Abstract

This paper explores the relationship between temperature and the human capital production process. Using student-level administrative data for a major US public school district and quasi-experimental variation in weather, I estimate the causal impact of hot temperature on educational outcomes. Heat exposure during a high stakes exam reduces performance by up to 15%, controlling for individual ability and accounting for avoidance behaviors. This leads to persistent impacts on educational attainment (e.g. high school graduation). The evidence also indicates that cumulative heat exposure may affect the rate of learning, and highlights the role of institutions in determining optimal adaptive responses to climate change.
1 Introduction

It is well-known that hotter countries tend to be poorer: a country that is 1°C warmer on average has roughly 8% lower GDP per capita. It is less well-known that hotter places tend to have substantially lower educational attainment, both across and within countries. Figure 1 graphs standardized math scores against average annual temperature for the 60 countries who participated in the 2012 Program for International Assessment (PISA). A strong negative association is apparent, even when controlling for per capita income. What role – if any – does climate play in explaining this relationship? The central objective of this paper is to shed light on the ways temperature may affect the human capital production process.

While the comparative economics literature has emphasized the role of institutions in the correlation between geography and living standards, recent empirical work highlights the causal impact of hot weather on contemporaneous economic activity such as health or labor productivity. A growing body of evidence also underscores the centrality of human capital accumulation in determining long-run economic mobility, whether in the context of macroeconomic growth or individual income. And yet, few studies have explored the relationship between climatic factors and the human capital production process, particularly in environments where adaptive responses can be taken into account.

Whether or not temperature meaningfully affects educational outcomes may be important for at least two reasons. First, if hot temperature adversely affects the rate of human capital accumulation, then this link may contribute to cross-national living standards comparisons given the centrality of human capital to long-run growth. Second, a temperature-human capital link may be important in estimating the magnitude of the carbon externality and the design of associated economic policies. Current social cost of carbon estimates do not reflect potential human capital impacts from climate change, and there is considerable disagreement over the scope of climate adaptation, particularly regarding the role of institutional factors in realizing (or hindering) efficient adaptation responses.

Assessing the relationship between climate and human capital presents formidable challenges. First, unobserved factors that give rise to the first correlation noted above (between temperature and income) may also affect the second one (between temperature and educe-
Given the well-documented relationship between average climate and institutional quality, cross-sectional analyses of temperature and human capital are likely to suffer from substantial omitted variable bias. Second, it is important to take adaptive responses and avoidance behaviors into account when assessing the welfare impact of environmental shocks, especially when thinking about potential long-run mechanisms linking climate to welfare (DJO, 2014; Graff Zivin and Neidell, 2014). This is difficult in the lab or in voluntary, low-stakes settings, where the proper incentives may not be present. Evidence suggests that the realized impact of weather shocks may depend on the quality of institutions (Kahn, 2005; Dell et al., 2012), the adoption of technologies such as air conditioning (Barreca et al., 2016; Park, 2016), and even sector-specific policies such as crop insurance (Annan and Schlenker, 2015). All of these factors may themselves be endogenous to climate, and may also drive a wedge between realized adaptation and the hypothetical efficient adaptation frontier.

In light of these challenges, this study uses school- and student-level administrative data from New York City which, in conjunction with a quasi-experimental empirical strategy and unique institutional features, allows for identification of the causal impact of hot temperature on educational outcomes net of adaptive responses (given current technologies). The research design is based on a simple premise: that short-run (day-to-day, year-to-year) variations in temperature are not caused by unobserved determinants of educational performance.

I focus on four empirical research questions. First, does acute heat stress meaningfully affect contemporaneous student performance? That is, do the findings from survey and laboratory contexts – wherein cognitive performance declines with elevated temperatures – extend to actual school settings, where the economic stakes are higher? Second, can short-run heat exposures, which presumably do not reduce the stock of human capital per se, nevertheless affect longer-term outcomes? Depending on the degree of institutional flexibility, the costs of retaking exams, or the presence of dynamic complementarities in the human capital production process, short instances of heat stress may have lasting economic consequences.

Third, is it possible for cumulative heat exposure to influence the efficacy of learning over time, as suggested by recent work (Cho, 2017), reducing the rate of human capital accumulation? Fourth, how do we expect relevant economic agents to adaptively respond, and how might such responses interact with existing institutional environments?

Using data from NYC public schools offers several distinct advantages. Having mul-

---

4 I direct the reader to a rich human capital literature on the many challenges and opportunities of using student outcomes as measures of human capital, including Hanushek and Woessmann, 2012.

5 This has been found to be the case in the context of air pollution (Ebenstein et al., 2016) and grade manipulation (Dee et al., 2016, Diamond and Persson, 2016). For instance, Ebenstein, Lavy and Roth (2016, henceforth ELR) find that air pollution exposure during high-stakes exams leads to lower post-secondary schooling attainment and reduced earnings.
tiple observations for each student within a given testing window allows robust causal inference, including models with student fixed effects. Because all students are assigned to test dates and locations without prior knowledge of temperature (and without the ability to reschedule), temperature on the day of an exam is unlikely to be correlated with student quality. The fact that NYC public schools represents the most resource-intensive public school district in one of the world’s richest nations suggests that observed gaps between potential and realized adaptive investment will likely not be due (entirely) to income constraints or lack of access to technology, which may be the case in many developing countries [Gertler et al., 2016]. Finally, the richness of the data set – which comprises 4.5 million individual exam observations from roughly 1,000 high schools – permits an illustration of the subtle ways in which adaptive responses to environmental shocks may depend on institutional quality.

Focusing on NYC offers one notable disadvantage: namely that of out-of-sample prediction. Whereas the United States is one of the richest, most highly air conditioned countries in the world, most of the countries for whom a climate-human capital link may exert first-order welfare consequences are poor and have very little AC. Moreover, New York’s climate, while subject to periodic hot weather (e.g. several days above 90°F per year), is not regularly subject to the sort of extreme heat that much of the developing world experiences (e.g. many days above 100°F or 110°F). The results from this study are thus presented in the spirit of illustrating the potential for a new and important climate-development link, and may be interpreted as establishing a likely lower bound for welfare impacts in hotter, poorer settings.

The first main result is that heat exposure during a high-stakes exam exerts a causal and economically meaningful impact on student performance, even when controlling for individual ability. Taking an exam on a hot day leads to -0.22% lower performance per °F above room temperature (72°F). Put another way, a 90°F day reduces exam performance by 15 percent of a standard deviation relative to a more optimal 72°F day. For a sense of the magnitude: the within-school Black-White achievement gap is approximately 25% of a standard deviation. At least 18% of the students in my sample experience a hot exam with outdoor temperatures exceeding 90°F.

Looking at longer-run outcomes, I find that acute heat exposure during exams subsequently affects a student’s chances of graduating from high school. For the median student, taking an exam on a 90°F day leads to a 10.9% lower likelihood of passing any given subject. This means that a one standard deviation increase in average exam-time temper-

---

6Because the exams are effectively mandatory, selection bias is further minimized.

7It is possible that, due to greater physiological acclimatization, responses in poorer countries are actually smaller, though there is relatively little evidence on the relative protective effect of physiological versus technological adaptation.

8Heat exposure also substantially reduces students’ chances of achieving key performance thresholds
ature over the student’s high school career reduces the likelihood of graduating on time by roughly 2.5 percentage points, despite the fact that students are able to retake failed exams. These results are consistent with a world in which acute heat exposure nudges some students to achieve less schooling overall due to institutional rigidities and opportunity costs of time similar to those documented by Dee et al (2016) and ELR (2016), or dynamic complementarities in the human capital investment process (Cunha and Heckman, 2007 and Diamond and Persson, 2016).

I provide further but somewhat more speculative evidence on the relationship between cumulative heat exposure and human capital accumulation. Leveraging year-to-year variation in the number of hot days during the school year, I find that repeated heat exposure may reduce the rate of learning as evidenced by end-of-year exam performance – in addition to and controlling for the short-run impact documented above. A one standard deviation increase in the number of days above 80°F reduces Regents performance by approximately 4% of a standard deviation. Though this estimate is relatively under-powered, sign is consistent with emerging findings [Cho, 2017, Garg et al., 2017], and the coefficients are similar in magnitude to eliminating the gains associated with having a teacher with half a standard deviation higher value-added for one grade – an intervention which has been shown to increase cumulative lifetime incomes of NYC students by approximately $14,800 per student, or $445,000 per classroom [Chetty et al., 2014].

Two pieces of evidence highlight the importance of institutions in determining the efficacy of adaptive responses to environmental shocks. First, building-level AC installation data suggests that public school air conditioning may be sub-optimally provided, due perhaps to principal-agent problems or other market failures in educational settings. Fewer than 62% of NYC public schools had any form of air conditioning as of 2012, compared to residential and commercial AC penetration rates of 90% or higher. Comparing schools that do and do not have some form of AC, I find limited evidence for protective effects of having AC equipment present. One possibility is that, in the presence of informational asymmetries in education production, thermal comfort is perceived by administrators to be a consumption amenity as opposed to a production input, and thus inefficiently utilized from the social planner’s perspective.

That are used by local universities in college admissions decisions.

9While 90°F days are not infrequent during June exams, they are far less frequent during the school year, motivating the choice of independent variable to measure cumulative impacts. Prior evidence in the context of labor productivity suggests adverse impacts begin to occur around 80°F.

10There are many reasons why the later-life impacts of better teaching may be different from those associated with fewer heat-related disruptions, as discussed in the appendix.

