

Temperature, Test Scores, and Human Capital Production

Jisung Park
Harvard University*
February 26, 2017

Abstract

How does temperature affect the human capital production process? Evidence from 4.5 million New York City high school exit exams indicates that heat exposure may affect educational performance in both the short and long run. Taking an exam on a 90°F day relative to a 72°F day results in a reduction in exam performance that is equivalent to a quarter of the Black-White achievement gap, and meaningfully affects longer-run educational outcomes as well, leading to a 12.3% higher likelihood of failing a subject exam and a 2.5% lower likelihood of on-time high school graduation. Furthermore, cumulative heat exposure over the course of the preceding school year may reduce the rate of learning as seen in exit exam scores, controlling for the short-run effect of exam day temperature. Teachers try to offset some of the impacts of exam day heat stress by selectively boosting grades for students who experience particularly hot exam sittings, perhaps in response to low levels of classroom air conditioning.

*The author would like to thank Larry Katz, Andrei Shleifer, Robert Stavins, Geoffrey Heal, Raj Chetty, Joe Aldy, Claudia Goldin, Edward Glaeser, Jim Stock, Sendhil Mullainathan, Melissa Dell, Michael Kremer, Josh Goodman, Jonah Rockoff, Jeff Miron, Peter Huybers, Daniel Schrag, Max Auffhammer, Sol Hsiang, Olivier Deschenes and numerous seminar participants at Harvard, Columbia, UC Berkeley, Oxford, IZA, Seoul National, the NYC Department of Health, the NBER Summer Institute and the Bill and Melinda Gates Foundation for valuable comments and feedback. Thanks also to the NYC Department of Education for data access, and to Nicolas Cerkez, Rodrigo Leal, Kevin Eskici, and William Xiao for excellent research assistance. All remaining errors are my own. Funding from the Harvard Environmental Economics Program, the National Science Foundation, and the Harvard Climate Change Solutions Fund is gratefully acknowledged.

1 Introduction

A longstanding literature explores the relationship between schooling inputs, educational achievement, and economic welfare.¹ A wave of recent studies highlights the causal impact of temperature on a range of outcomes including health and labor productivity, suggesting that the physiological and cognitive effects of heat stress may have economically meaningful consequences.² And yet, few studies have documented the role that temperature plays in the human capital production process, especially in school settings.

To assess whether and how heat stress may affect educational outcomes, I use administrative data from the nation’s largest school district: New York City public schools. I focus on three empirical research questions. First, does acute heat stress affect performance on high-stakes exams? That is, do early lab-based findings – wherein cognitive performance declines with elevated temperatures – extend to school contexts, where the economic stakes are higher? Second, do these short-run heat exposures, which presumably do not reduce the stock of human capital *per se*, meaningfully 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, one might expect even short instances of heat stress during an exam to have lasting economic consequences.³ Third, is it possible for cumulative heat exposure to influence the efficacy of learning over time? Whether or not temperature affects the rate of human capital accumulation may have first-order welfare and policy implications, especially given the longstanding relationship between geography and growth (Acemoglu, Johnson, and Robinson, 1999; Gallup, Sachs, and Mellinger, 2000; Dell, Jones, Olken, 2014) as well as the prospect of future climate change (Greenstone et al, 2013; Kahn, 2016).

My research design is based on a simple premise: that short-run variations in temperature are not caused by unobserved determinants of educational performance. This empirical strategy, in conjunction with institutional features requiring NYC students to take a series of high-stakes exams spanning 10-15 days each June, allows for identification of the causal impact of heat stress on performance using within-student variation. Since all students are assigned to test dates and locations without prior knowledge of temperature (and without

¹For a review of the literature on wage returns to human capital, see Card (1999). For some recent examples of the later-life impacts of schooling interventions, see Angrist and Lavy (1999), Chetty et al (2014), or Deming et al (2014).

²See Dell, Jones, and Olken (2014) for an excellent review of this emerging literature. See Barrecca et al (2016) for health impacts, Zivin and Neidell (2014) for labor supply, Sudarshan and Tewari (2013) for labor productivity, and Dell, Jones, and Olken (2012), Hsiang (2010) and Park and Heal (2013) for output impacts of hot weather.

³As has been found to be the case in the context of air pollution (Ebenstein, Lavy and Roth, 2016) and grade manipulation (Dee et al, 2016; Diamond and Persson, 2016). For instance, Ebenstein, Lavy and Roth (2016) find that air pollution exposure during high-stakes exams leads to lower post-secondary schooling attainment and reduced earnings.

the ability to reschedule), temperature on the day of an exam is unlikely to be correlated with student quality. Similarly, year-to-year fluctuations in the incidence of hot days by neighborhood are unlikely to be systematically correlated with student performance when comparing the same schools over time. The richness of the data set, which comprises over 4.5 million individual exam observations from 990 thousand high school students, allows for an in-depth analysis of the potential mechanisms through which temperature may enter into the human capital production function, as well as analyses of how economic agents may adaptively respond.

I find that heat exposure during an exam exerts a causal and economically meaningful impact on educational achievement, even when controlling for individual student ability. For the average student, taking a NY State Regents exam on a hot day leads to -0.22% lower performance per °F above room temperature (72°F), such that a one standard deviation increase in test-time temperature has a negative effect that is 1/8th the size of the Black-White score gap.⁴ 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. These results are consistent with the existing ergonomic literature, as well as recent empirical work on the causal impacts of heat stress in a variety of welfare-relevant settings (Dell, Jones, and Olken, 2012; Zivin and Neidell, 2014; Barecca et al, 2016). The effect sizes are comparable to the impact of temperature on voluntary home math assessments (Zivin, Hsiang, and Neidell, 2015) and the impact of air pollution on high-stakes exam performance (Ebenstein, Lavy and Roth, 2016). These 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, cooling technology adoption is incomplete, and where high-stakes exams pose real hurdles to further schooling.

Perhaps in response, teachers seem to have used their discretion in grading to selectively boost grades around passing thresholds, particularly when students have experienced hot exam sittings. Building on work by Dee et al (2016) who use data from the same district to document grade manipulation by teachers prior to city-wide grading reforms, I estimate the relationship between the extent of 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 between 1999 and 2011, the extent of manipulation varied systematically according to the temperature students experienced during the exam, with hot takes exhibiting approximately 1.5% more bunching behavior per °F. Such “adaptive grading” represents a hitherto undocumented and likely sub-optimal channel of climate adaptation. A possible unintended consequence of eliminating teacher discretion in New York City public schools in 2011 may

⁴Regents exams are required high school exit exams that determine diploma eligibility and can influence college admission.

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.

Looking at longer-run outcomes, I find that acute heat stress during a high-stakes exam reduces the likelihood that a student passes any given subject, which subsequently affects her chances of graduating from high school. Taking an exam on a 90°F day leads to a 10.9% lower likelihood of passing that exam for the median student. This means that a one standard deviation increase in average exam-time temperature over the student’s high school career (up until senior year) reduces the likelihood of graduating on time by roughly 2.5 percentage points. Heat stress also substantially reduces students’ chances of achieving key performance thresholds that are used by local universities in college admissions decisions. 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 Ebenstein, Lavy and Roth (2016), and/or due to dynamic complementarities in the human capital investment process as suggested by Cunha and Heckman (2007) and Diamond and Persson (2016).

Leveraging quasi-random variation in cumulative heat exposure over the course of the school year, I find that repeat heat stress may reduce the rate of learning and human capital accumulation – in addition to and controlling for the short-run impact documented above. Hot days during the preceding school year reduce end-of-year exam performance, though the effects are less precisely estimated given limited spatial temperature variation and a panel length of 13 years. A one standard deviation increase in the number of days above 80°F reduces Regents performance by approximately 4% of a standard deviation. The effect is similar in size 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 the same NYC students by approximately \$14,800 per student, or \$445,000 per classroom (Chetty et al, 2014) – though there are many reasons why the later-life impacts of better teaching may be different from those associated with fewer heat-related disruptions. While these estimates are measured with considerable error, they are consistent with a model of human capital accumulation in which heat stress during class time reduces effective pedagogical engagement by students and teachers, and suggests promising avenues for further research, particularly in developing countries.

Finally, I find evidence suggesting that structural adaptation to heat stress in New York City public schools is incomplete, and that existing air conditioning units may be only partially effective. Building-level AC installation data suggests that less than 62% of schools had any form of air conditioning as of 2012. Of those that do, over 40% were deemed to have defective components by independent building inspectors (BCAS, 2012). Comparing schools that do and do not have some form of AC, I find limited evidence for protective effects of

being in a school that has AC. Performance impacts of heat stress in schools with central air conditioning are smaller than those in schools without air conditioning equipment, but not significantly so. This may be in part due to data constraints – AC installation status may be a noisy predictor of actual AC utilization at the classroom level – but is also consistent with previous findings which suggest 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; Mendell and Heath, 2005).⁵ These results suggest that more research is needed in ascertaining the true cost-benefit of installing or improving AC equipment as an input to school production.

This paper contributes to a growing literature exploring the causal impact of climate on economic outcomes, including impacts of temperature shocks on human health (Barecca et al, 2016), labor productivity and supply (Zivin and Neidell, 2014; Cachon et al, 2013), violent crime (Anderson, 1987; Hsiang et al, 2013), and local economic output (Hsiang, 2010; Dell, Jones, and Olken, 2012; Park and Heal, 2013), as well as the nascent empirical literature on climate adaptation (Mendelsohn, 2000; Deschenes and Greenstone, 2011; Burke and Emerick, 2015).

It also contributes to a long literature documenting the efficacy and welfare implications of various inputs to schooling, including teacher value added (Chetty et al, 2014), reductions in class size (Angrist and Lavy, 1999; Chetty et al, 2014), and school choice and desegregation (Sandstrom and Bergstrom, 2005; Deming et al, 2014).

While more careful research is needed to verify whether repeated heat exposure reduces the rate of human capital accumulation in the long run, the findings presented here suggest that the interplay between climate and human capital could be an additional contributing factor to the long-debated relationship between hotter climates and slower growth (Mankiw, Romer, and Weil, 1992; Gallup, Sachs, and Mellinger, 1999; Acemoglu, Johnson, and Robinson, 2000; Rodrik et al, 2004; Dell, Jones, and Olken, 2012; Burke et al, 2015). It is worth noting that while the average New Yorker is exposed to approximately 11 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. To the extent that future climate change will likely 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.

The rest of this paper is organized as follows. Section 2 provides an overview of the relevant ergonomic and economic literature on heat and human welfare. Section 3 presents a simple conceptual framework. Section 4 describes the data and institutional context and presents key summary statistics. Section 5 presents the main results for the short-run impact

⁵Most NYC public schools are located in very old buildings. According to NYC Open Data, the median school building in NYC was constructed in 1932.

of heat stress on exam performance. Section 6 explores whether these short-run shocks have economically meaningful consequences. Section 7 explores the impact of cumulative heat exposure on learning. Section 8 discusses implications and concludes.

2 Heat Stress and Human Welfare

Three stylized facts from the existing scientific literature are of particular relevance in thinking about the impact of temperature on human capital production: first, that heat stress directly affects physiology in ways that can be detrimental to cognitive performance; second, that 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, that the inverted U-shaped relationship between temperature and performance documented in the lab has been confirmed in the context of a variety of welfare-relevant outcomes including health and labor productivity, but not yet for educational performance in high-stakes school settings.

2.1 The Physiology of Heat Stress

Heat stress has well-known physiological consequences. At extreme levels, heat exposure can be deadly, as the body becomes dehydrated and hyperthermia begins to cause dizziness, muscle cramps, and fever, eventually leading to acute cardiovascular, respiratory, and cerebrovascular reactions. Exposure to heat is also associated with increases in blood viscosity and blood cholesterol levels, which can eventually cause increased morbidity in the form of heat exhaustion and stroke, the latter most acutely for the elderly.

Even at relatively mild temperatures, heat can affect human behavior through its subtle effects on physiology and psychology. The human brain produces a disproportionate amount of body heat – by some estimates originating up to 20% of total body heat despite comprising 2% of total mass (Raichle and Mintun, 2006) – and has been shown to experience reduced neural processing speed and impaired working memory when brain temperature is elevated (Hocking et al, 2001).⁶

Not surprisingly then, core body temperature can reduce cognitive and physical function, as has been shown in a wide range of lab and field experiments discussed below.

⁶Heat stress has also been shown to increase negative affect and reduce concentration, which may further diminish cognitive and/or physical performance (Anderson and Anderson, 1984). For instance, Kenrick and MacFarlane (1986) find a strong positive correlation between higher temperature and aggressive horn honking frequency and duration in Phoenix, with significantly stronger effects for subjects without air-conditioned cars.

2.2 A Revealed Preference for Avoiding Temperature Extremes

All else being equal, individuals prefer not to be exposed to extreme temperatures. Revealed preference techniques such as hedonic price estimation have long confirmed the general intuition that most experience non-trivial direct disutility from being exposed to temperature extremes, and are willing to pay non-trivial amounts for such climate amenities when markets allow.⁷

The willingness to avoid acute heat stress is perhaps most directly evident in energy markets. On the intensive margin, annual expenditures on electricity for air conditioning are highly sensitive to hot days (Greenstone and Deschenes, 2013), as well as to average climates (Mansur, Mendelsohn, and Morrison, 2008). On the extensive margin, and conditional on sufficient income levels, residential air conditioning ownership is closely linked to average climate (Davis and Gertler, 2015).⁸

The preference for avoiding heat exposure is evidenced also by data on time-use decisions of Americans. Using ATUS data, Zivin and Neidell (2014) show that individuals working in highly exposed industries such as construction or transportation report spending substantially less time (up to 18 percent fewer hours per day) working outdoors on days with maximum temperatures above 90°F.

