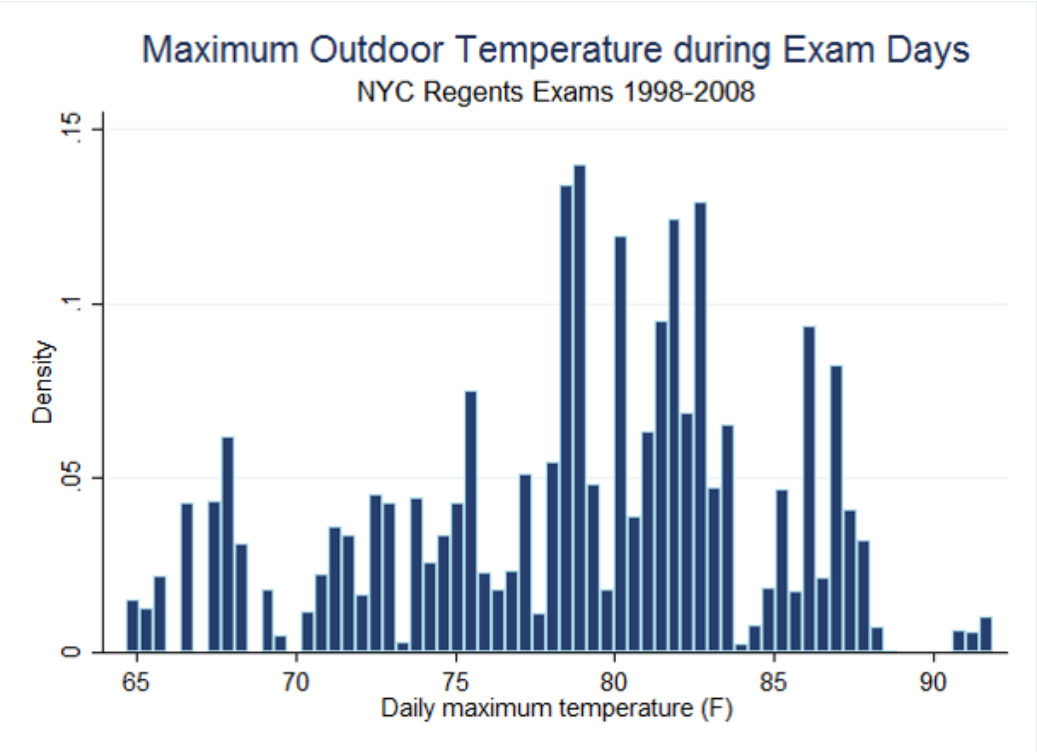
Table 1

Table 1: Characteristics of New York City Regents Exam Takers by Borough (1999-2009)

Variable	Borough					
	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average/ Total
Mean # of subjects taken per student	4.99	4.64	4.57	4.57	4.46	4.64
Mean scaled score (0-100)	60.81	64.76	66.38	65.70	67.42	64.71
Standard deviation	18.63	17.99	18.27	17.39	17.34	17.92
Absent on exam day	0.002	0.002	0.003	0.005	0.0003	0.002
Proportion with failing scores	0.52	0.42	0.40	0.40	0.36	0.42
Proportion below college proficiency	0.75	0.67	0.64	0.66	0.63	0.67
Number of students	95,964	163,060	111,420	146,285	34,880	551,609
Number of observations	479,318	761,593	504,307	658,173	155,471	2,558,862