11Of the schools that do have AC, over 40% were deemed to have defective components by independent building inspectors.

12It is also consistent with findings suggesting that partial air conditioning retrofits in old buildings can in some cases do more harm than good due to reduced air quality and increased noise [Niu, 2004], and may also be due to data constraints since AC installation status may be a noisy predictor of actual AC
Second, and possibly in response to inadequate air conditioning, teachers appear to have selectively manipulated grades upward when students experienced hot exam sittings. Previous studies have used bunching estimators to document grade manipulation by teachers [Dee et al., 2016]. Following this approach, I estimate the relationship between grade manipulation and exogenous variation in exam-time temperature using a school, subject, and date-specific bunching estimator at passing cutoffs. I find that, while on average 6% of pre-reform Regents exams exhibit upward grade manipulation, the extent of manipulation varied systematically according to the temperature students experienced during the exam, and may have mitigated some of the adverse impacts on long-run educational outcomes. Such “adaptive grading” represents a hitherto undocumented and likely sub-optimal channel of climate adaptation.\footnote{A possible unintended consequence of eliminating teacher discretion in New York City public schools in 2011 may have been to expose more low-performing students to climate-related human capital impacts, eliminating a protection that applied predominantly to low-achieving Black and Hispanic students.}

This paper is broadly related to a growing literature exploring the causal impact of climate on economic outcomes, including impacts of temperature shocks on human health [Barreca et al., 2016], labor supply [Graff Zivin and Neidell, 2014], violent crime [Kenrick and MacFarlane, 1986; Hsiang et al., 2013] and local economic output [Dell et al., 2012; Heal and Park, 2013; Deryugina and Hsiang, 2014], as well as the literature on climate adaptation [Deschénes and Greenstone, 2011; Burke et al., 2015]. In particular, this paper shares with a smaller set of papers – including Graff-Zivin, Neidell, and Hsiang (henceforth GNH, 2015), ELR (2016), Garg et al (2017) and Cho (2017) – an emphasis on understanding the mechanisms through which environmental shocks such as temperature may affect educational outcomes. In comparison to the existing literature, this paper establishes a causal connection between hot temperature and both short- and long-run educational outcomes in high-stakes environments, and is able to shed light on important interactions between institutional settings and adaptive responses. The findings suggest that classroom temperature may be an important factor for policymakers to consider when allocating public resources, especially in contexts where heat exposure is frequent, high-stakes exams pose hurdles to further schooling, and where market failures in infrastructure investment are likely to be present.

The paper also contributes to a long human capital literature documenting the efficacy and welfare implications of various inputs to schooling, including teacher value added [Chetty et al., 2014], reductions in class size [Chetty et al., 2011; Angrist and Lavy, 1999], and school choice and desegregation [Deming et al., 2014].

Finally, it provides empirical support for the view that climate and human capital may interact in a way that contributes to the long-debated relationship between hotter climates and slower growth, though more careful research is needed to verify whether repeated heat utilization at the classroom level.
exposure from future climate change may reduce the rate of human capital accumulation in the long run.\textsuperscript{14}

To the extent that future climate change may result in a disproportionate increase in realized heat exposure for the poor within and across countries, these findings lend further support to the notion that climate change may have distributively regressive impacts. It is worth noting that while the average New Yorker is exposed to approximately 14 days above 90°F per year, the average resident of New Delhi experiences over 80 such days annually, with climate forecasts suggesting up to 190 such days per year in New Delhi by 2100.

The rest of this paper is organized as follows. Section I provides a brief overview of the relevant literature on heat and human welfare and describes the underlying conceptual framework. Section II describes the data and institutional context and presents key summary statistics. Section III presents the main results and various robustness checks. Section IV discusses implications and concludes.

2 The Economics of Heat Exposure

Three stylized facts from the existing scientific literature are of particular relevance in thinking about the impact of temperature on human capital production. First: heat stress directly affects physiology in ways that can be detrimental to cognitive performance. Second: most individuals demonstrate a revealed preference for mild temperatures close to room temperature, commonly taken to be between 65°F and 74°F, or 18°C and 23°C. Third: the inverted U-shaped relationship between temperature and performance documented in the lab has been confirmed in a range of welfare-relevant contexts including health and labor, but evidence of impacts on human capital remains thin, particularly in high-stakes school environments.\textsuperscript{15}

2.1 Heat Exposure and Human Capital

In seminal work, GHN (2015) provide evidence of the adverse impacts of temperature stress on cognitive performance in US households. They use NLSY survey data which includes short, voluntary assessments that were administered to several thousand students at home, and find evidence of contemporaneous impacts of hot days on math performance but not verbal performance, and little evidence of cumulative or persistent impacts over time.

\textsuperscript{14}This is especially true given competing mechanisms in the existing literature, and the potential for endogenous innovation in adaptation technologies.

\textsuperscript{15}Seminal papers on the physiology of heat exposure [Mackworth, 1946] Seppanen et al., 2006], hedonic analyses of a revealed preference for mild temperatures [Roback, 1982] Sinha et al., 2015], and impacts of temperature on labor and task productivity [Grether, 1973] Graff Zivin and Neidell, 2013] Sudarshan and Tewari, 2013] and how they relate to this paper are discussed in detail in the online appendix.
Empirical evidence from school settings – where students spend the majority of pedagogically engaged hours and where potentially welfare-enhancing public policy interventions might occur most directly – is limited mostly to low stakes cognitive assessments. Schoer and Shaffran (1973) compare the performance of students in a pair of classrooms set up as a temporary laboratory, with one classroom cooled and one not. They find higher performance in cooled environments relative to hot ones, but do not test for persistent impacts or avoidance behaviors. Garg et al (2017) study the impact of hot days during the year on Indian cognitive assessments and find that years with more days above 29°C reduce subsequent cognitive performance. Noting larger impacts for hot days during the growing season, they suggest that this is mostly due to reduced agricultural yield.

In contrast, Cho (2017) uses variation in summer-time temperature to predict November college entrance exam performance by Korean high school students, and finds evidence of cumulative heat exposure reducing subsequent performance. While these exams are very high stakes, they occur in November, meaning that Cho (2017) is unable to jointly examine the impact of hot temperature on contemporaneous exam performance and longer-run outcomes. Moreover, the primary identification in Cho (2017) is driven by school fixed effects (student-level identifiers are not present in the data), making the results vulnerable to selection bias from changes in student composition as well as endogenous changes in school AC status, neither or which are observed in the data.

This study seeks to explore the ways in which temperature may affect the human capital production process – and ultimately contribute to differences in living standards across countries and individuals – by using evidence from high-stakes exams in New York City public schools. Unlike the existing literature, it assesses contemporaneous and cumulative impacts of heat exposure simultaneously, and focuses on the role of adaptive investments and avoidance behaviors. Using data on high stakes exams taken in June allows for the joint estimation of short and medium run impacts, as well as an analysis of persistent impacts on educational attainment. The richness of the administrative dataset permits causal identification using within-student variation in temperature. School-level air conditioning data and analysis of bunching behavior permits an in-depth analysis of the potential extent of (and possible constraints to) efficient adaptation responses.

In related work using weather shocks as instruments for educational impacts, Goodman (2014) shows that snowfall can result in disruptions to learning by increasing absenteeism selectively across different student groups. Peet (2014) uses temperature, precipitation, and wind variation as instruments for pollution exposure in a sample of Indonesian cities and finds evidence of persistent impacts on student performance and labor market outcomes, though it is unclear to what extent temperature exerts a direct impact, and through what channels. A growing number of studies explores the impact of air pollution on student outcomes [Currie et al., 2009; Roth, 2016], and consistently find large impacts on absenteeism and exam performance. In the case of pollution during high-stakes exams in Israel (ELR, 2016), there is evidence for persistent and economically meaningful impacts that extend well beyond formal schooling.
2.2 Conceptual Framework and Empirical Predictions

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

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

1. We expect acute heat exposure for student $i$ in time $t$ to reduce contemporaneous exam performance, $\frac{\Delta h_{it}}{\Delta T_{it}} < 0$, if any of (a) direct flow utility, (b) marginal cost of effort, or (c) cognitive performance are adversely affected by temperature.

2. Short instances of heat exposure during an exam should not in principle reduce the stock of human capital $h_{it}$, but might nevertheless reduce overall human capital attainment, $\frac{\Delta h_{it+\tau}}{\Sigma T_{it}} \leq 0$, if the schooling environment features high costs of retaking exams, or in the presence of dynamic complementarities due to policies such as tracking.

3. Repeated heat exposure may reduce human capital accumulation and educational attainment over time: $\frac{\Delta h_{i}}{\Sigma_{0} \Delta T_{it}} \leq 0$. This may be due to reduced student cognition during class or study time, reduced teacher performance, or other heat-related illnesses. They may also arise from reduced attendance or increased violent behavior.

4. Any combination of information asymmetries between students and school administrators (or teachers and school administrators), market failures in the provision of local public goods (e.g. collective action problems in providing adequate electrification), or liquidity constraints in the context of air conditioning or other infrastructure investment, can lead to socially sub-optimal provision of thermal comfort in educational settings. Furthermore, we expect that even “optimal adaptation” need not imply zero residual impacts of thermal stress given non-zero costs of adaptation.
3 Institutional Context, Data, and Summary Statistics

3.1 New York City Public Schools

The New York City public school system (NYCPS) is the largest in the United States, with over 1 million students. The median student is relatively low-performing and low-income, though a substantial minority attend high-achieving magnet schools including Stuyvesant Academy and Bronx Science which consistently rank among the nation’s best.\(^{17}\)

The average 4-year graduation rate, at 68%, is below the national average of 81% but comparable to other large urban public school districts (e.g. Chicago, at 67%). System-wide averages mask considerable discrepancies in achievement across neighborhoods. Schools in the predominantly Black or Hispanic neighborhoods of Brooklyn and the Bronx have four-year graduation rates as low as 35% per year.