Taken together, these studies suggest that individuals experience direct disutility from heat stress, may experience increased marginal disutility of effort when temperatures are elevated, and are willing to pay non-trivial amounts to avoid this non-pecuniary impact.⁹

2.3 Temperature and Task Performance

Beginning with the early experiments of Mackworth (1946), wherein British naval officers were required to perform physical and mental tasks such as deciphering Morse Code under varying degrees of heat stress, a long series of lab experiments have subsequently documented a single-peaked relationship between temperature and human task performance in highly controlled environments (Grether, 1973).¹⁰ Whether in the context of guiding a steering wheel, running

⁷For instance, see Roback (1982) or Sinha and Cropper (2015).

⁸Even in the United States, where average air conditioning penetration is above 80%, households in warmer areas exhibit substantially higher rates of ownership – and tend to invest in more expensive central AC – than those in cooler climates (Energy Information Administration, 2009). For instance, over 85% of households in the US South had central AC as of 2009, compared to 44% of households in the Northeast and 76% in the Midwest. The fact that air conditioning penetration varies substantially across countries (Park and Heal, 2013) and across households within countries according to income level suggests either that the marginal utility of climate control is dependent on overall income and/or that many poorer households face substantial liquidity constraints in purchasing cooling appliances, as suggested by Davis and Gertler (2015).

⁹An additional motivation of this study is to assess whether there are indirect pecuniary impacts of heat stress which operate through the channel of human capital accumulation.

¹⁰See Seppanen, Fisk, and Lei (2006) for a meta-review of the literature on temperature and task performance.

on a treadmill, or performing arithmetic, heat stress has been shown to reduce accuracy and endurance substantially on a wide range of physical and cognitive tasks.¹¹

A more recent econometric literature documents causal impacts of heat stress on a variety of welfare-relevant outcomes *in situ*. Leveraging quasi-experimental variation in local weather, these studies find clear impacts of hot days on mortality (Deschenes and Greenstone, 2011; Barecca et al, 2016), labor supply (Zivin and Neidell, 2014), labor productivity (Cachon et al, 2013), violent crime (Anderson, 1987; Hsiang et al, 2013), and even local output and GDP (Dell, Jones, and Olken, 2012; Park and Heal, 2013; Deryugina and Hsiang, 2015). There is also evidence suggestive of long-lasting welfare impacts of heat stress in-utero and in early childhood, including impacts of hot days during pre- and early-natal periods on later-life earnings (Isen, Rossin-Slater, Walker, 2015).

2.4 Heat Stress and Human Capital

Despite the emerging literature on the economics of extreme heat stress, the role that temperature plays in education and human capital development remains poorly understood.¹²

There is some early evidence that the lab-based findings of adverse cognitive impacts from heat stress also occur in home environments. Zivin, Hsiang, Neidell (2015) use NLSY survey data which includes short, voluntary assessments that were administered to several thousand students at home, and find evidence for contemporaneous impacts of hot days on math performance but not verbal performance.

Empirical evidence from school settings – where students spend the majority of pedagogically engaged hours and where potentially welfare-enhancing public policy interventions might take place most directly – is limited, however, apart from a few qualitative case studies which do not permit causal identification (e.g. Duran-Narucki, 2008) or early classroom experiments (Schoer and Shaffran, 1973).¹³ In contrast, there are a number of studies exploring the impact of air pollution on student outcomes (Currie et al, 2009; Ebenstein, Lavy and

¹¹There is also experimental evidence suggesting cold effects human cognition and task productivity as well. In general, the evidence is stronger and more consistent for adverse impacts of heat stress, especially when it comes to impacts in situ, where heating and cooling technologies may be present.

¹²This is not for lack of anecdotal evidence, or complaint on part of students, parents, and teachers. For instance, in 2015, the New York Times published an article decrying the lack of adequate air conditioning in its public schools, suggesting that heat stress in classrooms were reducing student engagement and impeding learning. Mayor Bloomberg’s response to media critiques on this issue is suggestive of possible financial, institutional, and cultural constraints to full adaptation: “Life is full of challenges, and we don’t get everything we want. We can’t afford everything we want. I suspect that if you talk to everyone in this room, not one of them went to a school where they had air conditioning.” See New York Times, 2015: <http://mobile.nytimes.com/2015/06/24/nyregion/new-yorks-public-school-students-sweat-out-the-end-of-the-semester.html>.

¹³To the best of my knowledge only one study uses an experimental or quasi-experimental research design to assess the impact of temperature on student performance in the classroom. Schoer and Shaffran (1973) assess the performance of students in a pair of classrooms set up as a temporary laboratory, with one classroom cooled and one not, and found higher performance in cooled environments relative to hot ones.

Roth, 2016; Roth, 2016).¹⁴ These studies consistently find large impacts on absenteeism and exam performance. In the case of pollution during high-stakes exams in Israel (Ebenstein, Lavy and Roth, 2016), there is evidence for persistent and economically meaningful impacts that extend well beyond formal schooling.

This study seeks to expand on the nascent literature exploring whether and how temperature affects the human capital production process by using evidence from high-stakes exams in public schools.

3 Conceptual Framework

Motivated by the evidence linking temperature and human task performance presented above, this section provides a simple conceptual model which illustrates the mechanisms through which heat stress may affect the human capital production process.

3.1 Definitions and Setup

Define human capital, h_i , as a measure of skills or knowledge accumulated through schooling. Let e_i represent composite schooling investment and comprise all pecuniary costs of schooling, including schooling time and effort investment.¹⁵ The pecuniary returns to schooling investment are summarized in terms of labor market wage returns to human capital or skill: $w \cdot h_i(e_i)$, where w denotes wages.

In the classical Mincerian framework and derivative models that have followed, optimal schooling investment, e_i^* , depends on student characteristics such as income, ability, or opportunity costs/discount rates, which in turn determine the relative costs and benefits to incremental investments in schooling.¹⁶ Here, we are interested in understanding the consequences of heat exposure while a student is in school, allowing for optimizing responses. In this simple setup, students determine how much time and effort to invest in schooling based on a utility function that is increasing in consumption and decreasing in effort.

Let T represent the extent of temperature elevation above the optimal zone, and define $a(T) = (1 - \beta_T T)$ as a measure of the effectiveness of any given unit of schooling effort or time,

¹⁴Two existing studies assess the impact of weather variation on student performance. 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.

¹⁵ e_i may also include direct costs of schooling such as the cost of books and tuition.

¹⁶For instance, lower ability individuals may suffer greater disutility from being in school for an incremental year (more negative U_e), or may experience lower pecuniary returns from an incremental unit of effort (low $\delta h_i / \delta e_i$), leading to a lower optimal level of schooling attainment given the opportunity costs.

such that $h_i = h_i(e_i, a(T)) = (1 - \beta_T T) \cdot e_i$.¹⁷ As suggested by the experimental literature described above, let us assume that $a'(T) = -\beta_T < 0$: that is, cognitive effectiveness is declining in the extent of heat stress (i.e. a single-peaked function of ambient temperature).¹⁸

Similarly, one might expect any given exam score to be influenced by this short-run cognitive impact of heat stress if temperature in the classroom is elevated during an exam. Let $s_{it}(h_{it}) = (1 - \beta_T T_t) \cdot h_{it} + \epsilon_t$ denote an exam score associated with student i who has accumulated human capital of level h by the time of exam t , where $\epsilon_t \sim N(0, \sigma_\epsilon)$ is white noise capturing the fact that, with or without temperature-stress, most realized exam scores provide an imperfect signal of underlying knowledge, and may be influenced by other idiosyncratic factors.

The student's utility function can be represented as:

$$U_i = U_i(C_i, e_i, T) = U_i(w(1 - \beta_T T) \cdot e_i, e_i, T) \quad (1)$$

where

$$U_c > 0; \quad U_e < 0; \quad \text{and} \quad U_T < 0. \quad (2)$$

T is an exogenously determined parameter depending on the local climate and its manifestation as weather on any given school day or year. The student optimally chooses e_i subject to the consumption budget constraint: $C_i = w(1 - \beta_T T) \cdot e_i$ and a given climate or temperature.¹⁹

3.2 Adaptive Responses to Heat Exposure

In response to heat stress – especially prolonged or persistent heat stress – individuals can engage in a wide range of adaptive responses. One may in principle reschedule strenuous activities during cooler times of day, as many in hotter climates routinely do as a matter of cultural norm (e.g. the Spanish Siesta). When resources allow, they may install and utilize cooling technologies such as ceiling fans or air conditioning.

In school settings, however, it is unclear how much adaptive behavior is feasible given common constraints on student activities. A typical secondary school student cannot install an air conditioner in her classroom, even if she can afford it financially. Nor, in most cases, can

¹⁷For instance, temperatures of 90°F and 80°F will correspond to $T(90) > T(80) > 0$, whereas 72°F, often considered to be the optimal room temperature, corresponds to $T(72) = 0$.

¹⁸Apart from $a'(T) < 0$, we can remain agnostic as to the specific functional form of $a(T)$, though it is likely the case that $a''(T) < 0$, given the fact that at some point heat stress becomes deadly. Note that it is possible for the realized effectiveness of schooling effort to be adversely affected by temperature because of temperature's impact on teacher cognition or effort, as well as other relevant actors (e.g. parents, school administrators).

¹⁹I abstract away from the temporal distinction between short-run weather and long-run climate for simplicity, and leave an explicit dynamic treatment, where agents' knowledge (or lack of knowledge) regarding shifts in future climate distributions may be relevant, as suggested by Kahn (2016), for future work.

her parents, however active in their parent-teacher engagement they may be. In fact, given the complex capital budgeting procedures in most US public school districts, it is possible that teaching staff or school administrators also cannot install air conditioning equipment at will, even if they divine a clear preference or need on part of their students due to perceived effects on learning.²⁰

At the same time, some margins of adaptation may be unique to school environments. To the extent that teachers and administrators have discretion in grading exams or applying institutional rules regarding graduation, it is possible that they can buffer some of the random and transient shocks that affect short run performance but do not reduce human capital: a form of second-best response given institutional constraints.²¹ Teaching staff may in principle reschedule classes or exams in anticipation of – or in response to – acute weather shocks.

Suppose students can engage in avoidance behaviors which reduce the negative impact of heat while incurring some pecuniary cost. Let us denote this investment k_i and define it such that $h_i = (1 - (\frac{\beta_T}{1 + k_i})T) \cdot e_i$, and $U_k < 0$. For instance, suppose students are able to respond to heat stress by purchasing a cool beverage, installing a desk fan, or initiating more structural responses by lobbying teachers and administrators to open classroom windows or turn on or install air conditioners. Let us assume also that heat adversely affects cognition $a'(T) \leq 0$, has a weakly negative direct effect on utility $U_T \leq 0$, or increases the marginal cost of additional effort $U_{eT} \leq 0$, all of which are suggested by the existing literature.

The student’s value function can be expressed as:

$$V_i(e_i^*, T, k_i^*) = \max_{e_i, k_i} U_i(w(1 - \frac{\beta_T T}{1 + k_i}) \cdot e_i, e_i, T, k_i) \quad (3)$$

where e_i^* and k_i^* denote optimal investments in schooling effort and adaptive capital respectively.

Intuitively, students trade off pecuniary costs of schooling with pecuniary and non-pecuniary benefits, while investing in adaptive capital such that the marginal benefit in terms of increased skill creation equals the added cost, which we summarize simply by U_k . The magnitude of U_k will depend on institutional flexibility, available technologies, and/or the respon-

²⁰For instance, in the case of New York City public schools, air conditioners must meet efficiency standards and be obtained from and installed by a specific vendor chosen by the city, in addition to receiving city approval with regard to a variety of safety regulations, contractual obligations and energy considerations. In some cases, school “sustainability” policies prohibit administrators from investing in new infrastructure unless it can be demonstrated that it has a net neutral impact on carbon emissions, a bar that new air conditioning cannot clear unless electricity is obtained completely from renewable sources.

²¹Dee et al, (2016) document evidence for substantial grade manipulation behavior on part of teachers in NYC public schools on NY State Regents exams – the primary performance metrics used in this analysis. I explore the possibility that positive grade manipulation by NYC teachers – selectively applied for exams that were subject to excess heat exposure – may have acted as a buffer against the negative impacts of heat stress in section 5.

siveness or punitiveness of parents, teachers, and school administrators.²²

Note that, unless adaptive technologies involve zero costs, $U_k = 0$, the existence of adaptive margins does not imply that the welfare, exam performance, or human capital impacts of heat stress in school will necessarily be eliminated or even minimized; i.e. the availability of adaptive technologies in principle does not imply their full utilization in response to environmental stressors in practice.