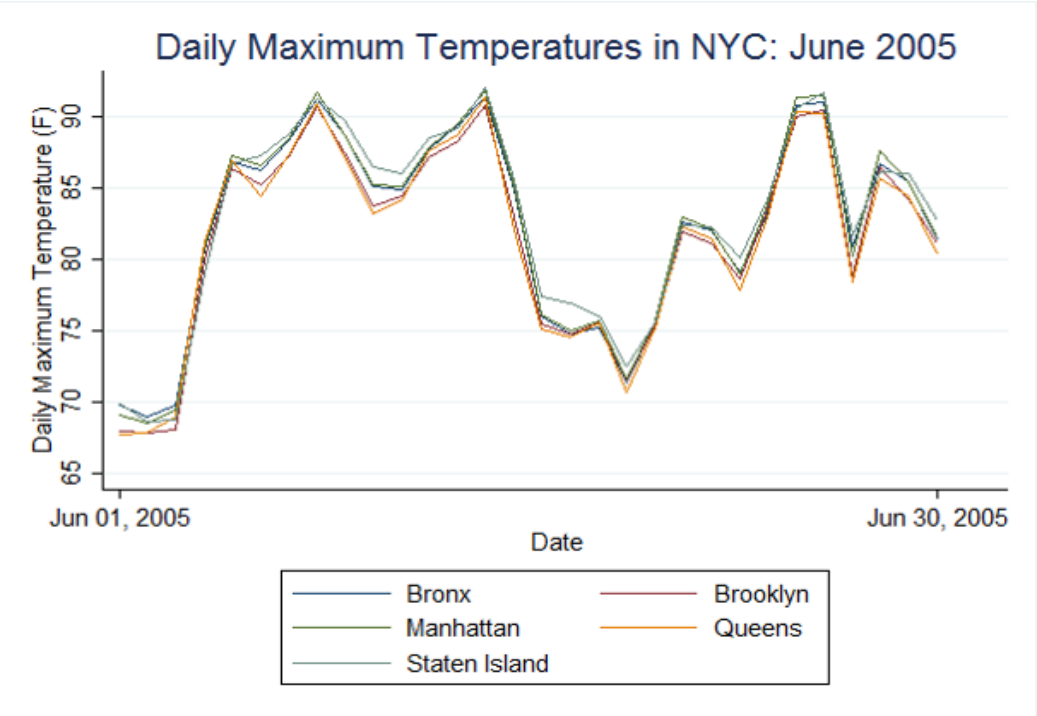
Notes: This table presents summary statistics for June Regents exams taken during the 1998-1999 and 2008-2009 school years, taken from NYC DOE administrative records. The sample includes all students taking core Regents subjects at the conclusion of the spring semester. Core subjects include: Math (Algebra, Geometry, Trigonometry), Science (Physics, Chemistry, Living Systems), English, US History & Government, and World History and Geography. Scores are recorded on a 0 — 100 scale by teachers in each school, with subject-year-specific scaling guidelines provided by NYSED. Failing thresholds are 65 out of 100 for all subjects included here. Scores above 75 are deemed 'College proficient'.

Figure 2



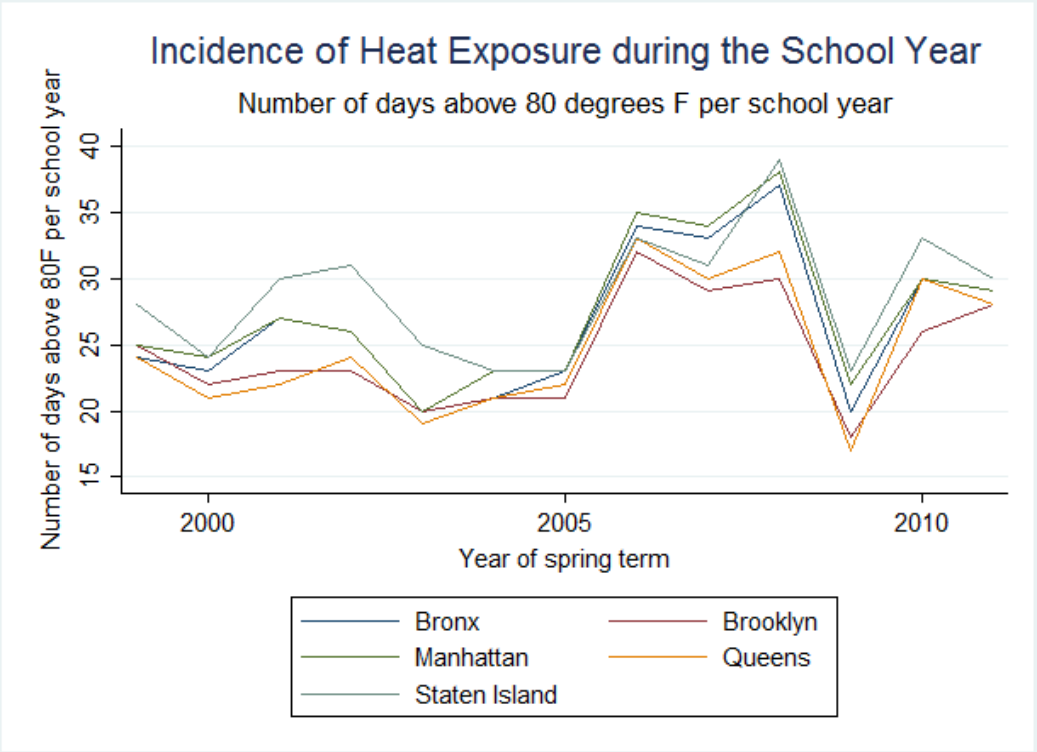
Notes: This figure presents a histogram of daily maximum temperatures on NYC Regents Exam dates during the primary June exam waves (1998-2009)