3.2 New York State Regents Exams

Each June, students in the state of New York take a series of high-stakes exams called “Regents exams”. These standardized subject assessments are administered by the New York State Education Department (NYSED) and are used to determine high school diploma eligibility as well as college admissions.

Regents exams are high-stakes for the average NYCPS student. Students are required to meet minimal proficiency status – usually a scale score of 65 out of 100 – in five “core” subject areas to graduate from high school.\(^{18}\) Many local universities including City University of New York (CUNY) use strict Regents score cutoffs in the admissions process as well: for instance, requiring that students score above 75 on English and Math simply to apply.

The vast majority of students take their Regents exams during a pre-specified two-week window in mid-to-late June each year.\(^{19}\) The test dates, times, and locations for each of

---

\(^{17}\) Approximately 19% of NYC students attend private schools – in particular, residents of the Upper East Side of Manhattan (70-80%). These students are not included in our sample.

\(^{18}\) The core subject areas are English, Mathematics, Science, U.S. History and Government, and Global History and Geography. In the data, these five core areas consist of 11 different subjects: Math (Integrated Algebra, Geometry, and/or Trigonometry), English, Science (Physics, Earth Science, Living Environment, or Chemistry), US History & Government, and Global History & Geography. In the analyses that follow “subject” will refer to this 11 category classification, as these subjects are taken on different dates within any given exam administration. The passing threshold is the same across all core subjects. Students with disabilities take separate RCT exams, and are evaluated on more lenient criteria. 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.

\(^{19}\) For any given student, exam takes are spread out across multiple days and years though, in effect, most exams are taken junior and senior year. Apart from the fact that most students take English their junior year, and Living Environment and Global History prior to other “advanced” sciences and US History
these Regents exams are determined over a year in advance by the NY State education authority (NYSED), and synchronized across schools in the NYCPS system to prevent cheating. Each exam is approximately 3 hours long, with morning and afternoon sessions each day, and are taken at the student’s home school.

All exams are written by the same state-administered entity and scored on a 0-100 scale, with scaling conducted according to subject-specific rubrics provided by the NYSED in advance of the exams each year. All scores are therefore comparable across schools and students within years, and the scaling designed in such a way that is not intended to generate a curve based on realized scores, which would complicate identification. I use standardized performance at the subject level as the primary measure of exam performance in this study, though the results are robust to using scale scores. Though centrally administered, Regents exams were locally graded by teachers in the students’ home schools, at least until grading reforms were implemented in 2011 in response to a series of media reports suggesting grade manipulation.

In summary, using scores from Regents exams to explore the impact of heat on human capital production offers several distinct advantages. First, they are high-stakes exams used to determine diploma eligibility and possibly affecting college enrollment. This means that, in addition to direct welfare relevance, they may also provide information about compensating behavior that is not available in low-stakes laboratory studies or voluntary assessments such as those in the NLSY and routine cognitive tests. Second, they are offered at a time of year when temperatures are likely to be hot sometimes but not always, due to the considerable day-to-day variability in June temperatures. Because they occur at the end of the school year, they are also more likely than periodic assessments or college entrance exams to reflect cumulative impacts of hot temperature that may have accrued over the course of the school year. Finally, Regents exams were centrally administered and compulsory for all public school students during the study period, meaning there is relatively little possibility of anticipatory alteration of exam timing based on weather forecasts, or for bias due to selection into taking the exam.

respectively, there do not seem to be any clear patterns in the timing of various subject exams throughout students’ high school careers. Some advanced students may take Regents subject exams during middle school or during early January waves, while failing students are required to retake exams in August.

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

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

As has been documented by Dee et al (2016), a substantial portion of NYC Regents exams featured bunching at passing cutoffs, clear evidence for discretionary grade manipulation by teachers, a phenomenon I verify in the analysis below.
3.3 Student Outcome Data

I obtain student-level information from the New York City Department of Education (NYC DOE). The data includes the universe of all public school students who took one or more Regents exams over the period 1999 to 2011. I obtain information on exam dates from archived Regents exam schedules from the NYC DOE database, and match observations according to exam code and year information.

3.4 Weather Data

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

3.5 School Air Conditioning Information

Information on building-level air conditioning equipment comes from records originally compiled by New York City School Construction Authority (SCA), which administers

---

23 I also use data from standardized math and English language and arts (ELA) exams administered in 3rd through 8th grade from NYC DOE to provide a measure of previous ability. Specifically, I calculate the average combined z-score of each student for whom previous standardized ELA and math exam records are available. Combined z-scores are constructed by computing standardized z-scores by subject and year, and computing the annual average by student. For students who are missing these records, I assign imputed average z-scores by decile of the realized Regents score distribution. These exams are generally taken in January and March, and feature substantially less temperature-related variation, due presumably to the lack of hot exam days. Cold days do not appear to affect ELA and math scores.

24 A detailed description of the matching procedure and subsequent sample restrictions can be found in the online appendix.

25 A detailed description of these corrections is presented in the online appendix. The primary results reported below are not sensitive to either of these corrections. The corrections reduce standard errors but leave implied point estimates relatively unchanged.

26 The air quality controls used here are admittedly crude, especially for localized pollutants such as ozone. Given the focus of the study, the relatively low levels of particulate matter during the sample period, and the high correlation between ozone and summertime temperature, I run analyses with and without controls for air quality but cannot separately identify the precise effects of pm, ozone, and temperature simultaneously.
detailed, building-level surveys for NYC public schools. While a centralized database was not publicly available, a web-scrape of individual school websites resulted in matchable records for 644 of the 890 middle and high school buildings in the study sample. The records include information on air conditioning equipment presence and maintenance status as of the year 2012. Unfortunately, the data does not provide AC installation or usage status by year, nor does it provide information regarding where within a school AC was available. As such, I take AC installation status as a rough proxy for the true variable of interest, which is effective AC utilization at the classroom level.

### 3.6 Summary Statistics

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

Table [1](#) presents summary statistics for the key outcome variables that form the basis of this analysis. The average student scores just around the passing cutoff, with a median score of 65 (sd = 17.9), though there is considerable heterogeneity by borough as well as student type. The student body is 40% Latino, 31% African American, 14% Asian and 13% White, and approximately 78% of students qualify for federally subsidized school lunch.

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

Figure [2](#) illustrates the source of identifying variation for short-run impacts, with temperatures weighted by exam observation and school location. Outdoor temperature during exams range from a low of 60°F to a high of 98°F [27](#). Cumulative heat exposure during the school year can be non-trivial as well, as suggested by Figure [3](#) which presents the incidence of days with maximum temperatures above 80°F by school year and borough. On average, NYC students experience between 19 and 39 days above 80°F per school year, with a mean value of 26.7 and a standard deviation of 5.6. There are on average 2.5 days during the school year above 90°F. Most of these days occur during the months of September, October, and June [28](#). Summer school students, which comprise roughly the bottom third of the student body but are not included in this

---

13

---

27 Day-to-day variation within the June exam period can be considerable, as suggested by panel [1](#) in Figure 2, which shows the variation in outdoor temperature by school and exam take across two consecutive test dates within the sample period.

28 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).
analysis, are subject to an additional 9 days above 90°F on average.

In 2012, 62% of all NYC public school buildings were reported as having some kind of air conditioning equipment on its premises, including window units, which means that fully 38% of school buildings did not have any form of air conditioning equipment available (Figure 7a). Of the 62% that were reported as having air conditioning, 42% (259 out of 644) were cited as having defective components, according to the third-party auditors conducting the BCAS assessments.

4 Empirical Strategy and Primary Results

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

4.0.1 1. Does Exam-Time Heat Exposure Affect Exam Performance?

Figure 4a presents a binned scatterplot that motivates this analysis. It shows the relationship between standardized exam score and percentile of observed exam-day temperature, plotting residual variation after controlling for school fixed effects and average differences across subjects and years, in addition to controls for daily precipitation. It suggests that exams taken on hot days exhibit lower scores.

To further isolate the causal impact of short-run temperature fluctuations on student performance, I exploit quasi-random variation in day-to-day temperature across days within student-month-year cells, focusing on the main testing period in June. While it is unlikely that temperature is endogenous to student behavior, nor is it likely for students to select into different temperature treatments given the rigidity of exam schedules, time-varying unobservables may still be correlated with weather realizations. This motivates a specification that includes year, time of day, and day of week fixed effects. For instance, if certain subjects tend to be scheduled more often in the afternoon when students are relatively fatigued (as in Sievertson et al, 2016) or toward the end of the exam period (Thursday as opposed to Monday), we may expect mechanical correlation between temperature and test scores that is unrelated to the causal effect of temperature on student cognition or effort.