Some students and schools may be more able to invest in adaptive responses than others, due, for instance, to different income endowments. These and other reasons discussed in the online appendix suggest that, similarly to the case of traditional environmental pollutants such as air quality or toxic chemicals, the adverse welfare impacts of heat stress may accrue disproportionately to the poor (Currie, 2009; Aizer et al, 2015).²³

3.3 Empirical Predictions

The main empirical predictions are as follows:

1. We expect acute heat stress to reduce exam performance, $\frac{\Delta s_{it}}{\Delta T_{it}} < 0$, if any of (A) direct flow utility, (B) marginal cost of effort, or (C) cognitive performance are adversely affected by temperature, even if effective adaptive technologies and techniques are available.
2. Repeated heat stress may reduce human capital accumulation and educational attainment, $\frac{\Delta h_{it}}{\Sigma_0^{t-1} \Delta T_{it}} \leq 0$.²⁴
3. We expect agents (e.g., students, teachers, and/or parents) to adapt along the most cost-effective of available margins, but only partially given non-zero costs of adaptation.

²²This cost may come in the form of financial costs if students are able to invest in their own cooling equipment at school, on the way to and from school (e.g. taking a taxi as opposed to walking), or at home during homework hours. It may alternatively come in the form of political capital or time/effort costs incurred in appealing to parents, teachers, or school administrators to lower the thermostat if air conditioning equipment is present or install air conditioning if equipment is not present. Even if students and teachers are able to reschedule classes or exams to dates and times that are not as hot, there may still be some cost associated with coordinating the makeup session or engaging in pedagogy out of original sequence.

²³In the language of the model presented above: to the extent that the relative magnitudes of U_k and U_c depend on income endowments, we would expect students from disadvantaged backgrounds to invest in less “optimal” k .

²⁴This may be through a variety of channels including (possibly) reduced effort on part of students in response to heat stress or institutional rigidities that permit random score shocks to affect subsequent investment in schooling time or effort.

4 Institutional Context, Data, and Summary Statistics

4.1 New York City Public Schools

The New York City public school system is the largest in the United States, with over 1 million students across the five boroughs. 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.²⁵

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.

4.2 New York State Regents Exams

Each June, all students in the state of New York take a series of standardized 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 NYC student. Public school 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.²⁶ 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-determined two-week window in mid-to-late June each year.²⁷ The test dates, times, and locations for each

²⁵Approximately 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.

²⁶The 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.

²⁷For any given student, exam takes are spread out across multiple days and years though, in effect, most exams are taken junior and senior year. Apart from the fact that most students take English their junior year, and Living Environment and Global History prior to other “advanced” sciences and US History respectively,

of these Regents exams are determined over a year in advance by the NY State education authority (NYSED), and synchronized across schools in the NYC public school system to prevent cheating.²⁸ Each exam is approximately 3 hours long, with morning and afternoon sessions each day, and are taken at the student's home school.²⁹

All Regents 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.³⁰ 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 NYC schools. As has been documented by Dee et al (2016), a substantial portion (approximately 6%) of NYC Regents exams featured bunching at passing cutoffs, clear evidence for discretionary grade manipulation by teachers. I document this manipulation as well and describe the ways in which it affects this analysis in further detail below.

In summary, using scores from NY State 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. 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 to reflect cumulative impacts of heat stress that may have accrued over the course of the school year. Third, they are taken by a diverse

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.

²⁸The exact dates and ordering of subjects within testing period vary from year to year, allowing for additional identification of possible temperature impacts using the interaction between daily temperature and afternoon/morning status, as reported below.

²⁹Throughout 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.

³⁰In principle, they 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 (Tan and Michel, 2011). The primary identification of short-run impacts include year fixed effects, and thus do not rely on this cross-year comparability.

mix of students, as opposed to by high- (or low-) performing subgroups alone, more likely supporting out-of-sample validity. 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.

4.3 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 2014.³¹

While the data set is incredibly rich, exam dates are not provided in the student-level data. As such, I obtain exam dates and times for each of the 120 main Regents exam sessions that were administered between June 1998 and June 2014 from publicly available exam schedules. These archived schedules provide the date and time of each NY Regents exam taken by NYC public school students over the past two decades (sample provided in the online appendix). Due to inconsistencies in the way exam subjects and terms are coded in the student-level data during later years (arising from changes in grading regimes), I drop exam records for years 2012 through 2014 and use only the records of exams taken in the years 1999 to 2011.³²

4.4 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.

In order to best approximate ambient temperatures experienced by students during their exams, which are taken from 9:15am to 12:15pm and 1:15pm to 4:15pm for morning and

³¹I 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.

³²The exact matching process, in addition to the rationale for limiting the sample, is described in the online appendix.

afternoon sessions respectively, I generate predicted outdoor test-time temperatures by fitting a fourth-order polynomial on observed daily min and max temperature data. Doing this for all June days over the sample period allows me to impute AM and PM temperatures by station on exam days. To account for possible spatial heterogeneity in experienced temperatures due to urban heat island effects, I assign spatial correction factors generated by satellite re-analysis data, which provides 30m by 30m resolution temperature readings for a representative summer day in the New York City (Rosenzweig et al, 2006). These variables are matched geographically using street addresses for 890 school buildings in my sample. The results are robust to using the raw (uncorrected) station readings as well as the spatially and temporally corrected temperature data.³³ The results reported below are also robust to the inclusion of controls for daily air quality, including particulate matter (pm2.5) and ozone, which I take from EPA monitor data from Manhattan.

4.5 School Air Conditioning Information

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

In 2012, a series of Building Condition Assessment Surveys (BCAS) were carried out by independent contractors who recorded detailed information about each school building’s mechanical and electrical systems. Using these records, I am able to match information on air conditioning availability in 2012 for 644 of the 890 middle and high school buildings in the study sample. 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 in 2012 as an admittedly rough proxy for the true variable of interest, which is effective AC utilization at the classroom-by-date/year level.

4.6 Summary Statistics

The final working dataset consists of 4,509,102 exam records for 999,582 students. The sample comprises data from 91 different exam sessions pertaining to the core Regents subjects (11 in total) over the 13 year period spanning the 1998-1999 to 2010-2011 school years.

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.

Tables 1 and 2 present 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

³³Additional details regarding spatial and temporal corrections to the weather data are provided in the online appendix.

well as student type. African American and Hispanic students, with average scores of 61.2 and 61.5, tend to perform substantially worse than Whites and Asians, who average 72.9 and 74.7 respectively.³⁴ Girls tend to perform slightly better than boys, as do students who are not eligible for federally subsidized school lunches (higher SES). NYC students tend to score consistently higher on some subjects relative to others – for instance, the average score on Earth Science, at 62.6 over the study period, is considerably lower than that for US History, at 67.6 – motivating a preferred regression specification that includes subject fixed effects.

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

Cumulative heat exposure during the school year can be substantial 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 5.6.³⁵ Most of these days occur during the months of September, October, and June.³⁶

Figure 4 provides a map of the schools in NYC, coded by air conditioning status. According to the available data, 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. 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.

³⁴The within-school differences are, not surprisingly, smaller. Blacks and Hispanics performed on average 4.9 and 4.4 points below Whites respectively.

³⁵The average number of days during the regular school year with temperatures above 90°F is 2.5, with a standard deviation of 1.3. Summer school students are, on average, subject to an additional 9 days above 90°F.

³⁶Despite 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).

Empirical Strategy and Primary Results

The following sections present the empirical strategy and results. First, I describe the strategy for identifying causal impacts of acute heat exposure on contemporaneous student performance and present the results from this “short-run.” I then present an analysis of the potential long-run consequences of these short-run impacts, focusing on pass and proficiency rates and graduation status. Finally, I present suggestive evidence for the long-run impact of repeated heat exposure on the rate of learning, using a measure of cumulative heat stress during the school year and end-of-year exam performance as a measure of learning.

5 Does Heat Stress Affect Exam Performance?

5.1 The Impact of Exam-Time Temperature on Student Performance

Figure 5 presents a binned scatterplot that motivates this analysis. It shows the relationship between scaled exam score (0-100) 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. Figure 6 presents a similar binned scatterplot that uses standardized Regents performance as the dependent variable, standardized by subject over the study period (1999-2011). Both plots strongly suggest that exams taken on hot days exhibit lower scores.

To 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.

It is possible in principle that hot temperature and unobserved determinants of student performance are correlated in the cross-section. This might be the case if low- (high-) performing schools tend to be located in areas that are more likely to experience greater (less) heat stress on any given exam day, due, for instance, to urban heat islands and unobserved infrastructure quality variation. Using location fixed effects (e.g. school) solves potential biases arising from omitted variables in the cross section.

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 as well. 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 (see 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.

To address these concerns, I 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, though, as reported in Table 3 below, the main effect is robust to variations that replace student-by-year fixed effects with school-by-year, student and year, and school and year fixed effects as well:

$$Y_{ijsty} = \gamma_{iy} + \eta_s + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \beta_3 Time_{sty} + DOW_{sty} \beta_4 + \epsilon_{ijsty} \quad (4)$$

Here, Y_{ijsty} denotes standardized exam performance (z-scores, standardized by subject) for student i taking subject exam s in school j on date t in year y . The terms γ_{iy} and η_s denote student-by-year and subject fixed effects respectively. T_{jsty} is the outdoor temperature in the vicinity of school j during the exam (subject s on date t , year y). X_{jsty} is a school-by-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 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.

To the extent that temperature variation within student-month-year cells are uncorrelated with unobserved factors influencing test performance, one would expect the coefficient β_1 to provide an unbiased estimate of the causal impact of temperature on exam performance (β_T from the model presented in section 3), subject to attenuation bias due to measurement error in weather variables, as well as downward bias from upward grade manipulation.³⁷

Table 3 presents the results from running variations of equation (4), for the subset of students who take at least 2 exams in any given year. (Table 4 presents the results using un-standardized scale scores, 0-100, as the dependent variable). As suggested by the first row of columns (1)-(4), exam-time heat stress exerts a significant causal impact on student performance. 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. The estimates are robust to allowing for arbitrary autocorrelation

³⁷It is worth noting that it is possible in principle 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 β_1 may actually be biased upwards.

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 Regents scale scores by approximately -0.008 standard deviations ($se=0.002$) per $^{\circ}F$.³⁸ 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.³⁹

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 to Zivin, Hsiang, and Neidell (2015), who find a $90^{\circ}F$ day to reduce NLSY math assessment scores by approximately -0.12 standard deviations. Both are slightly smaller than the effect sizes found in laboratory studies, which, according to a recent meta-review clustered around -0.6% per $^{\circ}F$ (Seppanen, Fisk, and Lei, 2006). The effect documented here is also similar in magnitude to the effects on Israeli high school exit exams of a standard deviation increase in pm2.5 and CO pollution found by Ebenstein, Lavy and Roth, (2016).⁴⁰

The results are robust to a model that replaces student fixed effects with student-by-year fixed effects, or a model with school-by-year fixed effects and a full suite of observable demographic controls including ethnicity, gender, federally subsidized school lunch eligibility, and previous ability (in the form of combined ELA and Math z-scores). In this case, students who move schools are assigned the modal school ID – that is, the school in which they spend the most years.⁴¹ These robustness checks 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 across the board.

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' desire or ability to maintain focus or concentration during a three-hour exam.

³⁸In 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

³⁹Precipitation 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.

⁴⁰I 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 Ebenstein, Lavy and Roth 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.

⁴¹If students attend more than one school for an equal amount of time, I assign the last school in which she was enrolled and took a Regents exam.

5.2 Heterogeneity by Demographic Sub-Group

Table 5 presents the results from running the a version of the short-run regression with interaction terms by demographic sub-group: specifically dummy variables denoting race, gender, federally subsidized school lunch eligibility (SES), and previous achievement as measured by 3rd through 8th grade standardized ELA and Math performance. The estimates suggest that Asians are relatively unaffected by exam-day heat stress, and that Blacks and Hispanics may be more susceptible to short-run performance impacts than Whites. Boys seem to respond better than girls, while there seems to be no meaningful differences between students who do and do not qualify for federally subsidized school lunch. Surprisingly, the performance of students in the bottom quintile of the previous achievement distribution appear to be *less* sensitive to heat stress compared to performance of top quintile students, due perhaps to grade manipulation by teachers, which is relevant primarily for lower-performing students, as described in greater detail below.

5.3 Adaptive Responses

If heat stress affects student performance, we would expect students, parents, and teachers to respond to mitigate this impact, presumably along the most cost-effective margins. It is worth noting, however, that adaptation to climatic stressors often takes place in a constrained context, for instance due to pre-existing built infrastructure (Hallegatte, 2009), suggesting that adaptive responses to temperature stress will likely be sub-optimal or second-best at least in the short run. This section explores two types of adaptive responses: air conditioning and teacher responses.

5.3.1 Air Conditioning

Central air conditioning as part of a well-designed combined HVAC (heating, ventilation, and air conditioning) system would seem to be a first-best adaptation strategy, given ergonomic assessments and existing evidence on the effectiveness of air conditioning at mitigating adverse health impacts from heat stress. As noted above, however, less than 62% of NYC public schools reported having some form of AC equipment as of 2012. This, in principle, provides an opportunity to compare the impact of heat stress on high-stakes exams across schools with and without AC units.

I estimate equation 4 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 6. 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. Point estimates are 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 small, however, suggesting either that existing AC units have an only mildly protective effect, or that AC status is measured with substantial noise. On the one hand, AC should mitigate some of the adverse temperature-related impacts on cognition, if it successfully maintains cooler classroom temperatures on hot days. On the other hand, some 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 (Niu, 2004).