Figure 3



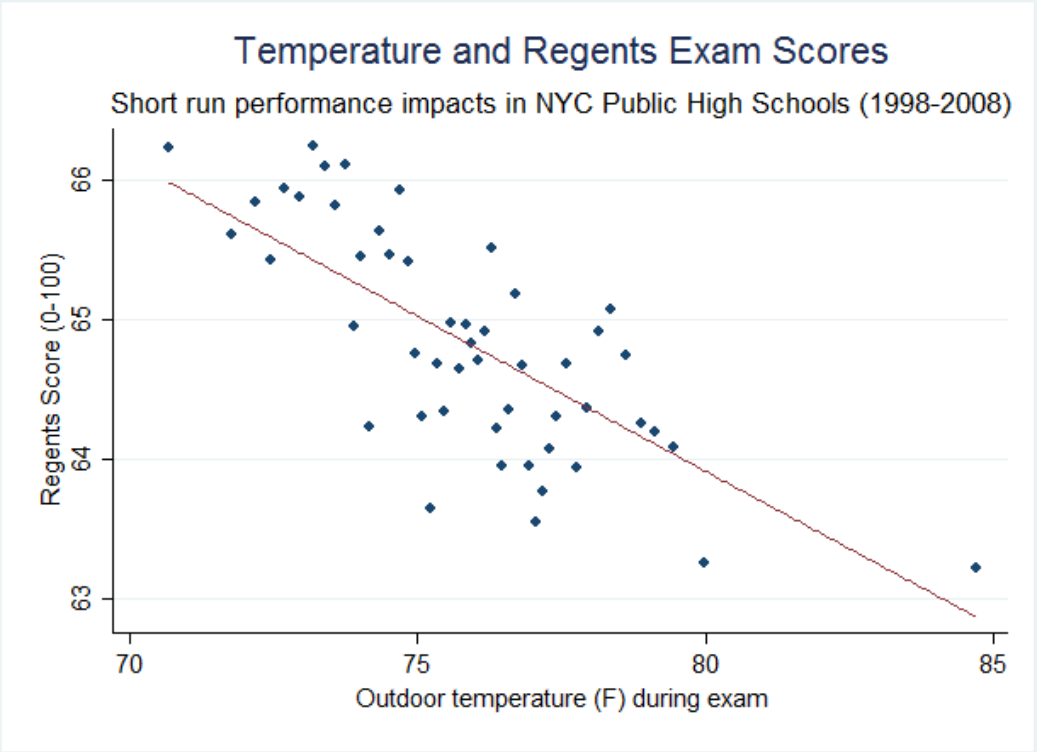
Source: This figure presents the source of identifying variation for contemporaneous performance impacts. Daily maximum temperatures by Borough of NYC during a typical June (2005).

Figure 4



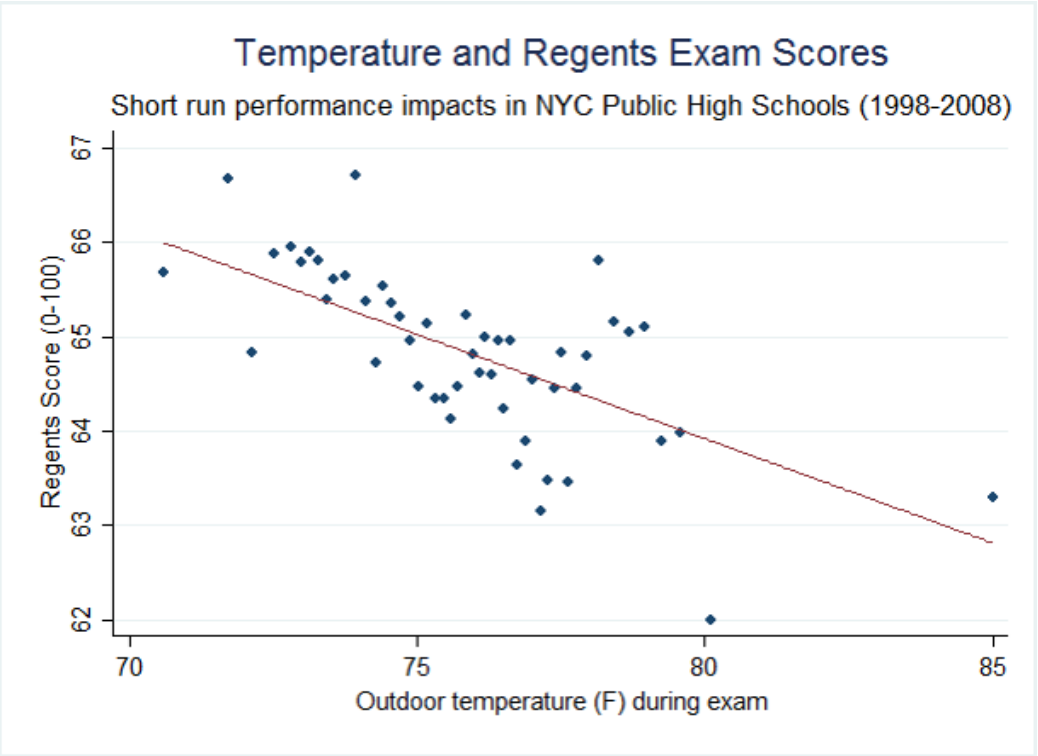
Source: This figure presents the source of identifying variation for contemporaneous performance impacts. Daily maximum temperatures by Borough of NYC during a typical June (2005).