I therefore estimate a baseline model that includes student-by-year and subject fixed effects, as well as controls for time of day and day of week:

\[
Y_{ijsty} = \gamma_i + \eta_s + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \beta_3 Time_{st} + DOW_{sty} \beta_4 + \epsilon_{ijsty} 
\]  

(1)

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

Student-by-year fixed effects ensure that we are comparing the performance of the same student across different exam sittings within the same year, some of which may be taken on hot days, others not, leveraging the fact that the average student takes 7 June Regents exams over the course of their high school career (on average between 3 and 4 per year). Subject fixed effects control for persistent differences in average scores across subjects. Year fixed effects control for possible spurious correlation between secular performance improvements and likelihood of hotter exam days due, for instance, to climate change.\(^{29}\)

Table 4b presents the results from running variations of equation (4) for the subset of students who take at least 2 exams in any given year.\(^{30}\) As suggested by the first row of columns (1)-(4), exam-time heat stress exerts a significant causal impact on student performance. The estimates are robust to allowing for arbitrary autocorrelation of error terms within boroughs and test dates, which is the level of exogenous temperature shock recorded in the data, as well as to using the full sample of scores (i.e. retaining all students with 1 or more exam records and replacing student fixed effects with school fixed effects and a vector of demographic controls).

Taking an exam on a hot day reduces performance by approximately \(-0.008\) standard deviations (se=0.002) per \( ^\circ F \).\(^{31}\) This amounts to \(-5.2\) percent of a standard deviation in performance per standard deviation increase in temperature, or \(-14.7\) percent of a standard deviation if a student takes an exam on a 90\( ^\circ F \) day as opposed to a more optimal 72\( ^\circ F \) day.\(^{32}\)

The effect of a 90\( ^\circ F \) day is thus comparable in magnitude to roughly 1/4 of the Black-White score gap, or 3/4 of the within-school Black-White score gap. This effect is comparable in raw magnitude to the impacts on mathematical reasoning found by GHN (2015),

\(^{29}\)To the extent that temperature variation within student-month-year cells are uncorrelated with unobserved factors influencing test performance, one would expect the coefficient \( \beta_1 \) to provide an unbiased estimate of the causal impact of temperature on exam performance, subject to attenuation bias due to measurement error in weather variables, as well as downward bias from upward grade manipulation. It is worth noting that it is possible for measurement error to be non-classical in a way that biases the estimates upward. For instance, if (1) the average classroom has more students in lower performing schools, (2) experienced classroom temperature scales non-linearly with outdoor temperature, and (3) students in lower performing schools are more susceptible to heat stress, then \( \beta_1 \) may actually be biased upwards.

\(^{30}\)Results using un-standardized scale score as the dependent variable are presented in the appendix. In models where student fixed effects are replaced by school fixed effects, a vector of demographic control variables are included to control for possible selection into subjects by student type.

\(^{31}\)In terms of scale scores, the effect is -0.13 points (se=0.04) per \( ^\circ F \), or -0.20\% per \( ^\circ F \) relative to a sample mean of 64.8 points.

\(^{32}\)Precipitation has a slightly positive effect, and ozone has a negative but insignificant effect, with a 1 standard deviation increase in ozone corresponding to a point estimate roughly 1/5th the size of a 1 standard deviation temperature effect.
who find a 90°F day to reduce NLSY math scores by approximately -0.12 standard deviations.\footnote{The effect documented here is also similar in magnitude the effects on Israeli high school exit exams of a standard deviation increase in pm2.5 and CO pollution found by ELR (2016). I find little evidence for adverse impacts of pm2.5, perhaps because average concentrations are much lower in NYC than in Israel, as well as the fact that the performance impacts documented by ELR (2016) are highly non-linear, driven mostly by heavily polluted days with pm2.5 above 100 micrograms per cubic meter. The maximum recorded value of pm2.5 in my data set is 38.8 micrograms per cubic meter.}

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

These results provide strong evidence that heat stress affects student performance, either by reducing raw cognitive ability and/or by increasing the disutility of effort which in turn affects students’ willingness or ability to maintain focus during a three-hour exam. They suggest that temperature in the learning environment may play an important role in determining student outcomes, and that whatever compensatory effort is exerted by students due to the high stakes nature of some exams may not be enough to offset the physiological impacts of temperature documented in the lab. However, these results are not in themselves directly informative of welfare impacts or effects on the stock of human capital.

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

Heat exposure during an exam, while reducing cognitive ability or concentration temporarily, presumably does not reduce the stock of knowledge or human capital \textit{per se}, at least not immediately through the physiological impact of heat stress itself.\footnote{In theory, it is possible that acute periods of stress can lead to the rewiring of neurons in such a way that alters one’s memory semi-permanently, which could mean that acute heat stress during a high-stakes exam could lead students to “forget” material they already knew, or become more confused or less confident about it in future applications, though in practice this seems unlikely.}

In an idealized, friction-less world with fully flexible educational institutions, unlucky students who fail a subject due to a hot exam sitting would immediately retake the exam until she believes her “true ability” has been reflected in the exam score: $s_{it} = h_{it}$. In this world, random heat exposure during exams should not affect the final amount of schooling achieved.

However, in the presence of institutional rigidities that limit the effective number of
possible retakes or impose time and effort costs to retaking an exam (by, for instance, requiring students to attend remedial courses), it is possible that even short-run heat exposure can lead to ripple effects on long-run educational attainment. In addition, employers may treat students who graduated from high school in five or six years differently from those who graduated “on-time” for a variety of reasons, reducing the pecuniary return to education once a subject has been failed the first time. Similarly, exam scores may serve as important signals within the education system – to the student herself, to her peers, or to her parents and teachers – leading to dynamic complementarities in human capital investment [Cunha and Heckman, 2007, Diamond and Persson, 2016].

Recent evidence suggests persistent effects of temporary score shocks in the context of teacher manipulation [Dee et al., 2016, Diamond and Persson, 2016] and air pollution [Ebenstein et al., 2016], with as yet inconclusive evidence regarding the specific mechanisms by which these may occur. 

Pass Rates and College Proficiency

If heat exposure during an exam pushes some students below important (cardinal) score thresholds that affect access to further educational opportunities, one might expect even small “doses” of heat exposure to potentially lead to lasting consequences for educational attainment.

In NYC, students must score a 65 or above to pass a given subject exam and thus have it count toward receiving a HS diploma. This cutoff does not change based on the realized distribution of performance in any given year. Students are also assigned “college ready” or “proficient” status on each of the subjects in which they receive a grade of 75 or higher and “mastery” status for scores of 85 or higher. Beyond any personal motivational or within-school signalling value, these designations carry real weight externally in the sense that many local colleges and universities such as City University of New York (CUNY) use strict score cutoffs in their admissions decisions.

---

35 [Dee et al., 2016] find substantial impacts of upward score manipulations on graduation status, especially for students who scored in the manipulable zone. Using administrative records from Swedish middle schools, [Diamond and Persson, 2016] also find substantial effects of upward score manipulations on subsequent performance, graduation likelihood, and even later life income. [Ebenstein et al., 2016] find that Israeli high school students who receive lower scores on their Bagrut (high school exit) exams due to air pollution are less likely to receive Bagrut certificates (comparable to high school diplomas) and receive lower wages later in life.

36 Until 2005, low-performing students were allowed the option of applying to receive a “local diploma” which required scores of 55 and above for exams to count toward the diploma. In the following regressions, I use the more stringent and universally accepted standard of “Regents Diploma” as the definition of passing score, as do [Dee et al., 2016]. Results of running the regression analyses below using the “Local Diploma” cutoff feature similar (slightly more negative) point estimates.

37 The scale score needed to be considered “college ready” differs by subject. For instance, for admission to CUNY schools, a student can demonstrate the necessary skill levels in reading and writing by meeting any of the following criteria: SAT Critical Reading score of 480 or higher; ACT English score of 20 or higher; N.Y. State English Regents score of 75 or higher.
To estimate the impact of contemporaneous heat stress on the likelihood that a student scores at or above the passing and proficiency thresholds, I run variations of the following models:

\[
p_{ijst} = \gamma_{iy} + \eta_s + \beta_1 T_{jst} + X_{jst} \beta_2 + \beta_3 Time_{st} + DOW_{st} \beta_4 + \epsilon_{ijst} \tag{2}
\]

\[
c_{ijst} = \gamma_{iy} + \eta_s + \beta_1 T_{jst} + X_{jst} \beta_2 + \beta_3 Time_{st} + DOW_{st} \beta_4 + \epsilon_{ijst} \tag{3}
\]

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

The results suggest that acute heat exposure can have significant short term impacts on student performance, with potentially lasting consequences. Exam-time heat stress reduces the likelihood of passing by 0.31 (se=0.12) percentage points per °F, or -0.54% per °F from a mean likelihood of 0.57.\(^{38}\)

Impacts on the likelihood of achieving proficiency status are slightly larger in aggregate, with a magnitude of -0.31 (se=0.10) per °F, or -0.96% per °F hotter exam-time temperatures (relative to a mean likelihood of 0.32).\(^ {39}\)

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

**Graduation Status**

Perhaps more importantly, I find evidence that these short-run heat exposures affect final educational attainment by affecting the likelihood that students graduate from high school. Figure 5a presents a binned scatterplot of 4-year graduation status on average exam-time temperature by student during June Regents exams up through the student’s senior year. It plots residual variation in a dummy variable denoting graduation status, controlling for school-level averages, student-level demographic characteristics, and the number of June Regents exams taken by student. It suggests that students who experience greater exam-time heat stress are less likely to graduate on time.

Whereas short-run impacts of heat stress could be identified within student cells, long-run impacts on graduation status cannot, the outcome variable is no longer date-specific.\(^ {40}\)

\(^{38}\)These results are presented in tabular form in the appendix.

\(^{39}\)Unless higher-ability students are more sensitive to heat stress, this discrepancy seems likely to be driven in part by grade manipulation around the passing threshold.