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.

5.3.2 Teacher Responses

Using a similar dataset from 2003 to 2012, 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 NYC schools system.

While the authors suggest the most plausible explanation to be the goodwill of teachers who seek to offset the impact of “a bad test day”, the factors that may give rise to such bad test days are left as something of a mystery: *“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).”*

Could discretionary grade manipulation have been a response to perceived performance impacts of heat stress – a form of second-best adaptation given the institutional constraints imposed by the existing high-stakes exam regime?

Given that Regents exams are taken in home schools, it seems possible that teachers may be able to observe or at least intuit the disruptive impacts of elevated classroom temperatures on test day. If they 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 constraints.

5.3.3 Estimating Adaptive Grading: Bunching at Score Thresholds

Figure 7 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.⁴² Indeed, running equation 4 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 estimate for the impact of temperature is substantially reduced and no longer significant: β_T equals -0.0007 (se=0.0024) as opposed to -0.0082 (se=0.0021) in the full sample.

To assess the potential presence and magnitude of “adaptive grading”, wherein the extent of manipulation is a function of temperature during exams, I first estimate a version of Dee et al’s bunching estimator by school, subject, month, and year (in effect, by school, subject, and exam take, which is the level of 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

⁴²The 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 (and Dee et al’s analysis) is that the majority of grade manipulation seems to be positive in direction, making this unlikely in practice.

near the proficiency cutoffs with a set of indicator variables, using the following regression:

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

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

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.

In 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.⁴³

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. For 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

⁴³As 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.

and observed fractions of test results to the left of the cutoff (below the predicted curve):

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

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

5.3.4 Adaptive Grading after Hot Exam Takes

Figure 8 presents a binned scatterplot of the bunching estimator and exam-time temperature by subject-month-year-school cell. It suggests a clear positive relationship between the degree of grade manipulation and the ambient temperature during the exam being graded.⁴⁴

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

$$\zeta_{smj} = \delta_0 + \delta_1 T_{smj} + \delta_2 X_{smj} + \chi_j + \eta_s + \delta_3 Year_y + \delta_4 Year_y^2 + \delta_5 Year_y^3 + \epsilon_{smj} \quad (7)$$

where T_{smj} denotes temperature, X_{smj} denotes precipitation and humidity, χ_j , η_s , and θ_y denote school, subject, and year fixed effects respectively, and $Year_y \dots Year_y^3$ denotes a cubic time trend in scores. The parameter of interest is δ_1 , which represents the increase in grade manipulation due to exam-time temperature.⁴⁵

According to the estimates presented in Table 7, 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.

While these results are highly suggestive of adaptive grading, it is not possible to infer teachers' intentions based on these results alone. It could be the case that teachers have an intuitive sense of whether a particular student scored below his or her "true ability",

⁴⁴Recall 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.

⁴⁵It 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.

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. It is first worth noting that, 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 9 presents a similar binned percentile plot, but 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). The positive association between bunching estimate and temperature remains, suggesting that teachers were motivated to boost students' grades more often when students experienced hot testing environments.

6 Does Short-Term Heat Stress Affect Longer-Term Schooling Outcomes?

Heat exposure during an exam, while reducing cognitive ability or concentration temporarily, presumably does not reduce the stock of knowledge or human capital *per se*, at least not immediately through the physiological impact of heat stress itself.⁴⁶ In the language of the model in section 3, heat stress during any given exam sitting may adversely affect a student's test score, $\frac{\Delta s_{it}}{\Delta T_t} < 0$, but the lower score in this case is not reflective of a reduction in true underlying human capital at that point in time h_{it} , much less of permanently reduced human capital: $\Delta s_{it} \neq \Delta h_{it}$.

In an idealized, friction-less world with fully flexible institutions, unlucky students who fail due to a hot exam sitting would immediately retake the subject 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 achievement or human capital attainment.

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

⁴⁶In 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.

students to attend remedial courses, it is possible that even short-run heat exposure can have ripple effects on long-run educational attainment.⁴⁷ 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 studies have found evidence for long-run effects of temporary score shocks in the context of teacher manipulation (Dee et al, 2016; Diamond and Persson, 2016) and air pollution (Ebenstein, Lavy and Roth, 2016), with as yet inconclusive evidence regarding the specific mechanisms by which these may occur. Dee et al (2016) find substantial impacts of upward score manipulations on graduation status, especially for students who scored in the manipulable zone.⁴⁸ 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, Lavy and Roth (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. All suggest that random shocks to Regents exam performance due to heat stress may also have long-run impacts on educational attainment and other welfare-relevant later-life outcomes.

6.1 Pass Rates and College Proficiency

If heat stress during Regents exams 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.

Students must score a 65 or above on any given Regents subject exam to pass the subject and thus have it count toward receiving a HS diploma, a cutoff that does not change based on the realized distribution of performance in any given year.⁴⁹ They 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

⁴⁷In 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.

⁴⁸They find some evidence that females are more responsive in terms of longer-run impacts on educational attainment, as are students of higher ability as evidenced by previous standardized scores on ELA and math exams.

⁴⁹Until 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.

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.⁵⁰

Since a large mass of students in NYC are located near the pass/fail threshold (the median NYC public school student expects to receive an average score of 64.8 across all of her subjects), we might expect aggregate pass rates to be non-trivially sensitive to heat stress. At the same time, given the grade manipulation documented in previous work, which is most prevalent for scores just below the passing (65 point) cutoff, we would expect realized pass rates to be less sensitive to heat stress than an extrapolation of the β_T coefficient from section 5.1 might imply.

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

$$p_{ijsty} = \gamma_{iy} + \eta_s + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \beta_3 Time_{sty} + DOW_{sty} \beta_4 + \epsilon_{ijsty} \quad (8)$$

$$c_{ijsty} = \gamma_{iy} + \eta_s + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \beta_3 Time_{sty} + DOW_{sty} \beta_4 + \epsilon_{ijsty} \quad (9)$$

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

Tables 8 and 9 report the results from running variations of equations 8 and 9 that include subject, year, student, time of day fixed effects. 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 (column 1). 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).

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. Taken to-

⁵⁰The scale score needed to be considered "college ready" differs by subject. According to CUNY admissions, 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. Similarly, one can satisfy the mathematics skill requirement if you meet any of these criteria: SAT Math score of 500 or higher; ACT Math score of 21 or higher; N.Y. State Regents score of 70 or higher in Algebra I (Common Core) and successful completion of the Algebra 2/Trigonometry or higher-level course; score of 80 or higher in either Integrated Algebra, Geometry or Algebra 2/Trigonometry AND successful completion of the Algebra 2/Trigonometry or higher-level course; score of 75 or higher in Math A or Math B, Sequential II or Sequential III.

gether, these estimates suggest that experiencing hot ambient temperatures during a Regents 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 NYC student.

6.2 Long-Run Impacts of Acute Heat Exposure: Graduation Status

Figure 10 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 for 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, since, unlike exam scores, the outcome variable is no longer date-specific.⁵¹ 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.⁵²

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 would be to compare within exam-count and year-count cells controlling for observable factors.

To implement this strategy, I collapse the data at the student level, and estimate variations of the following model:

$$g_{ijcn} = \alpha_0 + \alpha_1 \overline{T}_{ij} + X_{ij}\alpha_2 + \chi_j + \theta_c + Z_i\alpha_3 + exams_n\alpha_4 + \epsilon_{ijc} \quad (10)$$

g_{ijcn} is a dummy for whether student i in school j and entering cohort c who takes n June Regents exams over the course of her high school career has graduated after 4 years in high school. \overline{T}_{ij} denotes the average temperature experienced by student i while taking June Regents exams in school j , up through her senior year. X_{ij} is a vector of weather controls

⁵¹Graduation 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.

⁵²That is, assuming that the average June climate in New York City can be represented by a (relatively stable) 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 further details).

averaged at the student-school level. χ_j denotes school fixed effects; θ_c denotes cohort fixed effects; Z_i is a vector of student-level controls including race, gender, federally subsidized school lunch eligibility, and previous ability (combined ELA and math z-scores); and $exams_n$ denotes a vector of number of June exams fixed effects.

The parameter of interest is α_1 , which captures the impact of an additional degree of heat exposure over all core June Regents exams on the likelihood of graduating on time.⁵³ School fixed effects account for potential bias arising from omitted variable bias due to unobserved determinants of graduation rates in the cross-section.⁵⁴ Cohort fixed effects in graduation rates allow for the possibility that heat exposure and graduation rates are correlated due to secular trends in both variables – though warming trends and average improvements in NYC schools would suggest this effect to lead to downward rather than upward bias in the estimate of α .

Table 10 presents the results from running variations of equation 10 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 is a conservative representation of the level at which quasi-random temperature variation occurs, but 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. This corresponds to a 3.12 to 3.34 percentage point decline in the likelihood of on-time graduation from a one standard deviation in average exam-time temperature (+4.4°F), or 4.59% to 4.91% decline relative to a mean on-time graduation rate of 68 percent.

7 Does Cumulative Heat Exposure Reduce the Rate of Learning?

The previous analyses suggest that short-run heat stress exerts a causal and statistically significant impact on student performance in high-stakes school settings. The following section assesses whether cumulative heat exposure during class time reduces the amount and rate of

⁵³The 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.

⁵⁴For instance, if low- (high-) performing schools tend to be located in areas of New York that are more- (less-) likely to experience heat stress, graduation rates and average exam-time heat exposure may be correlated in the cross-section.

learning achieved, as a window into understanding potential links between temperature and human capital accumulation.

7.1 Cumulative Learning Impacts of Repeated Heat Exposure

Figure 11 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 are likely reducing learning attainment, controlling for the impact of short-run heat exposure on contemporaneous cognitive 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 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.⁵⁵ 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.⁵⁶

I estimate variations of the following model:

$$y_{jsty} = \beta_0 + \beta_1 T_{jsty} + X_{jsty} \beta_2 + \sum_d^D \beta_3^d DD_{jT}^d + \chi_j + \eta_s + Z_{jsty} \beta_4 + \beta_5 Time_{sty} + DOW_{sty} \beta_6 + \beta_7 Year + \beta_8 Year^2 + \beta_9 Year^3 + \epsilon_{jsty} \quad (11)$$

where y_{jsty} 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_{jsty} denotes exam-time outdoor temperature at school j for subject s on date and time t , during year y ; γ_j denotes school fixed effects; η_s denotes subject fixed effects; and Z_{jsty} represents a vector of demographic controls averaged at the school by take (subject-month-year) level. $Time_{sty}$ represents a dummy for time of day (Time=1 denotes

⁵⁵This 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.

⁵⁶Recall that the main source of identifying variation for short-run impacts is day-to-day variation in exam-time temperature within student-year cells, which is best approximated in this case by variation across subjects within school-year cells.

afternoon exam), and DOW_{sty} is a vector of fixed effects for day of week. $Year...Year^3$ denotes a cubic time trend in scores.

School fixed effects account for possible spurious correlation between average school-level performance and local climate. Subject fixed effects control for average differences in performance across subjects. 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.

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.⁵⁷ I use a number of bin classifications for hot days, motivated by the existing literature (e.g. Barecca et al 2016; Hsiang and Deryugina, 2015) as well as the analyses presented in the previous section, which find negative impacts of heat stress beginning around 72°F. 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. Deschenes and Greenstone, 2011) suggests to be around 60°-70°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_{jsty} denotes a vector of contemporaneous 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 11 that allow for a flexible characterization of the reference category against which we can interpret the impact of hot days. Table 11 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.

As shown in Table 11, the results are highly 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, we can see that days between 80° and 90°F have a negative impact of between -0.011 (se=0.0031)

⁵⁷School 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.

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.⁵⁸

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). To the extent that the independent variables – notably cumulative temperature exposure in the classroom – are measured with error, these estimates are likely attenuated downward.

In conjunction with the findings from section 6, I take these results as preliminary evidence of long-run human capital impacts of heat exposure. 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 studies such as Chetty et al (2014) which examine the same population of NYC students and find significant impacts of improved learning on later-life outcomes.

8 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 a preliminary assessment of possible adaptive responses by students and teachers.

Taking a high school exit exam on a 90° F day results in a 4.5% (0.15 standard deviation) reduction in exam performance relative to a more optimal 72°F day, controlling for student specific unobservables. This amounts to roughly 1/4 of the Black-White test score gap. Given existing institutional constraints – namely, cardinal passing thresholds and graduation

⁵⁸Cold 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 (2015) who finds that the number of snow days and the amount of local snowfall adversely predict end-of-year performance in Massachusetts public schools.

requirements that depend on exam performance – these short-run performance impacts can have non-trivial longer-run effects on educational outcomes. I find that a 90°F day results in a 10.9% lower probability of passing an exam, 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, affecting at least 90,000 students, possibly many more.⁵⁹

Moreover, the evidence suggests that repeated heat stress can disrupt learning and reduce the rate of human capital accumulation. 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, though the effects are less precisely estimated. A year with five additional 80°F+ days is associated with a 2.1% (-0.05 standard deviation) reduction in learning on average, suggesting that heat exposure during class time can reduce long-run human capital and educational attainment. These effects are on par in magnitude with wiping out 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 comparable 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).