Figure 5



Source: This figure presents the source of identifying variation for contemporaneous performance impacts. Daily maximum temperatures by Borough of NYC during a typical June (2005).

Figure 6



Notes: This figure presents a binned scatterplot of scaled exam scores (0-100) and test-day temperature (F). Each observation is a quantile of residual temperature, approximately 50,000 exam records. Includes student, year, and subject fixed effects, as well as controls for precipitation, dewpoint, and time of day of exam.

Table 2

Table 2: Contemporaneous Performance Impacts of Temperature Stress
NYC Regents Exams 1998-2008
Dependent Variable is Regents Score (0-100)
MAIN SPECIFICATIONS

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Temperature (F)	-0.173***	-0.172***	-0.154***	-0.221***	-0.220***	-0.203***
	0.019	0.019	0.018	0.022	0.022	0.021
Precipitation	0.015	0.012	0.015	0.021	0.019	-0.005
	0.014	0.013	0.015	0.013	0.013	0.033
AM/PM correction				X	X	X
Subject FE	X	X	X	X	X	X
School FE	X	X		X	X	
Year FE	X	X	X	X	X	X
School X Year FE	X			X		
School-specific trends		X			X	
Student FE			X			X
N	2,539,540	2,539,540	2,539,540	2,539,540	2,539,540	2,539,540
r-squared	0.265	0.251	0.671	0.266	0.258	0.671

Notes: June Regents exams only. Subjects include Math A, Math B, Biology, Physics, Chemistry, English, World History, US History, and US Government. Robust standard errors clustered at school by year level. School-by-year and student-fixed effects specifications include 11,930 and 730,170 fixed effects respectively.

Table 3

Table 3: Contemporaneous Performance Impacts of Temperature Stress
NYC Regents Exams 1998-2008
Dependent Variable is Regents Score (0-100)
ROBUSTNESS OF STUDENT FIXED EFFECTS SPECIFICATION

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Temperature (°F)	-0.137***	-0.301***	-0.155***	-0.149***	-0.366***	-0.203***
	0.023	0.018	0.017	0.025	0.021	0.020
Precipitation	0.044**	0.032**	0.021	0.052***	0.038***	0.038
	0.014	0.012	0.014	0.014	0.012	0.011
AM/PM correction				X	X	X
Subject FE			X			X
Year FE		X	X		X	X
Student FE	X	X	X	X	X	X
N	2,539,540	2,539,540	2,539,540	2,539,540	2,539,540	2,539,540
r-squared	0.713	0.719	0.756	0.713	0.718	0.756

Notes: June Regents exams only. Subjects include Math A, Math B, Biology, Physics, Chemistry, English, World History, US History, and US Government. Robust standard errors clustered at school by year level. Student-fixed effects include 730,170 fixed effects.