\(^{40}\)Graduation status is student specific, and while NYC DOE data provides 4, 5, and 6-year graduation and dropout status by student, the way in which the data is coded does not allow reliable matching by year.
This poses additional challenges to causal identification. Computing a measure of average heat exposure across multiple exam sittings by student results in mechanical correlation between average experienced temperature and the number of exams such that students who take more exams are more likely to be assigned average temperature values closer to the climatic mean in that month, and students who take fewer exams are more likely to be assigned extreme values.\footnote{That is, assuming that the average June climate in New York City can be represented by a distribution of daily temperature realizations, the average temperature across multiple days will exhibit a form of mean-reversion as one increases the number of draws from the underlying climate distribution (see online appendix for a visual representation).}

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

$$g_{ijcn} = \alpha_0 + \alpha_1 T_{ij} + X_{ij} \alpha_2 + \chi_j + \theta_c + Z_i \alpha_3 + \text{exams}_n \alpha_4 + \epsilon_{ijc}$$ (4)

\(g_{ijcn}\) is a dummy denoting whether student \(i\) in school \(j\) and entering cohort \(c\) who takes \(n\) June Regents exams over the course of her high school career has graduated after 4 years in high school. \(T_{ij}\) denotes the average temperature experienced by student \(i\) while taking June Regents exams in school \(j\), up through her senior year. \(X_{ij}\) is a vector of weather controls averaged at the student-by-school level. \(\chi_j\) denotes school fixed effects; \(\theta_c\) denotes cohort fixed effects; \(Z_i\) is a vector of student-level controls including race, gender, federally subsidized school lunch eligibility, and previous ability (combined ELA and math z-scores); and \(\text{exams}_n\) denotes a vector of number of June exams fixed effects.

The parameter of interest is \(\alpha_1\), which captures the impact of an additional degree of heat exposure over all June Regents exams on the likelihood of graduating on time.\footnote{The intuition is that variation in experienced temperature among students in the same school and cohort will be plausibly uncorrelated with residual variation in graduation status within school and cohort cells. Suppose there are two students, Jill and Karen, who entered high school in 2000. In 2001, because of differences in the sequence of subjects that Jill and Karen took, Jill takes Regents exams on Monday, Wednesday, and Thursday, and Karen takes Regents exams on Monday, Tuesday, and Friday. Suppose a similar phenomenon occurs during their sophomore, junior, and senior years, such that they take the same overall number of June exams. The variation in overall experienced temperature between Karen and Jill in 2001 will likely be exogenous to any unobserved differences in Jill and Karen's likelihood of graduating from high school.} School fixed effects account for potential omitted variable bias due to unobserved determinants of graduation rates being correlated with average temperature in the cross-section (e.g. if urban heat island effects are stronger in poorer neighborhoods). Cohort fixed effects in graduation rates allow for the possibility that heat exposure and graduation rates are correlated due to secular trends in both variables – though warming trends and average...
improvements in NYC schools would suggest this effect to lead to downward rather than upward bias in the estimate of $\alpha_1$.

Table 5a presents the results from running variations of equation 4 with and without school and cohort fixed effects, as well as flexible controls for the number of exams. Standard errors are clustered at the borough by date and time level, based on the intuition that this represents a conservative estimate of the level at which quasi-random temperature variation occurs, though the results are once again robust to alternative levels of clustering.

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

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

4.0.3 3. Does Cumulative Heat Exposure Reduce the Rate of Learning?

Given the effects documented above, and the vastly uneven distribution of hot days across countries, one might be interested in understanding the extent to which repeated heat exposure over time may reduce the effectiveness of learning. Emerging evidence [Cho, 2017, Garg et al., 2017] suggests that cumulative heat exposure can hinder skill formation – or at least exam preparation – though the exact mechanism(s) through which this may occur is as yet unclear.

Figure 6a presents a binned scatterplot of Regents score on the number of days with max temperatures between 80°F and 90°F, controlling for exam-day temperature and precipitation, as well as school-, subject- and time of day fixed effects. The figure suggests that hot days may be reducing learning attainment, at least as reflected in end-of-year exam performance.

Because Regents exams are subject-specific and usually administered at the end of the school year during which that subject was taken, they provide a suitable opportunity for uncovering potential cumulative learning impacts of heat exposure during the school year. On the other hand, because each subject exam is usually only taken once per year and

---

43See online appendix for methodology used in calculating the number of students and exams affected. These figures do not explicitly account for upward grade manipulation, and therefore likely represent a lower bound.
observed over the course of 13 years in my data set, and because cross-sectional variation in heat exposure within New York City is relatively limited, the analysis is likely to exhibit limited precision compared to the estimates of short-run exam-day effects.

To identify the impact of cumulative heat exposure on learning, I collapse the data to the school by subject and month (year) level. I retain subject-level variation in order to estimate the impact of cumulative heat stress while controlling for the short-run impacts of contemporaneous heat stress documented above.

I estimate variations of the following model:

\[
y_{jst} = \beta_0 + \beta_1 T_{jst} + X_{jst} \beta_2 + \sum_d \beta_d D_{jT}^d + \chi_j + \eta_s + Z_{jst} \beta_4 \\
+ \beta_5 T_{t} \gamma_{jst} + DOW_{st} \beta_6 + \beta_7 Y_{ear} + \beta_8 Y_{ear}^2 + \beta_9 Y_{ear}^3 + \epsilon_{jst}\]  

where \(y_{jst}\) denotes the average Regents z-score (standardized once again by subject over the study period) for students in school \(j\) taking subject \(s\) on date and time \(t\), during year \(y\); \(T_{jst}\) denotes exam-time outdoor temperature at school \(j\) for subject \(s\) on date and time \(t\), during year \(y\); \(\gamma_j\) denotes school fixed effects; \(\eta_s\) denotes subject fixed effects; and \(Z_{jst}\) represents a vector of demographic controls averaged at the school by take (subject-month-year) level. \(T_{t}\) represents a dummy for time of day (Time=1 denotes afternoon exam), and \(DOW_{st}\) is a vector of fixed effects for day of week. \(Y_{ear}...Y_{ear}^3\) denotes a cubic time trend in scores.

The variable \(DD_{jy}\) denotes a vector of day counts in a series of degree-day bins during the preceding school-year \((y)\), beginning with the first day of the fall semester up to the first day of the testing period the following June. I use a number of bin classifications for hot days, motivated by the existing literature (e.g. [Barreca et al., 2016]) as well as the analyses presented in the previous section, which find negative impacts of heat stress beginning around 72\(°\)F. The coefficients of interest correspond to the “hot” degree day bins, around or above 80\(°\)F, and represent the correlation between the number of hot days in a school year and end-of-year exam scores. \(X_{jst}\) denotes a vector of contemporaneous heat exposure. 

---

44 This is for a couple of reasons. Because I do not observe student-specific measures of cumulative heat exposure over the preceding school year, keeping student-level exam observations will likely introduce additional measurement error, since cumulative heat exposure during preceding school years is measured at the school level and some students may have been present for more days than others or live in neighborhoods that are more prone to heat stress than others.

45 Time trends are included in lieu of year fixed effects to account for possible secular changes in performance over time that may be spuriously correlated with shifts in climate over the study period. The results presented below are robust to including school-specific trends as well as subject-specific trends.

46 School years are defined such that the test year corresponds to the year in which the spring semester of the academic year occurs. For instance, \(y=2000\) corresponds to the 1999-2000 school year; \(y=2001\) to the 2000-2001 school year, etc.

47 The preferred analysis flexibly divides temperature days into 10 degree bins, beginning with 10\(°\)F to 20\(°\)F up to 100\(°\)F and above, omitting the “optimal” bin, which the data and previous work using similar approaches (e.g. [Deschênes and Greenstone, 2011]) suggests to be around 60\(°\)-70\(°\)F.
and cumulative weather controls, including precipitation and dewpoint on exam day as well as during the preceding school year, and annual snowfall, which is taken from weather station readings in Central Park and assigned uniformly across all schools in the city.

I run variations of equation 5 that allow for a flexible characterization of the reference category against which we can interpret the impact of hot days. Table 6b presents the results from these analyses, with columns (1), (2), (3), and (4) corresponding to specifications that control for (1) hot (70-80°F and above) days only, (2) hot and cold (30-40°F and below) days only, (3) hot days only and average daily maximum temperature over the school year, and (4) all degree day bins from 0°F to 100°F omitting the 60-70°F bin respectively.

The results are suggestive of cumulative learning impacts due to heat exposure during the school year. First, note that the short-run impacts persist in all specifications, with relatively stable point estimates of similar magnitude from the results presented above. Focusing on column (3), which controls for average daily max temperatures during the school year as well as for the contemporaneous effect of exam-time temperature, days between 80° and 90°F have a negative impact of between -0.011 (se=0.0031) and -0.012 standard deviations (se=0.0033), or approximately 1% of a standard deviation per hot day. Estimates for days above 90°F are much noisier given the relatively limited number of such days during term. Results in columns (1), (2) and (4) suggest a similar pattern of hot days during the preceding school year reducing exam performance.

These estimates suggest that a one standard deviation (3.91 day) increase in the number of days with maximum temperatures above 80°F can reduce learning by approximately 0.04 standard deviations, as measured by end-of-year exam performance. These impacts are on par with the learning impacts of a 0.4 standard deviation reduction in average teacher value-added [Chetty et al., 2014], or 1/2 of the impact of reducing class size from 31 to 25 [Angrist and Lavy, 1999].