Based on building level air-conditioning installation data gathered from engineering surveys, school air conditioning penetration in New York City public schools (at less than 62%) seems to be far below residential and commercial sector averages for the region. Perhaps because of this, and due to the high stakes nature of these exams, teachers seem to have responded by boosting grades of students who experience hot exam sittings and score just below passing cutoffs. Building on evidence of grade manipulation by NYC teachers presented by Dee et al (2016), I find that the extent of bunching at passing score cutoffs is highly correlated with temperature during the test session in which a given subject exam was administered, suggesting that may have been trying to offset some of the potential long-term consequences of momentary 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 (particularly air conditioning) 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 heat stress should be included among the long list of relevant inputs to schooling. For instance, the timing of

⁵⁹See 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.

high-stakes exams 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, Lavy and Roth (2016). They might 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 (Davis and Gertler, 2015).

Second, this study raises new questions in an old debate regarding geography and economic prosperity (Gallup, Sachs, Mellinger, 1999; Acemoglu, Johnson, and Robinson, 2000; Rodrik et al, 2004; Nordhaus, 2008; Dell, Jones, and Olken, 2012). Figure 12 presents a simple binned scatterplot of standardized PISA math scores and average annual temperature for the 60 countries who participated in the 2012 Program for International Assessment (PISA).⁶⁰ It shows a strong negative correlation between standardized achievement and average temperature, even when controlling for per capita income. How much of this residual variation 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 and lack of protective capital?⁶¹

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 et al, 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 change, especially when using short-run, weather-based estimates to inform projections about the distant future.

⁶⁰Data on population-weighted average temperature by country is taken from Dell, Jones, and Olken (2012).

⁶¹The extent to which cross-country differences in standardized scores such as the PISA assessments is a subject that has received considerable attention. See Woessman (2016) and Hanushek and Woessman (2012) for a review of this literature as well as a discussion of how scores are standardized across countries and exam waves.

References

1. Acemoglu, Daron, Simon Johnson, and James A. Robinson. The colonial origins of comparative development: An empirical investigation. No. w7771. National bureau of economic research, 2000.
2. Acemoglu, Daron, and Melissa Dell. "Productivity differences between and within countries." *American Economic Journal: Macroeconomics* 2.1 (2010): 169-188.
3. Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. "Lead Exposure and Racial Disparities in Test Scores." Brown Univ. Working Paper (2015).
4. Angrist, Joshua D., and Victor Lavy. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114.2 (1999): 533-575.
5. Anderson, Craig A., and Dona C. Anderson. "Ambient temperature and violent crime: Tests of the linear and curvilinear hypotheses." *Journal of personality and social psychology* 46.1 (1984): 91.
6. Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro. "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century." *Journal of Political Economy* 124, no. 1 (2016): 105-159.
7. Burke, Marshall, and Kyle Emerick. "Adaptation to Climate Change: Evidence from US Agriculture." (2015).
8. Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. "Global non-linear effect of temperature on economic production." *Nature* (2015).
9. Cachon, Gerard, Santiago Gallino, and Marcelo Olivares. "Severe weather and automobile assembly productivity." *Columbia Business School Research Paper* 12/37 (2012).
10. Cascio, Elizabeth U., and Douglas O. Staiger. "Knowledge, tests, and fadeout in educational interventions". No. w18038. National Bureau of Economic Research, 2012.
11. Card, David, and Jesse Rothstein. "Racial segregation and the black-white test score gap." *Journal of Public Economics* 91.11 (2007): 2158-2184.
12. Card, D. (1999) "The Causal Effect of Education on Earnings, in the *Handbook of Labor Economics*, Vol. 3A, O. Ashenfelter and D. Card, eds. Amsterdam: Elsevier Science
13. Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2010). "How does your kindergarten classroom affect your earnings? Evidence from Project STAR" (No. w16381). National Bureau of Economic Research.
14. Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2009). "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records". (No. w15617). National Bureau of Economic Research.

15. Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *The American Economic Review* 104.9 (2014): 2633-2679.
16. Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment". No. w21156. National Bureau of Economic Research, 2015.
17. Cunha, Flavio, and James Heckman. "The Technology of Skill Formation." *The American Economic Review*, vol. 97, no. 2, 2007, pp. 3147.
18. Currie, Janet. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* (2009): 87-122.
19. Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., & Rivkin, S. G. (2009). "Does pollution increase school absences?". *The Review of Economics and Statistics*, 91(4), 682-694.
20. Davis, Lucas W., and Paul J. Gertler. "Contribution of air conditioning adoption to future energy use under global warming." *Proceedings of the National Academy of Sciences* 112.19 (2015): 5962-5967.
21. Dee, T. S., Dobbie, W., Jacob, B. A., and Rockoff, J. (2016). "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations". No. w22165. National Bureau of Economic Research.
22. Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. "Temperature shocks and economic growth: Evidence from the last half century." *American Economic Journal: Macroeconomics* 4.3 (2012): 66-95.
23. Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. "What do we learn from the weather? The new climateeconomy literature." *Journal of Economic Literature* 52.3 (2014): 740-798.
24. Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. "School choice, school quality, and postsecondary attainment." *The American economic review* 104, no. 3 (2014): 991-1013.
25. Deryugina, Tatyana, and Solomon M. Hsiang. "Does the environment still matter? Daily temperature and income in the United States." (2014).
26. Deschenes, Olivier and Michael Greenstone. 2011. "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US." *American Economic Journal: Applied Economics*, 3(4): 152-85.
27. Deschenes, Olivier, and Enrico Moretti. "Extreme weather events, mortality, and migration." *The Review of Economics and Statistics* 91.4 (2009): 659-681.
28. Diamond, Rebecca, and Petra Persson. "The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests". No. w22207. National Bureau of Economic Research, 2016.

29. Duran-Narucki, Valkiria. "School building condition, school attendance, and academic achievement in New York City public schools: A mediation model." *Journal of environmental psychology* 28.3 (2008): 278-286.
30. Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger. "Geography and economic development." *International regional science review* 22.2 (1999): 179-232.
31. Glewwe, Paul, and Michael Kremer. "Schools, teachers, and education outcomes in developing countries." *Handbook of the Economics of Education* 2 (2006): 945-1017.
32. Goodman, Joshua. "Flaking out: Student absences and snow days as disruptions of instructional time". No. w20221. National Bureau of Economic Research, 2014.
33. Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton. "Developing a social cost of carbon for US regulatory analysis: A methodology and interpretation." *Review of Environmental Economics and Policy* 7, no. 1 (2013): 23-46.
34. Greenstone, Michael, and B. Kelsey Jack. "Envirodevonomics: A research agenda for an emerging field." *Journal of Economic Literature* 53.1 (2015): 5-42.
35. Grether, W. (1973). "Human performance at elevated environmental temperatures." *Aerospace Medicine* 44(7): 747-755.
36. Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton. "Developing a social cost of carbon for US regulatory analysis: A methodology and interpretation." *Review of Environmental Economics and Policy* 7.1 (2013): 23-46.
37. Hallegatte, Stephane. "Strategies to adapt to an uncertain climate change." *Global Environmental Change* 19.2 (2009): 240-247.
38. Hanushek, Eric A., and Ludger Woessmann. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of Economic Growth* 17.4 (2012): 267-321.
39. Hocking, Chris, Richard B. Silberstein, Wai Man Lau, Con Stough, and Warren Roberts. "Evaluation of cognitive performance in the heat by functional brain imaging and psychometric testing." *Comparative Biochemistry and Physiology Part A: Molecular Integrative Physiology* 128, no. 4 (2001): 719-734.
40. Hsiang, Solomon M. "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America." *Proceedings of the National Academy of sciences* 107.35 (2010): 15367-15372.
41. Hsiang, Solomon M., Marshall Burke, and Edward Miguel. "Quantifying the influence of climate on human conflict." *Science* 341.6151 (2013): 1235367.
42. Isen, A., Rossin-Slater, M., & Walker, R. (2015). "Heat and long-run human capital formation". Unpublished Manuscript.

43. Kahn, M. "The Climate Change Adaptation Literature." *The Review of Environmental Economics and Policy* (2016).
44. Kenrick, Douglas T., and Steven W. MacFarlane. "Ambient temperature and horn honking a field study of the heat-aggression relationship." *Environment and behavior* 18.2 (1986): 179-191.
45. Ebenstein, Avraham, Victor Lavy, and Sefi Roth. "The long run economic consequences of high-stakes examinations: evidence from transitory variation in pollution." *American Economic Journal: Applied Economics* (2016).
46. Lavy, Victor. "Long run effects of free school choice: College attainment, employment, earnings, and social outcomes at adulthood". No. w20843. National Bureau of Economic Research, 2015.
47. Mackworth, Norman H. "Effects of heat on wireless operators." *British journal of industrial medicine* 3.3 (1946): 143.
48. Mankiw, N. Gregory, David Romer, and David N. Weil. "A Contribution to the Empirics of Economic Growth." *The Quarterly Journal of Economics* 107.2 (1992): 407-437.
49. Mansur, Erin T., Robert Mendelsohn, and Wendy Morrison. "Climate change adaptation: A study of fuel choice and consumption in the US energy sector." *Journal of Environmental Economics and Management* 55.2 (2008): 175-193.
50. Mendell, Mark J., and Garvin A. Heath. "Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature." *Indoor air* 15.1 (2005): 27-52.
51. Mendelsohn, Robert. "Efficient adaptation to climate change". *Climatic Change* 45.3-4 (2000): 583-600.
52. Niu, Jianlei. "Some significant environmental issues in high-rise residential building design in urban areas." *Energy and buildings* 36.12 (2004): 1259-1263.
53. Oreopoulos P. and K. Salvanes. (2011) "Priceless: The Nonpecuniary Benefits of Schooling," *Journal of Economic Perspectives* 25(1), 159-84.
54. Pindyck, Robert S. "Climate change policy: What do the models tell us?." *Journal of Economic Literature* 51.3 (2013): 860-872.
55. Peet, Evan D. "Environment and Human Capital: The Effects of Early-Life Exposure to Pollutants in the Philippines." (2014).
56. Raichle, Marcus E., and Mark A. Mintun. "Brain work and brain imaging." *Annual Review of Neuroscience* 29 (2006): 449-476.
57. Roback, J (1982) "Wages, Rents, and the Quality of Life". *Journal of Political Economy* Vol. 90 No. 6 pp1257-1278.

58. Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. "Institutions rule: the primacy of institutions over geography and integration in economic development." *Journal of economic growth* 9.2 (2004): 131-165.
59. Rosenzweig, Cynthia, William Solecki, and Ronald Slosberg. "Mitigating New York City's heat island with urban forestry, living roofs, and light surfaces." A report to the New York State Energy Research and Development Authority (2006).
60. Roth, S. (2016) "The Contemporaneous Effect of Indoor Air Pollution on Cognitive Performance: Evidence from the UK" Unpublished Manuscript.
61. Sandstrom, F. Mikael, and Fredrik Bergström. "School vouchers in practice: Competition will not hurt you." *Journal of Public Economics* 89.2 (2005): 351-380.
62. Schoer, L. and Shaffran, J. (1973) "A combined evaluation of three separate research projects on the effects of thermal environment on learning and performance". *ASHRAE Trans.*, 79, 97108.
63. Seppanen, Olli, William J. Fisk, and Q. H. Lei. "Effect of temperature on task performance in office environment." Lawrence Berkeley National Laboratory (2006).
64. Sievertsen, Hans Henrik, Francesca Gino, and Marco Piovesan. "Cognitive fatigue influences students' performance on standardized tests." *Proceedings of the National Academy of Sciences* (2016): 201516947.
65. Sinha, Paramita, and Maureen L. Cropper. "Household Location Decisions and the Value of Climate Amenities". No. w21826. National Bureau of Economic Research, 2015.
66. Sudarshan, Anant, and Meenu Tewari. "The economic impacts of temperature on industrial productivity: Evidence from Indian manufacturing". Working Paper, 2013.
67. Tan, Xuan, and Rochelle Michel. "Why do standardized testing programs report scaled scores." *ETS Research and Development Connections* 16 (2011): 1-6.
68. Woessmann, Ludger. "The importance of school systems: Evidence from international differences in student achievement." *The Journal of Economic Perspectives* 30.3 (2016): 3-31.
69. Zivin, Joshua Graff, and Matthew Neidell. "Temperature and the allocation of time: Implications for climate change." *Journal of Labor Economics* 32.1 (2014): 1-26.
70. Zivin, Joshua S. Graff, Solomon M. Hsiang, and Matthew J. Neidell. "Temperature and Human Capital in the Short-and Long-run". No. w21157. National Bureau of Economic Research, 2015.