Table 4

Table 4: Contemporaneous Performance Impacts of Temperature Stress
NYC Regents Exams 1998-2008
Dependent Variable is Pass Dummy (1=Pass)
MAIN SPECIFICATIONS

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Temperature (F)	-0.00548***	-0.00543***	-0.00401***	-0.00438***	-0.00434***	-0.00360***
	0.00057	0.00056	0.00043	0.00051	0.00049	0.00042
Precipitation	0.00067	0.00069	0.00108	0.00050	0.00052	0.00099
	0.00028	0.00027	0.00031	0.00028	0.00027	0.00031
AM/PM controls				X	X	X
Subject FE	X	X	X	X	X	X
School FE	X	X		X	X	
Year FE	X		X	X		X
School X Year FE	X			X		
School-specific trends		X			X	
Student FE			X			X
N	2,562,910	2,562,910	2,562,910	2,562,910	2,562,910	2,562,910
r2	0.176	0.170	0.588	0.176	0.170	0.588

Notes: June Regents exams only. Subjects include Math A, Math B, Biology, Physics, Chemistry, English, World History, US History, and US Government. Robust standard errors clustered at school by year level. Student-fixed effects include 730,170 fixed effects.

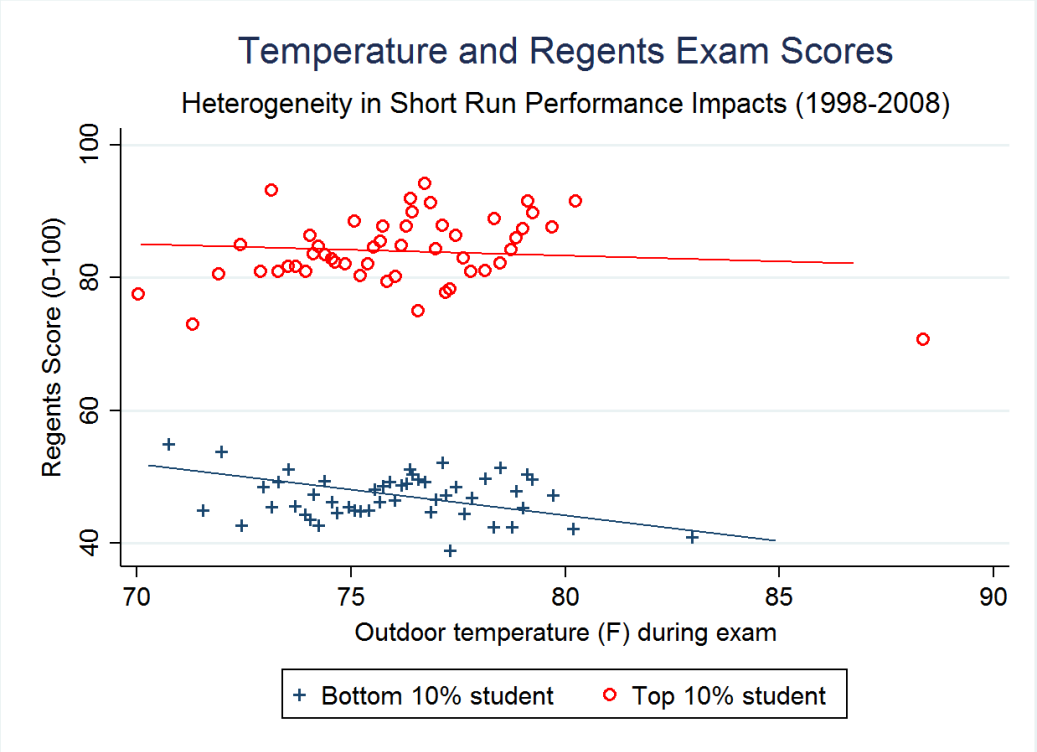
Table 5

Table 5: Average Regents Exam Performance by Quantile of Elementary-and-Middle School Achievement Distribution
NYC Public School Students (1998-2008)

	Average combined z-score (ELA/Math 3rd-8th grade)	Average Regents Score (0-100)	Average Pass Rate (0-1)
Bottom 10%	-2.49	44.64	0.13
	(0.93)	(16.14)	(0.34)
Top 10%	3.17	85.53	0.97
	(0.73)	(9.89)	(0.16)
Average	0.38	64.74	0.57
	(1.62)	(18.09)	(0.49)

Notes: Students are ranked on the basis of standardized ELA and Math z-scores during elementary and middle school years (3^{th} — 8^{th} grade). Each student is assigned a grade-by-year specific z-score for each of their ELA and Math exams. The 'Average combined z-score' denotes the average z-score of a given student, averaged over all available years and grades (median=4).