Though data limitations do not permit the analysis of the impacts on later-life outcomes such as wages or health directly, these results should be interpreted in light of recent analyses [Chetty et al., 2014] which examine the same population of NYC students and find significant impacts of improved learning on later-life outcomes.

4.0.4 4. How do agents adaptively respond?

If heat stress affects student performance in economically meaningful ways, standard micro theory would predict that students, parents, and teachers would respond to mitigate this

---

48Cold days appear to have a negative impact on end-of-year performance as well, particularly in the case of days with maximum temperatures between 30 and 40°F. This seems consistent with previous work by [Goodman, 2014] who finds that the number of snow days and the amount of local snowfall adversely predict end-of-year performance in Massachusetts public schools.
impact, presumably along the most cost-effective margins. However, efficient adaptation to environmental shocks may be contingent on the quality of existing institutions, as has been found in the context of mortality responses to environmental disasters such as earthquakes [Kahn, 2005]. In the short run, flexible institutions and smart policies may help to minimize the realized welfare impacts of a string of hot days or a particularly hot year. In the long run, one might expect institutions to mediate the responsiveness of school systems to a changing climate, with some adapting more quickly than others. Weak institutions may lead to greater distance from the efficient adaptation frontier in both the short and long run.

In the context of energy infrastructure, there may also be market failures arising from imperfect information, liquidity constraints, or principal agent problems that drive a wedge between realized and optimal adaptive investments. For instance, Gertler et al. (2016) find evidence for substantial liquidity constraints in the adoption of energy-intensive appliances, and a long literature has noted the potential for principal agent problems in the context of renter-occupied homes [Allcott and Greenstone, 2012]. It is possible that such market failures are more prevalent in school environments, where the link between educational inputs and market outcomes is imperfectly observable and often involves long lags. The fact that, in the US, many public school districts and teachers unions have clashed over installation of air conditioning suggests possible information problems associated with the efficacy of various adaptation options.49

This section explores two types of adaptive responses as a window into the ways in which institutional factors may or may not constrain the realization of first-best adaptation investments: air conditioning and teacher responses.

Air Conditioning

Air conditioning is a potential adaptation strategy which, at least in the context of health and labor productivity, has proven to be effective at mitigating the realized impacts of hot temperature [Barreca et al., 2016] [Park, 2016].

Average air conditioning penetration for commercial and residential buildings in New York City is around 90%, suggesting that historical rates of heat exposure may be frequent enough to merit investment in AC in the region. As noted above, however, fewer than 62% of NYC public schools reported having some form of AC equipment as of 2012. Among those that do, a substantial proportion are reported as having defective system components. This AC penetration gap – between public schools and residential and commercial buildings

49For instance, during a major teacher union strike in Chicago in 2012, “Timetable for air conditioning” was listed as one of four major contract demands. An agreement to provide universal air conditioning in Chicago public schools was not reached until 2016.
in the same region is consistent with the presence of institutional or other barriers to efficient adaptation.\footnote{Summer vacation means that school occupancy may be lower during hotter months, making this comparison less informative. However, it is worth noting that a substantial fraction (often 25\% or more) of public school students are required to attend summer school due to low achievement status.}

To compare the impact of heat exposure on high-stakes exams across schools with and without AC units, I estimate equation \ref{eq:1} separately for sub-groups of students who took exams in schools with and without central air conditioning, as well as for sub-groups in schools with and without any air conditioning at all as of 2012. The results from these regressions are reported in Table 7b. Column (1) reproduces the main effect on the entire sample. Columns (2), (3) and (4), (5) present results for sub-groups with central AC, with any AC, without central AC, and without any AC respectively. The point estimates are smaller and insignificant for sub-samples with AC units, -0.0053 (se=0.0029) and -0.00517 (se=0.0027), relative to sub-samples without AC: -0.0065 (se=0.0027) and -0.0062 (se=0.0026), for schools with and without central AC or any form of AC respectively.\footnote{Point estimates are slightly smaller (less negative) for schools with AC in the case of pass rates and proficiency status as well: for instance, -0.0029 (se=0.00097) versus -0.0039 (se=0.0010) for proficiency status.}

These differences are not statistically different, however, suggesting either that existing AC units have an only mildly protective effect (perhaps because they are not always in operation), or that AC status is measured with substantial noise.\footnote{The available data provide relatively crude proxies of the true variable of interest, which is effective air conditioning utilization: i.e. the amount of climate control functionally realized by students. While BCAS provides data on air conditioning installation at the school building level for the year 2012, it does not include information on which areas within a given school have working air conditioning, nor does it tell us during which years AC was present. The BCAS data also does not provide information on whether existing AC equipment was actually utilized on any particular day.}

It is also possible that the relatively old building stock in NYC places additional constraints to the effectiveness of some AC systems.\footnote{Engineering studies show that adding AC units to existing structures ad hoc (e.g. window units) can add substantial scope for disruption due to added noise and reduced air quality, since they often are not accompanied by integrated changes to ventilation and heating systems, especially in the case of older buildings \cite{Niu, 2004}.}

More research is needed to understand whether and to what extent school air conditioning represents an effective adaptation response to climate change, and what if any market failures may prevent efficient uptake.

\section*{Teacher Responses}

Using a similar dataset from 2003 to 2012, \cite{Dee et al., 2016} document systematic grade manipulation by NYC teachers on NY State Regents exams. They find that most of the manipulating behavior occurred at or around passing margin of 65 and that, while varied in magnitude across schools and student types, such manipulation was a near-universal phenomenon within the NYCPS system.
Upon careful analysis of competing explanations, the authors suggest this is most likely driven by the goodwill of teachers who seek to offset the impact of “a bad test day”:

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

A hot test day may be viewed as a bad test day, particularly if air conditioning is inadequately provided. In that case, could discretionary grade manipulation have been a response to perceived performance impacts of heat stress, a form of second-best adaptation given institutional constraints?

Teachers may be able to observe or at least intuit the disruptive impacts of elevated classroom temperatures on test day, especially since exams are taken in students’ home schools and graded by a committee of teachers from that school. If graders are benevolently motivated, as Dee et al suggest, they may be inclined to engage in more grade manipulation precisely for those exams that took place under disruptively or unusually hot conditions. One might call this selective response by teachers “adaptive grading”, a second-best adaptation strategy undertaken in the presence of institutional barriers to more efficient forms of (structural) adaptation.

**Estimating Adaptive Grading: Bunching at Score Thresholds**

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

We would expect any form of grade manipulation for students who initially score just below the passing cutoff, even “indiscriminate” grade manipulation uncorrelated with exam-time temperature, to downward attenuate the estimates of heat-related performance impacts uncovered above. Indeed, running equation 1 on the subset of grades that fall within the manipulable zone as established by [Dee et al., 2016] based on the institutional features of NY Regents exams and described in greater detail below, I find that the point

---

54The only case in which grading process may affect our interpretation of causality is if teachers grade differentially according to the temperatures they experience while grading, and the temperature during the exam is correlated with temperature during grading. If hot temperatures make teachers less productive and make more errors, this will simply add noise to the score variable. If hot temperature makes teachers irritable and more punitive in grading, then we might expect the beta coefficient to be picking up some of the correlation between test day temp and grading punitive-ness, although the most striking feature of the histograms above is that the majority of grade manipulation seems to be positive in direction, making this unlikely in practice.
estimate for the impact of temperature is substantially reduced and no longer significant: \( \beta_T = -0.0007 \) (se=0.0024) as opposed to \(-0.0082\) (se=0.0021) in the full sample.

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

\[
F_{ks} = \sum_{i=0}^{q} \psi_{ismyj}(Score)^i + \sum_{i\in-M_{cs},+M_{cs}} \lambda_{ismyj} \mathbb{1}[Score = i] + \epsilon_{ksmyj} \tag{6}
\]

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

As Dee et al point out, in other applications of “bunching estimates”, including constructing counterfactual distributions of taxable income around a kink in marginal taxes [Chetty et al., 2011], it has not generally been possible to specify an ex ante range of the variable in which manipulation might take place. Such ex ante designations are possible, however, in the case of NYC Regents exams because of known features of the NY Regents exams, including mandatory regrading policies (up until 2011) and published raw score to scale score conversion charts. Using this information, Dee et al are able to identify the range of potentially manipulable scores on both the left and right sides of the proficiency cutoffs (55 and 65). Following their strategy, I define a score as manipulable to the left of each cutoff if it is between 50 - 54 and 60 - 64, and manipulable to the right if it is between 55 - 57 and 65 - 67 as a conservative approximation of their subject-and-year-specific scale score-based rubric.

This generates a set of predicted fractions by score and subject. I verify that the average amount of bunching observed in my data is similar to that documented by [Dee et al., 2016], who find that approximately 6% of Regents exams between 2003 and 2011 exhibited grade manipulation.

55In practice, I use a fourth-order polynomial \((q=4)\) interacted with exam subject \(s\), but constant across years for the same exam subject. As [Dee et al., 2016] suggest, realized bunching estimates are not sensitive to changes in the polynomial order or whether one allows the polynomial to vary by year or subject. As a robustness check, I also estimate a linear approximation of the above estimator by generating predicted fractions using a linear spline between boundary points along the distribution that are known to be outside the manipulable range by subject. I then generate an estimate of the extent of bunching by school-subject-month-year cell, taking the absolute value of the distance between observed and predicted fractions by Regents scale score. The results are similar using this simplified measure of bunching.