Figures and Tables

Figure 1 Short-Run Identifying Variation in Temperature

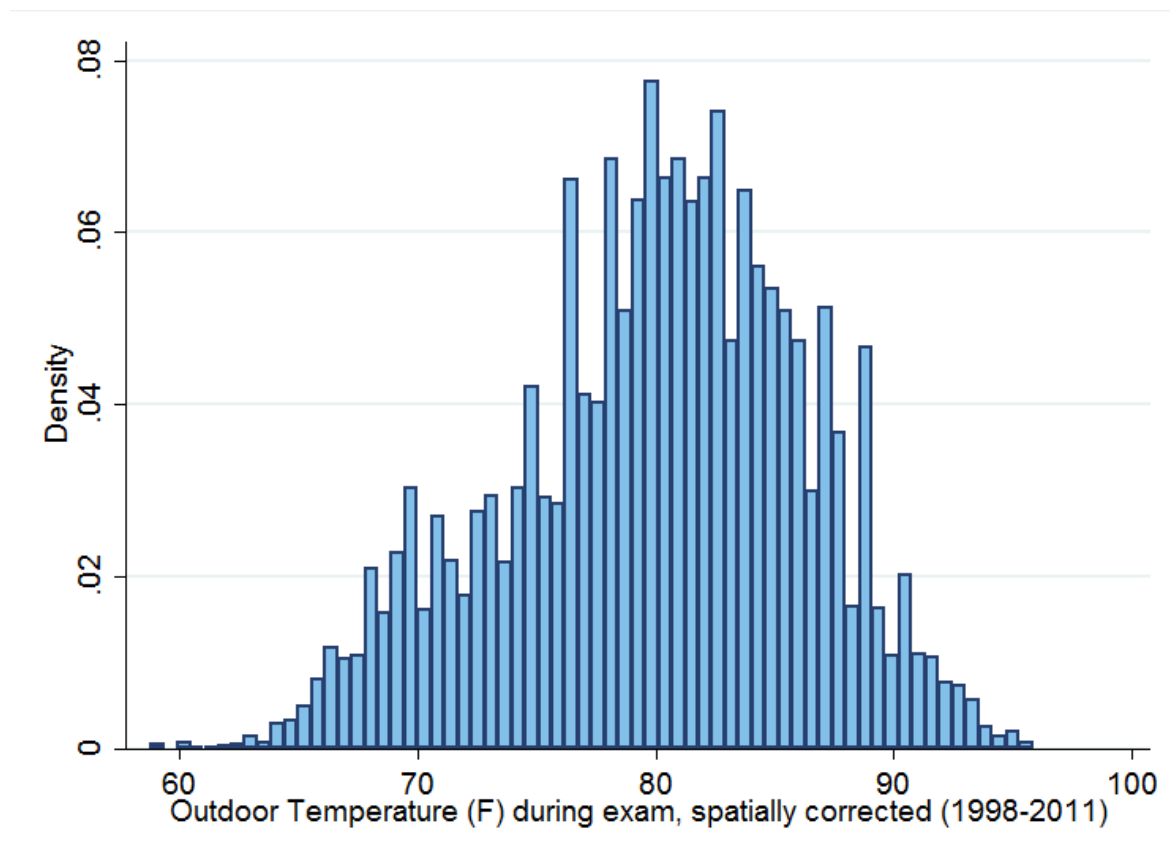


Figure 1: Temperatures measured at the school level, weighted by number of exam observations by date and time

Notes: This figure illustrates the source of identifying variation for short-run performance impacts of heat stress. It presents realized exam-time temperatures for all June Regents exam observations over the sample period (1999-2011), inclusive of spatial and temporal temperature corrections.

Figure 2 Temperature Variation on Two Consecutive Test Days

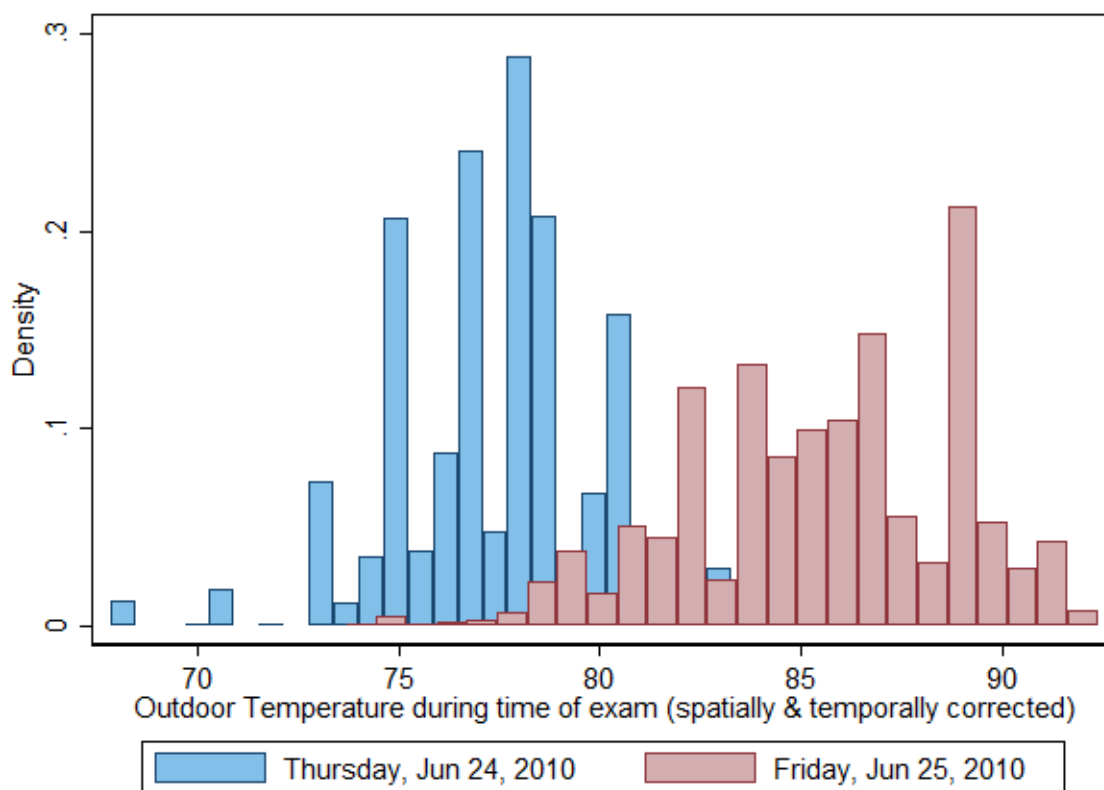


Figure 2: Temperatures measured at the school level, weighted by student population.

Notes: This figure illustrates the source of identifying variation for short-run performance impacts of heat stress. It presents realized exam-time temperatures 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, plotting observations at the student-exam level.

Figure 3 Variation in Cumulative Heat Exposure by Borough and Year in New York City

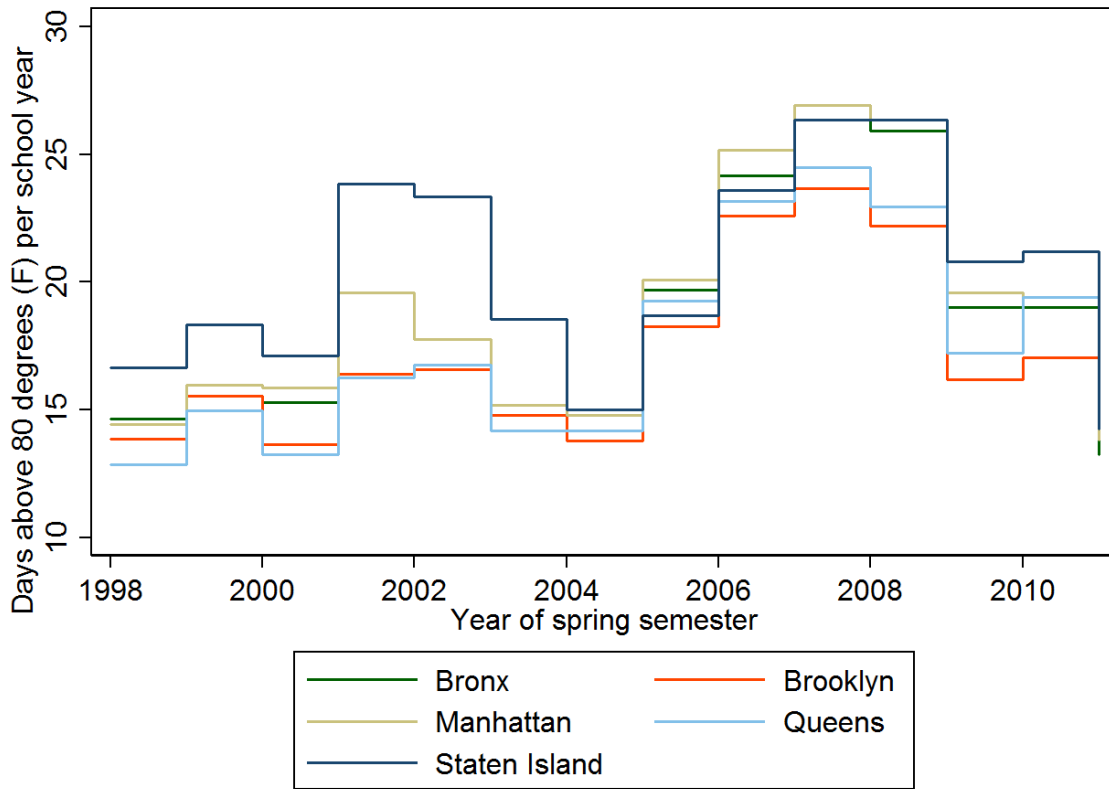


Figure 3: Temperatures measured at the nearest weather station (borough) level.

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 Borough

Borough	Regents Score	Pass Rate	Proficiency Rate	Previous Ability
Bronx	60.86 (18.17)	0.48 (0.50)	0.23 (0.42)	-0.19 (1.40)
Brooklyn	64.96 (17.81)	0.58 (0.49)	0.32 (0.47)	0.17 (1.41)
Manhattan	66.66 (18.08)	0.61 (0.49)	0.36 (0.48)	0.31 (1.49)
Queens	65.81 (17.42)	0.60 (0.49)	0.34 (0.47)	0.23 (1.33)
Staten Island	67.38 (17.23)	0.63 (0.48)	0.37 (0.48)	0.43 (1.44)
Total	64.86 (17.92)	0.57 (0.49)	0.32 (0.47)	0.16 (1.42)

Notes: This table presents summary statistics for the key outcome variables from June 1999 to June 2011, for the 4.5 million exam observations in the study sample. Previous ability denotes a combined z-score for standardized English Language and Arts (ELA) and math exams taken in 3rd-8th grade for the students in the Regents exam sample, averaged by student.

Table 2: Summary Statistics by Ethnicity

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

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.

Figure 4 Air Conditioning Status by NYC Public School Building

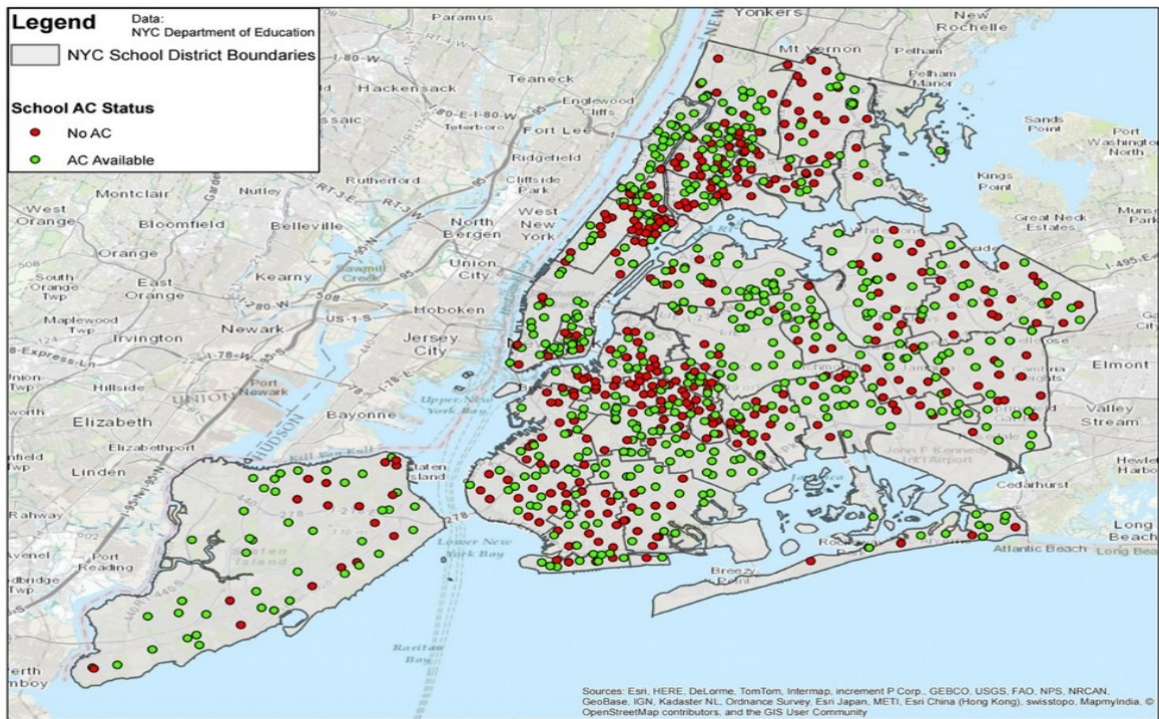


Figure 4: NYC public school buildings by AC status

Notes: This figure 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.