Figure 7



Notes: This is a binned scatterplot of Regents scores on test time temperature for top and bottom deciles of the student ability distribution. Fitted regression lines are drawn from running a model with school, subject, year, and school-by-year fixed effects. Students are ranked on the basis of standardized ELA and Math z-scores during elementary and middle school years (3^{th} — 8^{th} grade). Each student is assigned a grade-by-year specific z-score for each of their ELA and Math exams. The 'Average combined z-score' denotes the average z-score of a given student, averaged over all available years and grades (median=4).

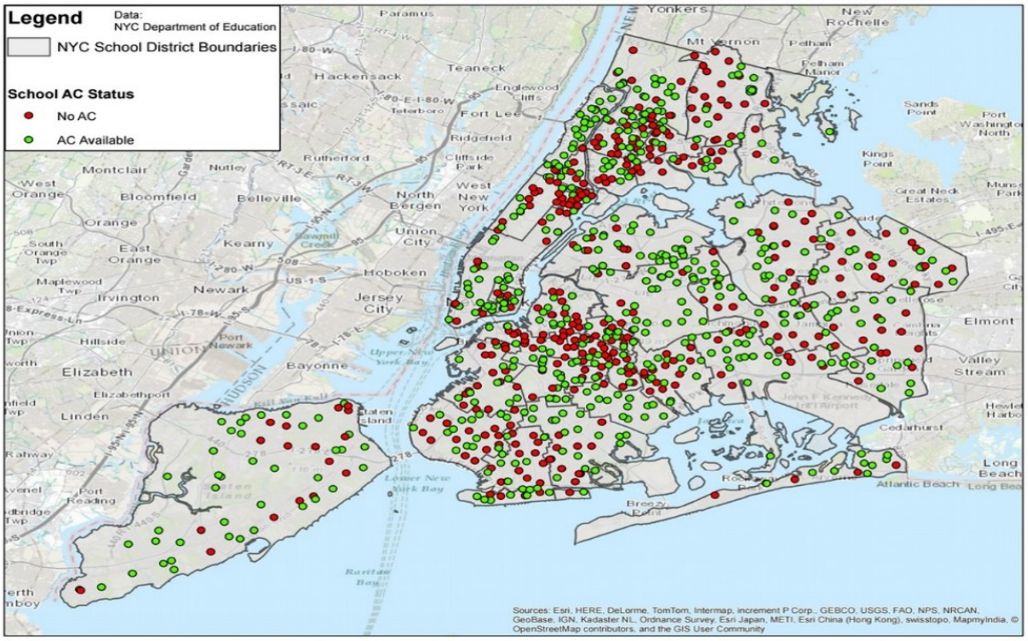
Table 6

Table 6: Building-Level Air Conditioning Installation Information
New York City Public Schools (2012)

Borough	Schools without (non-defective) AC	Schools with (non-defective) AC	Total
Bronx	163	66	229
Brooklyn	256	107	363
Manhattan	128	64	192
Queens	167	114	281
Staten Island	43	25	68
Total	757	376	1,133
Total Percentages	66.8%	33.2%	100%

Notes: This table presents summary statistics for air conditioning installation information for NYC public schools. Building condition assessments are reported by school building, and many NYC public schools share buildings with one or more other schools. The 1,133 buildings for which data were available cover approximately 94% of the school buildings in New York City as of 2012. Source: NYC School Construction Authority Building Condition Assessment Reports (BCAS, 2012)

Figure 8



Notes: Map of NYC Public Schools by air conditioning installation status as of 2012.