56For the years 1999-2011, and using the subject-specific bunching estimator above, I find that 5.8% of all June Regents exams exhibited bunching at or near the passing score cutoffs.
I then calculate observed fractions for each score from 0 to 100 by school, month, year, and subject, and generate a measure of bunching that integrates the differences between observed and predicted fractions: that is, summing the excess mass of test results that are located to the right of the cutoff (above the predicted curve) and the gaps between predicted and observed fractions of test results to the left of the cutoff (below the predicted curve):

$$\zeta_{smyj} = \frac{1}{2} \sum_{i \in +M_{ck}} (F_{ks} - \hat{F}_{k_{smyj}}) + \frac{1}{2} |\sum_{i \in -M_{ck}} (F_{ks} - \hat{F}_{k_{smyj}})|$$  \hspace{1cm} (7)

where $\zeta_{smyj}$ denotes the degree of bunching at the passing cutoff for subject $s$, month $m$, year $y$, and school $j$. I then examine the relationship between $\zeta_{smyj}$ and exam-time temperature in that cell, which corresponds to the temperature experienced by students taking subject $s$ in school $j$ in June of year $y$, with controls for precipitation and humidity.

**Adaptive Grading**

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

$$\zeta_{smyj} = \delta_0 + \delta_1 T_{smyj} + \delta_2 X_{smyj} + \chi_j + \eta_s + \delta_3 Year_y + \delta_4 Year^2_y + \delta_5 Year^3_y + \epsilon_{smyj}$$  \hspace{1cm} (8)

where $T_{smyj}$ denotes temperature, $X_{smyj}$ denotes precipitation and humidity, $\chi_j$, $\eta_s$, and $\theta_y$ denote school, subject, and year fixed effects respectively, and $Year_y...Year^3_y$ denotes a cubic time trend in scores. The parameter of interest is $\delta_1$, which represents the increase in grade manipulation due to exam-time temperature.

The amount of bunching increases by approximately 0.10-0.16 percentage points per degree F, or 1.7% to 2.8% per degree F hotter exam-time temperature relative to a mean of 5.8 percentage points, with significantly positive coefficients in specifications with and without school and year fixed effects. These estimates are presented in tabular form in the appendix.

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

---

57 Recall that Regents exams were, up until 2011, graded by teachers in students’ home schools. To the best of my knowledge, they were graded at the home school either in the evening following the exam or on a pre-specified day at the end of each month-specific exam period (for instance, the last Friday of the exam period), which means that it is possible that teachers remember which exams were subject to more heat stress even within a given exam period.

58 It is theoretically possible that if 1) teachers engage in punitive grade manipulation which is affected by negative affect, which has been shown to increase with high temperatures, and 2) temperature on the day of the exam is positively correlated with temperature during grading, then the above estimates are attenuated due to this reverse effect of punitive grading by hot-tempered teachers.
an intuitive sense of whether a particular student scored below his or her “true ability”, irrespective of whether or not this was due to temperature or other exam-time conditions, and that they respond by manipulating grades in the case of students on the passing margin.

It is also possible that, due to the distributional properties of most Regents exams, heat-related performance impacts may lead to a mechanical increase in the number of grades that fall within the manipulable zone, and thus a correlation between bunching behavior and exam-time temperature.

Irrespective of whether teachers’ explicit intentions are to compensate for heat-related impacts, the realized effect has been for this behavior to mitigate the adverse welfare impacts associated with exam-time heat exposure. I find however that, even when controlling for the potential increase in manipulable scores on hot exam days, there is evidence for more grade manipulation after hot exam takes.

Figure 8b presents a binned scatterplot of the bunching estimator and exam-time temperature by subject-month-year-school cell, adding school fixed effects to allow for arbitrary differences in the average amount of grade manipulation across schools, and expressing the bunching estimate as a proportion of scores within the manipulable zone (50-54, 60-64). It suggests a clear positive relationship between the degree of grade manipulation and the ambient temperature during the exam being graded. It is consistent with teachers boosting students’ grades more often when students experienced hot testing environments, and may suggest substantial institutional constraints to more efficient adaptation responses such as air conditioning or changes in exam format or timing.

5 Discussion and Conclusion

This paper explores the impact of heat stress on the human capital production process. Using administrative data from the largest public school district in the United States, I find that hot temperatures exert a causal, statistically significant, and economically meaningful impact on student outcomes by reducing performance on high-stakes exams as well as possibly reducing the amount of learning achieved over the course of the school year. The research design exploits quasi-random, within-student temperature variation to identify the causal impact of hot days on performance. The breadth and depth of the data set allows not only for credible causal estimation of the adverse impacts of heat stress, but also an assessment of how the optimality of adaptive responses by students and teachers may depend on institutional settings.

Taking a high stakes exam on a 90°F day results in 15% of a standard deviation reduction in exam performance relative to a more optimal 72°F day, controlling for student ability. This amounts to roughly 1/4 of the total Black-White test score gap. These short-
run performance impacts can have non-trivial effects on longer-run educational attainment. A 90°F day results in a 10.9% lower probability of passing a subject, and, for the average New York City student, a 2.5% lower likelihood of graduating on time. I estimate that, over the period 1998 to 2011, upwards of 510,000 exams that otherwise would have passed received failing grades due to hot temperature, affecting at least 90,000 students, possibly many more.

The evidence is also consistent with emerging evidence suggesting that repeated heat exposure can disrupt learning and reduce the rate of human capital accumulation [Cho, 2017], though the effects documented here are imprecisely estimated. Cumulative heat exposure over the course of the preceding school year, measured by the number of days where temperatures exceed 80°F, is associated with non-trivial reductions in end of year exam performance, controlling for the exam-day effects of heat stress noted above. A year with five additional 80°F+ days is associated with 2.1% of a standard deviation reduction in learning on average, effects that are similar in magnitude with previous findings [Cho, 2017]. This point estimate suggests an effect that is comparable to eliminating half of the gains from having a one standard deviation higher value-added teacher for a school year, though more careful research is needed to examine whether they result in similar effects on later-life outcomes, given well-documented fade-out in teacher-driven score effects [Cascio and Staiger, 2012] and the possibility that better teachers impart important skills not captured by subject exams [Cunha and Heckman, 2007].

A series of additional analyses provide a window into the ways in which institutional factors may affect adaptation. The evidence suggests that, as has been shown in the context of human health [Kahn, 2005], the quality of institutions may be an important factor that determines the realized welfare impacts of any given environmental shock, including hotter temperatures from climate change.

Building level air-conditioning data suggests that school air conditioning may be sub-optimally provided. At less than 62% (as of 2012), the AC penetration rate in New York City public schools seems to be far below residential and commercial sector averages for the region (over 90%). Existing AC does not appear to have a protective effect, though it is possible that this may be due to lack of timely utilization or other factors specific to NYC schools (e.g. old building stock).

Perhaps in response to the lack of first-best adaptation investment, teachers seem to have selectively boosted grades of students who experience hot exam sittings. Bunching estimators used in previous work [Dee et al., 2016], suggest a pattern of grade manipulation that is highly correlated with temperature during the test session in which a given subject exam was administered. One interpretation is that teachers may have been trying to offset some of the long-term consequences of short-term heat stress, which presumably affects students’ scores but does not reduce human capital per se. The findings underscore the need
for more careful research on the impact of school infrastructure on student performance, as well as potential market failures that may drive a wedge between realized adaptation strategies and the efficient adaptation frontier.

These results have several implications. First, they suggest that temperature should be included among the long list of relevant inputs to schooling. For instance, the timing of high-stakes exams (and the characteristics of the built environment in which they are administered) may affect social welfare: either from the standpoint of efficiency or distributional equity. Hot exam days may add noise to the signal-extraction process of high-stakes testing, thus leading to allocative inefficiencies in labor and higher education markets, as documented by Ebenstein et al., 2016. They may also affect distributional equity if individuals taking nationally or internationally standardized exams such as the SAT or Chinese National College Entrance Examinations in a hotter region or in an older school building are placed at a disadvantage relative to their peers in cooler regions or climate-controlled buildings. The latter dimension may be of particular importance in developing countries considering the well documented relationship between hotter climates and lower per capita incomes Acemoglu and Dell, 2010 as well as the strong links between income and air conditioning ownership at the household level Gertler et al., 2016. It may also be relevant in thinking about the persistence of racial achievement gaps in the United States, given the correlation between race, income, and average climate.

Second, this study raises new questions in an old debate regarding geography and economic prosperity Acemoglu et al., 2001, Rodrik et al., 2004. How much of the variation in student achievement across countries (Figure 1) can be explained by the cumulative influence of temperature stress on learning? Important caveats to cross-country comparisons in achievement notwithstanding, is it possible that hotter, poorer countries are subject to more challenging baseline learning conditions due to a combination of hot climate, lack of protective capital, and inflexible institutions?59

Finally, from the perspective of climate policy, this study lends further support to the notion that current social cost of carbon estimates omit important elements of the climate damage function: especially those channels, including reduced labor productivity, that operate through direct heat-stress of the human body Tol, 2009, Burke and Emerick, 2016 Heal and Park, 2016. These findings also support the notion that climate change may affect not only the level of economic activity but overall growth rates Pindyck, 2013, though more research on the impact of cumulative heat exposure on learning is needed.

At the same time, this study also underscores the importance of taking adaptive responses into account when thinking about the realized welfare consequences of climate

---

59 The extent to which cross-country differences in standardized scores such as the PISA assessments is a subject that has received considerable attention. See Woessmann, 2016 for a review of this literature as well as a discussion of how scores are standardized across countries and exam waves.
change, especially when using short-run, weather-based estimates to inform projections about the distant future. In particular, it points to important future research on the interactions between institutions, income, and historical climate and how they might affect the realized welfare impacts of climate change.