Figure 5 Short-Run Impacts of Heat Stress on Exam Performance: Regents Scale Scores

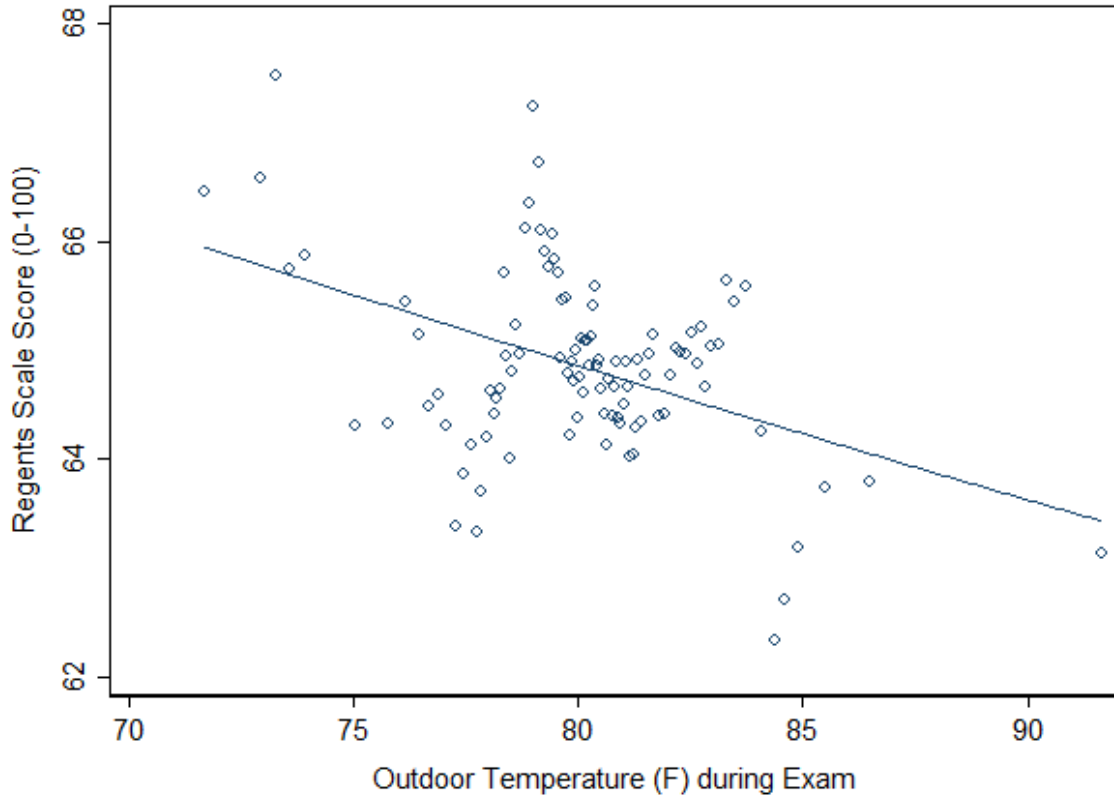


Figure 5: Dependent variable is Regents scale score (0-100)

Notes: This figure presents a binned scatterplot of Regents scores and exam-time temperature, by percentile of the realized exam-time temperature distribution, controlling for average differences across subjects, years, and schools, as well as exam-day precipitation and humidity (N=4,509,095). Each dot represents approximately 220,000 exam observations.

Figure 6 Short-Run Impacts of Heat Stress on Exam Performance: Standardized Regents Performance

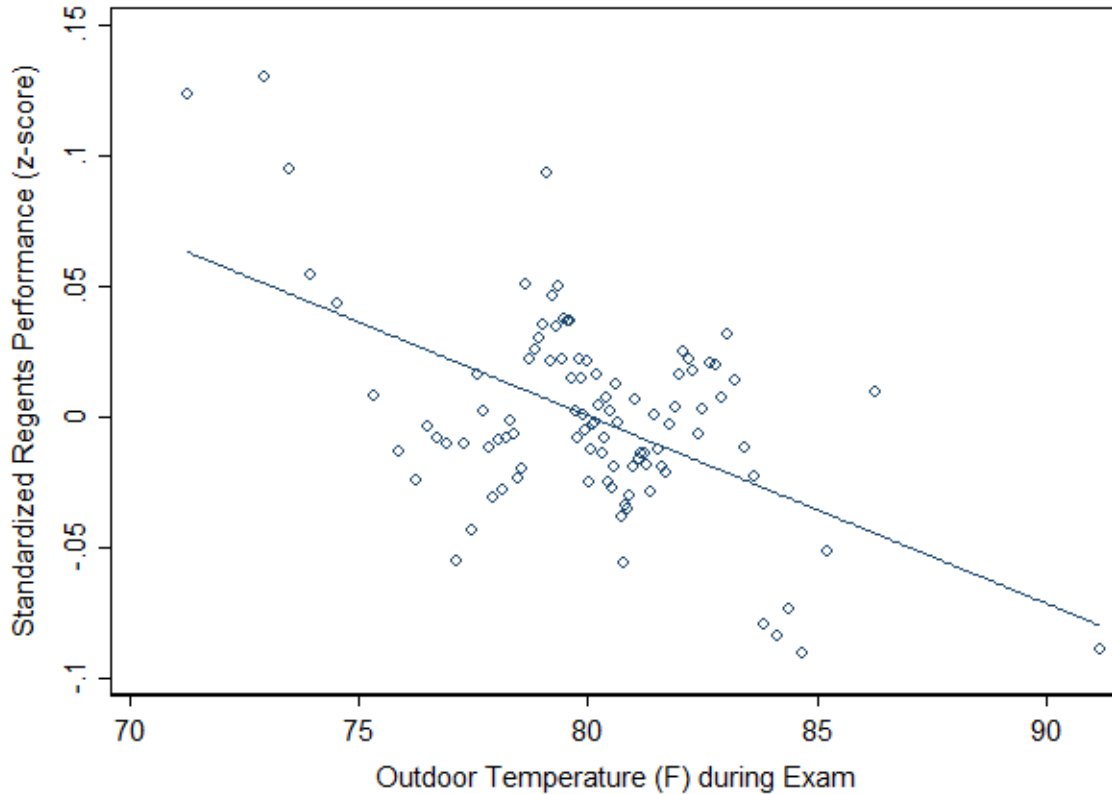


Figure 6: Dependent variable is Regents z-score

Notes: This figure presents a binned scatterplot of standardized Regents performance (all subjects and years) and exam-time temperature, by percentile of the realized exam-time temperature distribution, plotting residual variation after controlling for school-level averages, average differences across subjects, average differences across years, and exam-day precipitation and humidity (N=4,509,095). Each dot represents approximately 220,000 exam observations.

Table 3 Short-Run Impacts of Heat Stress on Exam Performance: Standardized Regents Performance

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

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

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

Table 3: Short-run impacts: Standardized performance for all subjects and years
Notes: The dependent variable is standardized Regents performance (standardized over the sample period 1999-2011). Observations are at student, exam, and date-and-time-level. Student, subject, school, year, student-by-year and/or school-by-year fixed effects are suppressed in output, and 919,067 singleton observations are dropped. Temperature is temporally corrected to account for diurnal fluctuations, and spatially corrected to account for urban heat island effects (see online appendix). All regressions include controls for daily dewpoint, precipitation, ozone, and pm2.5.

Table 4 Short-Run Impacts of Heat Stress on Exam Performance: Regents Scale Scores (0-100)

	(1)	(2)	(3)	(4)
	Score	Score	Score	Score
Temperature (F)	-0.152*** (0.0414)	-0.132*** (0.0371)	-0.183*** (0.0418)	-0.194*** (0.0405)
Afternoon	-0.532* (0.232)	-0.598** (0.214)	-0.322 (0.254)	-0.279 (0.228)
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.

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

Table 4: Short-run impacts: Regents scale scores

Notes: The dependent variable is Regents exam scaled score (0-100). Observations are at student, exam, and date-and-time-level. Student, subject, school, year, student-by-year and/or school-by-year fixed effects are suppressed in output, and 919,067 singleton observations are dropped. Temperature is temporally corrected to account for diurnal fluctuations, and spatially corrected to account for urban heat island effects (see online appendix). All regressions include controls for daily dewpoint, precipitation, ozone, and pm2.5.

Table5 Heterogeneity in Short-Run Impacts by Demographic Sub-Group

	(1) Regents z-score (by subject)
Temperature (°F)	-0.00825** (0.00301)
Asian × Temperature (°F)	0.00557* (0.00266)
Black × Temperature (°F)	-0.00214 (0.00200)
Hispanic × Temperature (°F)	-0.000458 (0.00192)
FSS Lunch × Temperature (°F)	0.000521 (0.000715)
Male × Temperature (°F)	0.00347** (0.00107)
Bottom Quintile (previous z) × Temperature (°F)	0.00445** (0.00105)
Top Quintile (previous z) × Temperature (°F)	0.00115 (0.00193)
Fixed Effects	
Year	X
Subject	X
Student	X
N	4347718
r2	0.710

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

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

Table 5: Heterogeneity in short-run impacts: standardized performance by subject
Notes: The dependent variable is standardized Regents performance (standardized by subject over the study period 1999-2011). Observations are at student, exam, and date-and -time-level. All regressions include student, subject, and year fixed effects, as well as controls for dewpoint. Student, subject, and year fixed effects suppressed in output, and singleton observations are dropped. Robust standard errors clustered at the borough by date-time level. Temperature is temporally corrected to account for diurnal fluctuations, and spatially corrected to account for urban heat island effects (see online appendix).

Table6 Heterogeneity in Short-Run Impacts by School Air Conditioning Status

	(1)	(2)	(3)	(4)	(5)
	All	Central AC	Any AC	No Central AC	No AC
Temperature (°F)	-0.00613** (0.00219)	-0.00530 (0.00293)	-0.00517 (0.00273)	-0.00649* (0.00274)	-0.00621* (0.00261)
Precipitation (mm)	0.00183 (0.00128)	0.00212 (0.00154)	0.00177 (0.00141)	0.00219 (0.00156)	0.00197 (0.00147)
Afternoon	-0.0417** (0.0149)	-0.0439* (0.0181)	-0.0473** (0.0169)	-0.0338* (0.0171)	-0.0369* (0.0163)
N	4347718	906451	1019279	1611336	1724670
r2	0.710	0.720	0.717	0.710	0.709

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

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

Table 6: Short-run impacts: Standardized performance by subject

Notes: The dependent variable is standardized Regents performance (standardized by subject over the study period 1999-2011). Observations are 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. Student, subject, and year fixed effects suppressed in output, and singleton observations are dropped. Robust standard errors clustered at the borough by date-time level. Temperature is temporally corrected to account for diurnal fluctuations, and spatially corrected to account for urban heat island effects (see online appendix).

Figure 7 Documenting Grade Manipulation: Bunching at Passing and Proficiency Thresholds

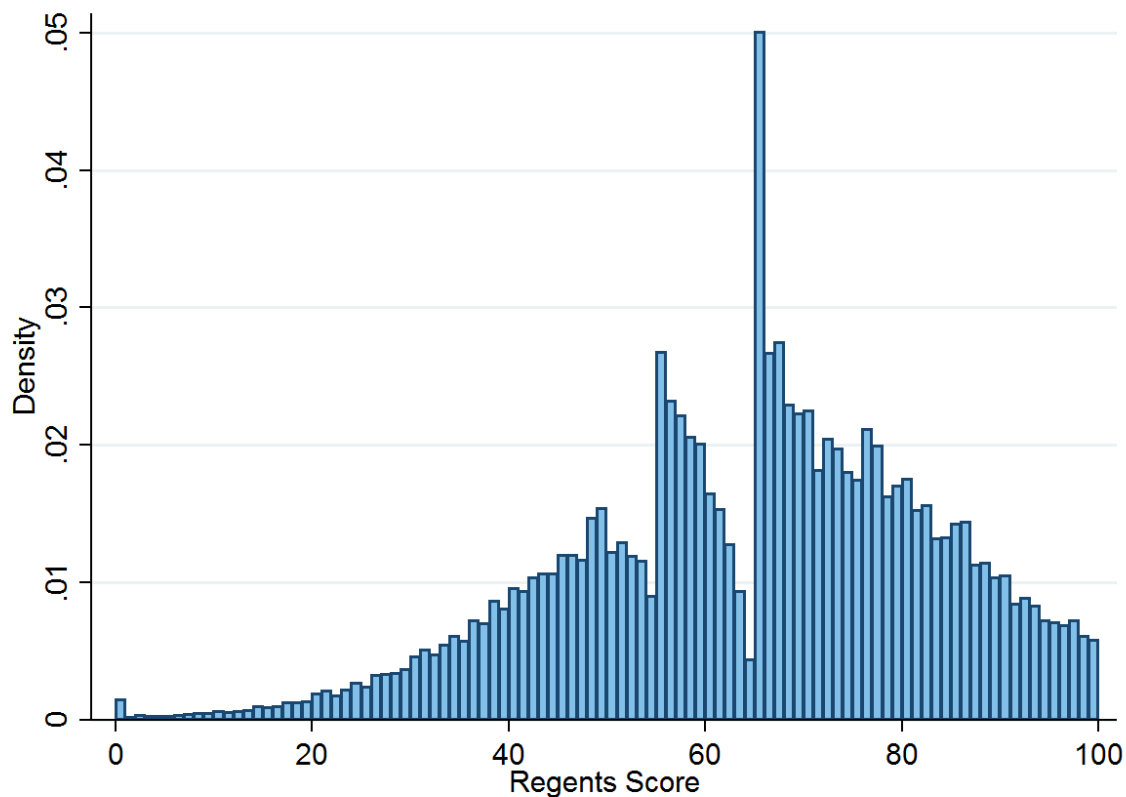


Figure 7: All Regents exams in core subjects prior to NYC grading reforms in 2011-2012.