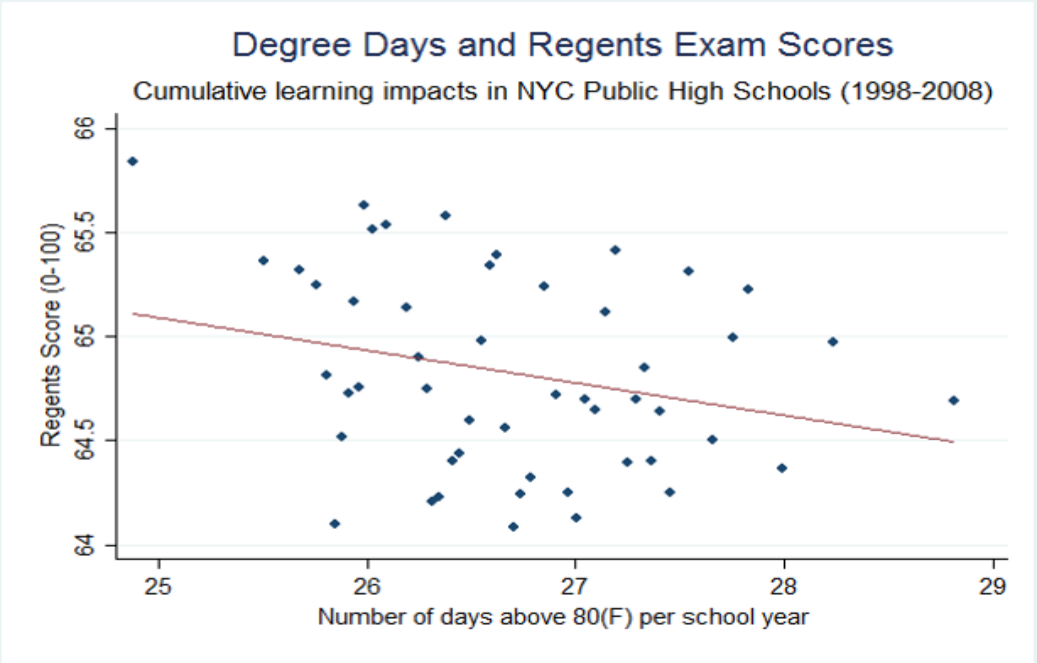
Source: NYC School Construction Authority Building Condition Assessment Reports (2012)

Table 7

Table 7: Effect Heterogeneity by Student Performance and School AC Status
Dependent Variable is Regents Exam Score (0-100)
Preferred School Fixed Effects Specification (Equation 1)

Outdoor Temperature during exam (°F) interacted with:	(1)	(2)	(3)	(4)
1	-0.228***	-0.211***	-0.254***	-0.195***
	0.021	0.021	0.021	0.020
Top 10% student & No AC school	-0.099***			
	0.021			
Top 10% student & AC school	-0.035***			
	0.021			
Bottom 10% student & No AC school		-0.457***		
		0.021		
Bottom 10% student & AC school		-0.438***		
		0.021		
Top Tercile student & No AC school			-0.087***	
			0.021	
Top Tercile student & AC school			-0.057***	
			0.021	
Bottom Tercile student & No AC school				-0.376***
				0.020
Bottom Tercile student & AC school				-0.366***
				0.021
Precipitation (inches)	0.021	0.023	0.026	0.026*
	0.013	0.013	0.013	0.012
Dewpoint	0.284	0.272	0.310	0.266
	0.043	0.043	0.042	0.042
AM/PM controls	X	X	X	X
Subject FE	X	X	X	X
School FE	X	X	X	X
Year FE	X	X	X	X
School X Year FE	X	X	X	X
N	2,539,540	2,539,540	2,539,540	2,539,540
R-squared	0.271	0.298	0.309	0.313

Figure 9



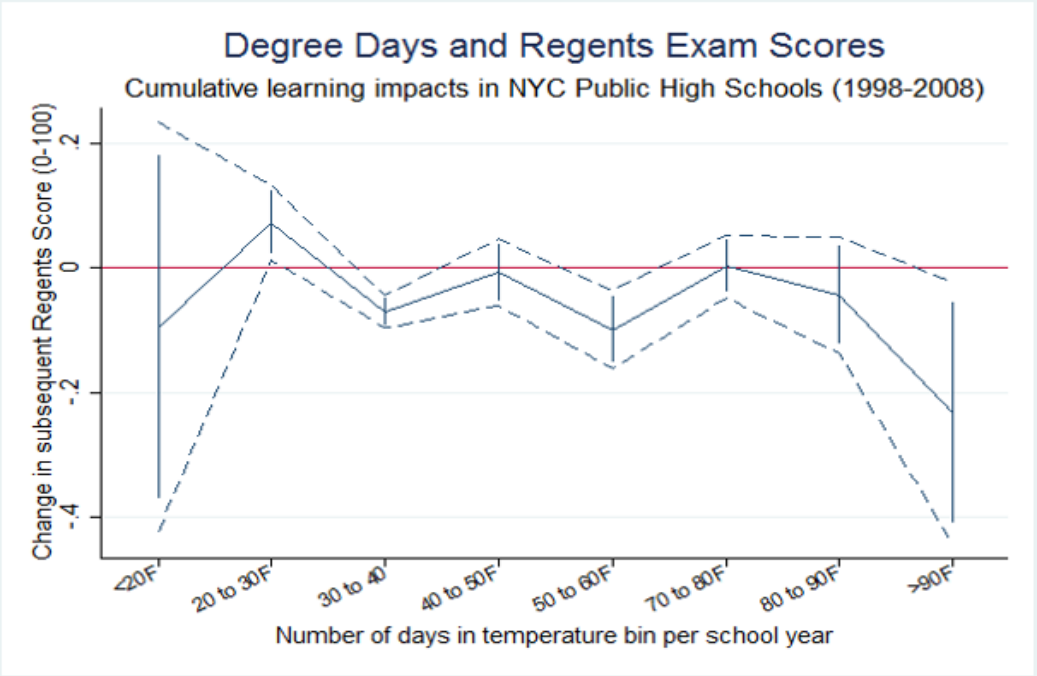
Notes: Binned scatterplot of residual exam score and number of hot days (80F and above) per school year. Includes school, subject fixed effects, linear time trends, controls for all other temperature bins, test day weather, annual precipitation, snowfall, and time of day, omitting 60-65F bin.