References


Pindyck, 2013] Pindyck, R. S. (2013). Climate change policy: What do the models tell us? *Journal of Economic Literature*, 51(3):860–872.


Figure 1: Climate and Standardized PISA Math Achievement Across Countries, controlling for per capita income (2012)

Notes: This figure presents a binned scatterplot of mean PISA math scores and average annual temperature by country in 2012, plotting residual variation after controlling for mean per capita income in 2012. Average annual temperatures are measured over the period 1980-2011. A standard deviation in PISA scores corresponds to approximately 100 points.

Figures and Tables
Figure 2: Short-Run Identifying Variation in Temperature

Notes: This figure illustrates the source of identifying variation for short-run performance impacts of heat exposure. It presents realized exam-time temperatures for (a) all June Regents exams (1999-2011) and (b) for two subsequent days within a Regents exam period – Thursday, June 24th, 2010, and Friday, June 25th, 2010 – inclusive of spatial and temporal temperature corrections. Temperatures measured at the school level, weighted by number of exam observations by date and time.
Figure 3: Cumulative Heat Exposure by School Year

Notes: This figure illustrates year to year variation in cumulative heat exposure during the school year, measured in terms of the number of days with max temperatures above 80°F per school year. Temperature readings are taken from USGS weather stations, one from each of the five boroughs of NYC.
Table 1: Summary Statistics by Ethnicity

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

Notes: This figure presents summary statistics for key outcome variables by demographic sub-group from June 1999 to June 2011, for the 4.5 million exam observations in the study sample. Standard deviations are in parentheses. Regents scaled scores range from 0 to 100. Previous ability is measured in terms of average z-scores from standardized math and verbal assessments in grades 3 through 8.
(a) Residualized variation in test performance

<table>
<thead>
<tr>
<th></th>
<th>(1) Z-score</th>
<th>(2) Z-score</th>
<th>(3) Z-score</th>
<th>(4) Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (F)</td>
<td>-0.00850***</td>
<td>-0.00736***</td>
<td>-0.0102***</td>
<td>-0.0108***</td>
</tr>
<tr>
<td></td>
<td>(0.00231)</td>
<td>(0.00207)</td>
<td>(0.00233)</td>
<td>(0.00226)</td>
</tr>
<tr>
<td>Afternoon</td>
<td>-0.0297*</td>
<td>-0.0334**</td>
<td>-0.0180</td>
<td>-0.0156</td>
</tr>
<tr>
<td></td>
<td>(0.0130)</td>
<td>(0.0119)</td>
<td>(0.0142)</td>
<td>(0.0127)</td>
</tr>
</tbody>
</table>

Fixed Effects

- Student by Year X
- Subject X X X X
- Time of Day, Day of week X X X X
- Student X
- Year X X
- School X
- School by Year X

N 3581933 3581933 3581933 3581933
r2 0.774 0.717 0.252 0.271

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

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

(b) Dependent variable is standardized performance by subject

Figure 4: Short-Run Impacts of Heat Exposure on Exam Performance

Notes: Panel (a) presents a binned scatterplot of residualized exam performance by percentile of the temperature distribution after controlling for school, subject, and year fixed effects. Each dot represents approximately 220,000 exam observations. Panel (b) presents the main regression results. Fixed effects are suppressed in output, and 919,067 singleton observations are dropped. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.
(a) 4-year graduation status and exam-time temperature

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg Exam-Time Temp (°F)</td>
<td>-0.00712***</td>
<td>-0.00758***</td>
<td>-0.00733***</td>
</tr>
<tr>
<td></td>
<td>(0.00173)</td>
<td>(0.00223)</td>
<td>(0.00231)</td>
</tr>
<tr>
<td>Number of June exams</td>
<td>0.193***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00688)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of June exams²</td>
<td>-0.0151***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000809)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of June exams³</td>
<td>0.000312***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000225)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fixed Effects

- School
- Number of June exams
- Cohort

N 515192
r² 0.232

Robust standard errors in parentheses
* p < 0.05, ** p < 0.01, *** p < 0.001

(b) Impacts on graduation status by regression specification

Figure 5: Long-Run Consequences of Short-Run Heat Exposure: Graduation Status

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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z-score</td>
<td>Z-score</td>
<td>Z-score</td>
<td>Z-score</td>
<td>Z-score</td>
</tr>
<tr>
<td>70°F-80°F days</td>
<td>0.000669</td>
<td>0.00506</td>
<td>0.00131</td>
<td>0.00412</td>
</tr>
<tr>
<td></td>
<td>(0.00171)</td>
<td>(0.00322)</td>
<td>(0.00187)</td>
<td>(0.00348)</td>
</tr>
<tr>
<td>80°F-90°F days</td>
<td>-0.0108***</td>
<td>-0.0121***</td>
<td>-0.0114***</td>
<td>-0.00337</td>
</tr>
<tr>
<td></td>
<td>(0.00310)</td>
<td>(0.00329)</td>
<td>(0.00279)</td>
<td>(0.00668)</td>
</tr>
<tr>
<td>90°F+ days</td>
<td>0.0209*</td>
<td>0.00459</td>
<td>0.00804</td>
<td>0.0198</td>
</tr>
<tr>
<td></td>
<td>(0.00937)</td>
<td>(0.0129)</td>
<td>(0.00973)</td>
<td>(0.0182)</td>
</tr>
<tr>
<td>Exam-Time Temp</td>
<td>-0.00580***</td>
<td>-0.00830***</td>
<td>-0.00790***</td>
<td>-0.00824***</td>
</tr>
<tr>
<td></td>
<td>(0.00152)</td>
<td>(0.00172)</td>
<td>(0.00165)</td>
<td>(0.00179)</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses, clustered at the station-by-year level
* p < 0.05, ** p < 0.01, *** p < 0.001

(b) Dependent variable is standardized Regents performance

Figure 6: Cumulative Learning Impacts of Heat Exposure

Notes: Panel (a): binned scatterplot of residualized exam performance at the school level on amount of cumulative heat stress, measured by the number of days per degree bin (days above 80°F on x axis), including controls for exam-day temperature and precipitation, as well as school-, subject-, time of day, day of week, and day of month fixed effects. All regressions in (b) include school, subject, and time of day fixed effects which are suppressed in the output. Robust standard errors are clustered at the borough by year level. Cumulative degree day variables are assigned by closest weather station and summed beginning on the first day of the preceding fall semester up through the first day of June Regents exams that year.
(a) AC status by school

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Central</th>
<th>Any AC</th>
<th>No Central</th>
<th>No AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp (°F)</td>
<td>-0.00613*</td>
<td>-0.00530</td>
<td>-0.00517*</td>
<td>-0.00649*</td>
<td>-0.00621*</td>
</tr>
<tr>
<td></td>
<td>(0.00219)</td>
<td>(0.00293)</td>
<td>(0.00273)</td>
<td>(0.00274)</td>
<td>(0.00261)</td>
</tr>
<tr>
<td>Precip (mm)</td>
<td>0.00183</td>
<td>0.00212</td>
<td>0.00177</td>
<td>0.00219</td>
<td>0.00197</td>
</tr>
<tr>
<td></td>
<td>(0.00128)</td>
<td>(0.00154)</td>
<td>(0.00141)</td>
<td>(0.00156)</td>
<td>(0.00147)</td>
</tr>
<tr>
<td>Afternoon</td>
<td>-0.0417**</td>
<td>-0.0439*</td>
<td>-0.0473**</td>
<td>-0.0338*</td>
<td>-0.0369*</td>
</tr>
<tr>
<td></td>
<td>(0.0149)</td>
<td>(0.0181)</td>
<td>(0.0169)</td>
<td>(0.0171)</td>
<td>(0.0163)</td>
</tr>
<tr>
<td>N</td>
<td>4347718</td>
<td>906451</td>
<td>1019279</td>
<td>1611336</td>
<td>1724670</td>
</tr>
<tr>
<td>r²</td>
<td>0.710</td>
<td>0.720</td>
<td>0.717</td>
<td>0.710</td>
<td>0.709</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses, clustered at the station-by-date-time level

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

(b) Short-Run Impacts by School Air Conditioning Status

Figure 7: Adaptive Responses: Air Conditioning

Notes: Panel (a) provides a map of New York City public schools, with green dots representing schools that had any air conditioning equipment as of 2012, and red dots representing schools that did not, according to the NY School Construction Authority. The dependent variable in (b) is standardized Regents performance, with observations at student, exam, and date-and-time-level. All regressions include student, subject, and year fixed effects, as well as controls for dewpoint, ozone, pm2.5 and precipitation. Fixed effects are suppressed in output, and singleton observations are dropped. Robust standard errors clustered at the borough by date-time level.
(a) All Regents exams in core subjects prior to NYC grading reforms in 2011-2012.

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

Figure 8: Adaptive Responses: Grade Manipulation

Notes: Panel (a) presents a histogram of Regents exam scores from June 1999 to June 2011. A large number of observations bunch at the pass/fail cutoffs, scores of 55 and 65 for local and Regents diploma requirements respectively. Panel (b) presents a binned scatterplot of bunching at the school-subject-date level by quantile of the exam-time temperature distribution, controlling for subject and year fixed effects and daily weather/air quality controls. The bunching estimator is calculated by integrating the distance between predicted and observed score fractions of scores within the manipulable zone. Included in the analysis are all June Regents exams in core subjects between 1999 and 2011. manipulable zone.