Table 8

Table 8: Long Run Cumulative Learning Impacts of Heat Exposure during school year
Dependent variable is Regents Score (0-100)
NYC Public High Schools (1998-2008)

Variable	(1)	(2)	(3)
Outdoor Temperature during exam (°F)	-0.197***	-0.194***	-0.207***
	0.022	0.022	0.022
Days during school year with max temperature of…			
80 to 90°F	-0.044	-0.024	-0.128
	0.048	0.043	0.058
90°F or above	-0.233**	-0.217**	-0.160
	0.109	0.097	0.151
15°F or below	-0.096	0.033	0.227
	0.168	0.169	0.217
20 to 30°F	0.073	0.056	
	0.032	0.020	
30 to 40°F	-0.071***	-0.060***	
	0.013	0.013	
40 to 50°F	-0.007	0.009	
	0.027	0.025	
50 to 60°F	-0.099	-0.070	
	0.033	0.028	
70 to 80°F	0.002	0.005	
	0.026	0.025	
Precipitation	-0.007	-0.010	0.001
	0.013	0.013	0.013
Time	0.264	0.247	0.368**
	0.154	0.154	0.156
Annual Snowfall	0.056***	0.061***	0.020
	0.015	0.011	0.020
Total precipitation during school year (inches)	-1.922***	-1.653***	-0.163
	0.449	0.421	0.558
N	2,539,540	2,539,540	2,539,540
R-squared	0.252	0.252	0.252
School FE	X	X	X
Subject FE	X	X	X
Global Quadratic Trend	X	X	X
District-Specific Trends	X		
Degree bin increments	10°F	10°F	5°F

Figure 10



Notes: Plot of coefficients corresponding to the impact of an additional day in any given degree day bins (x axis) on subsequent Regents exam scores (y axis). Plotted coefficients are obtained by running a model that includes school, subject fixed effects, linear time trends, test day weather, annual precipitation, snowfall, time of day, and controls for all other temperature bins, omitting the 60-70F bin. The dashed line indicates 99% confidence interval; solid line, the 95% confidence interval. Standard errors are clustered at school-by-year level.

A Data Linkages

Since we will be measuring the impact of temperature stress on exam performance we need to link exam scores to exam dates and ambient temperatures. To assess the possible mitigating impact of air conditioning, we need to link students and schools to BCAS reports. The latter step is straightforward, as the NYC DOE data and BCAS records both include multiple school identifiers (a district-borough-school number code, as well as school names and street addresses). The first step, of matching dates to exam records, was complicated by inconsistent record-keeping across schools and years. Exam subjects and terms (June, August, January) are not consistently recorded in all years. Specifically, there is evidence for inconsistency in reporting for some subjects and schools after 2009, which is when several subjects were discontinued. As such, I use the most conservative matching mechanism possible, by assigning test dates and temperature readings only to scores for which exam subject, exam year, and exam term are internally consistent. This results in 2.57 million matched June exams. Because all NYC public school students take any given Regents subject on the same date and time, it is possible to assign date and time information consistently across schools in the sample, so long as information on exam subject, year, and term are accurately provided. For instance, in June 2008, Regents exams in 9 subjects were taken over the course of six days between June 17 and June 25[2], beginning with 'Integrated Algebra' at 9:15am and 'Physical Setting: Chemistry' at 1:15pm on June 17 and ending with 'Physical Setting: Physics' at 9:15am on June 25. So an observation that has a subject code corresponding to Physics, term code corresponding to June, and year of 2008 can be assigned temperature corresponding to 9am on June 25, 2008. Through this procedure, I am able to assign daily weather information for 2.57 million student-exam observations over 95 different test sessions.