Notes: This figure presents a histogram of Regents exam scores from June 1999 to June 2011. As has been documented by Dee et al (2016), and is evident from visual inspection, a large number of observations bunch at the pass/fail cutoffs, scores of 55 and 65 for local and Regents diploma requirements respectively.

Figure 8 Adaptive Grading: Variation in Grade Manipulation by Exam-Time Temperature

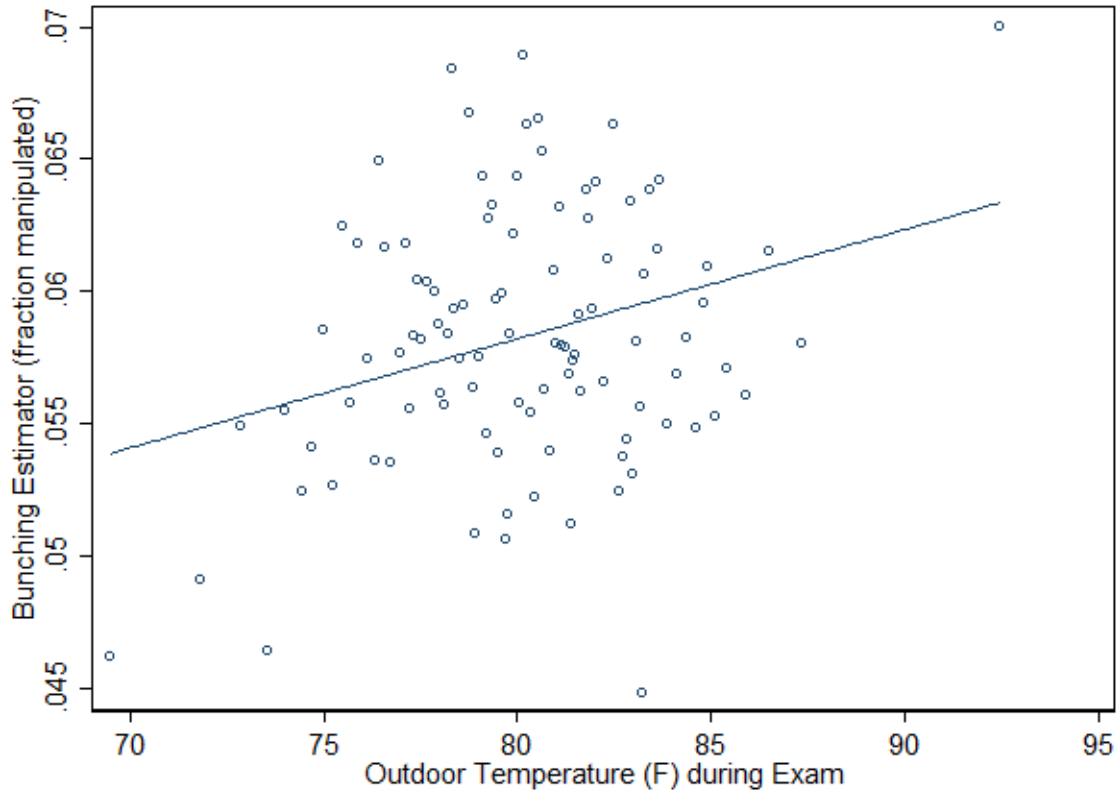


Figure 8: Grade Manipulation varies with exam-time temperature by subject, school, and take.

Notes: This figure presents a binned scatterplot of the residualized degree of bunching at the school-subject-date level by quantile of the exam-time temperature distribution, controlling for averages across subjects and years, as well as for exam-day precipitation. 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.

Table 7: Adaptive Grading: Grade Manipulation by Exam-Time Temperature

	(1)	(2)	(3)
	Bunching Estimator	Bunching Estimator	Bunching Estimator
Temperature (°F)	0.0013*** (0.0004)	0.0016*** (0.0005)	0.0010*** (0.0003)
Year			1.6705** (0.6591)
Year ²			-0.00042** (0.0002)
Fixed Effects			
School FE	X	X	X
Year FE		X	
Subject FE		X	X
N	30,731	30,731	30,731
r ²	0.018	0.082	0.081

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Dependent variable is degree of bunching by school-subject-date-and-time

Notes: The dependent variable is the extent of bunching by school-subject-date, measured as the fraction of exam observations that are missing from above and below the hypothetical score distribution, in the areas above and below the passing cutoffs that are subject to discretionary grade manipulation due to NY Regents grading rules respectively. Temperature is measured at the school-by-date (subject-month-year) level, accounting for temporal and spatial variation. All regressions include controls for exam-day precipitation, ozone, pm2.5 and dewpoint . School, subject, and year fixed effects are suppressed in output. Robust standard errors, in parentheses, are clustered at the borough by date and time level.

Figure 9 Proportion of Manipulable Exams Manipulated

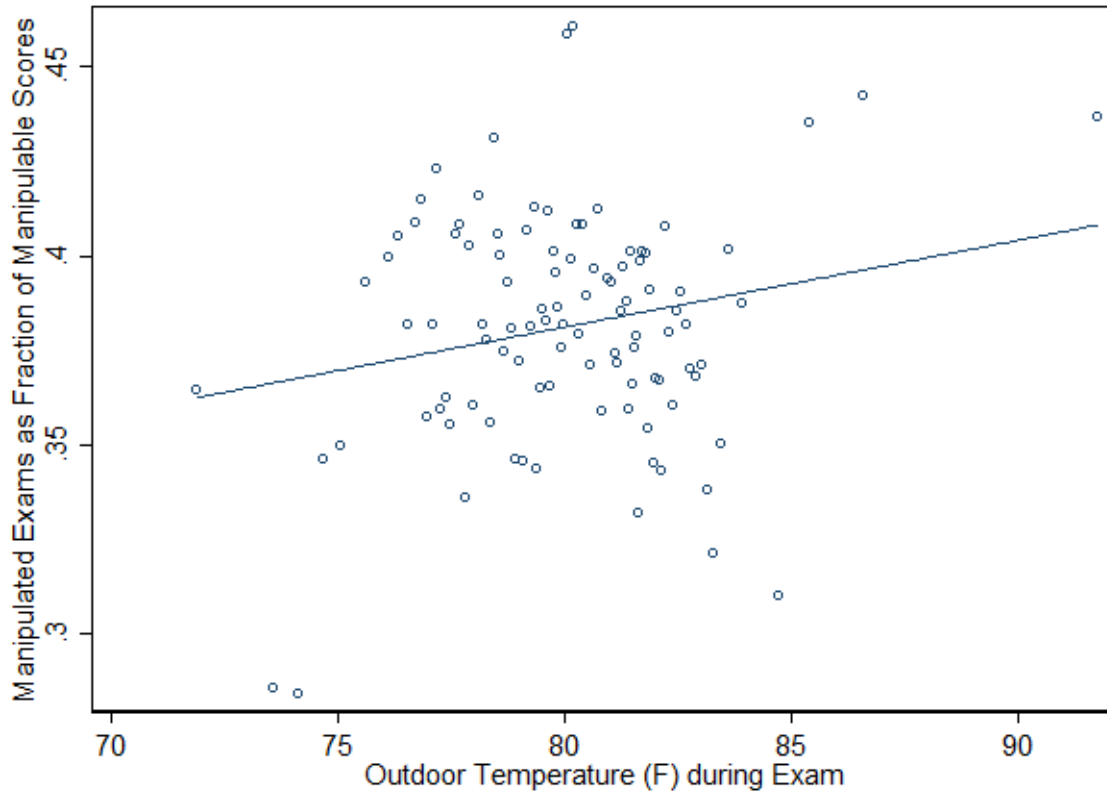


Figure 9: Grade Manipulation varies with exam-time temperature by subject, school, and take.

Notes: This figure presents a binned scatterplot of the residualized degree of bunching at the school-subject-date level by quantile of the exam-time temperature distribution, controlling for averages across subjects, years, and schools, as well as for exam-day precipitation. The bunching estimator is calculated by integrating the distance between predicted and observed score fractions of scores within the manipulable zone, and expressing this fraction as a proportion of the total number of manipulable scores for that school, subject, and take. The plot suggests that, not only was there more potential bunching after hot exam days due to mechanical shifts in the score distribution, but that teachers boosted a larger fraction of manipulable grades after hot exam days. Included in the analysis are all June Regents exams in core subjects between 1999 and 2011.

Table 8 Short-Run Impacts of Heat Stress on Exam Performance: Regents Pass Rates

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

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

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

Table 8: Dependent variable is Dummy for Passed Regents exam

Notes: Observations are at student, exam, and date-and-time-level. Student, subject, school, year, student-by-year and/or school-by-year fixed effects are suppressed in output, and 919,067 singleton observations are dropped. Temperature is temporally corrected to account for diurnal fluctuations, and spatially corrected to account for urban heat island effects (see online appendix). All regressions include controls for daily dewpoint, precipitation, ozone, and pm2.5.

Table 9 Short-Run Impacts of Heat Stress on Exam Performance: College Proficiency Status

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

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

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

Table 9: Dependent variable is dummy for Proficiency status achieved
Notes: Observations are at student, exam, and date-and-time-level. Student, subject, school, year, student-by-year and/or school-by-year fixed effects are suppressed in output, and 919,067 singleton observations are dropped. Temperature is temporally corrected to account for diurnal fluctuations, and spatially corrected to account for urban heat island effects (see online appendix). All regressions include controls for daily dewpoint, precipitation, ozone, and pm2.5.

Figure 10: 4-year Graduation Status and Average Exam-Time Temperature by student

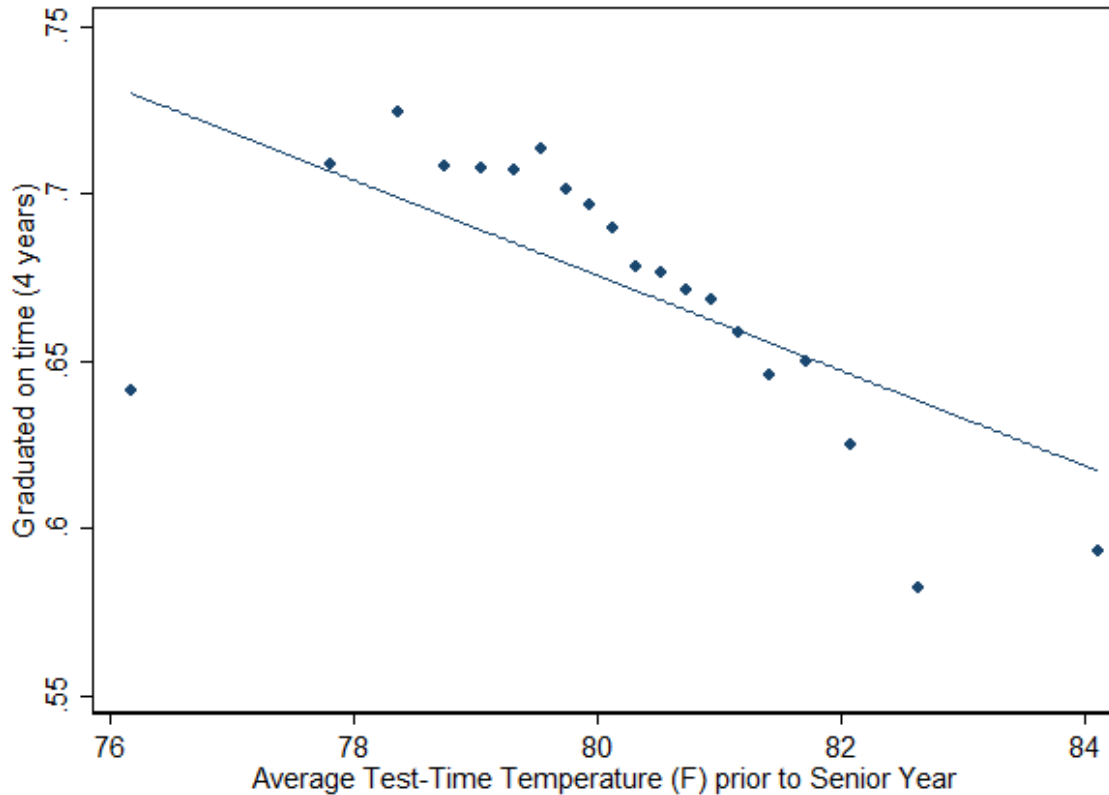


Figure 10: Dependent variable is dummy for graduated in 4 Years

Notes: This figure presents a binned scatterplot of 4-year graduation status by quantile of the observed average exam-time temperature distribution, where exam-time temperatures are averaged by student for all June Regents core subject exam sessions up through their senior year of high school. The figure plots residual variation after controls are included for school-level averages, the total number of exams taken by student, average exam-day precipitation and ozone, as well as student-level observables (ethnicity, gender, average previous z-scores, federally subsidized school lunch eligibility) are included.

Table 10: Longer-run Impacts, 4 Year Graduation Rates

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

Standard errors in parentheses

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

Notes: The dependent variable is a dummy for whether or not student graduated in four years, with observations at the student level. Robust standard errors are clustered at the borough by date and time level. All regressions include controls for daily precipitation, ozone, and dewpoint. Number of June exams, cohort, and school fixed effects are suppressed in output. Temperature is temporally corrected to account for diurnal fluctuations, and spatially corrected to account for urban heat island effects. Average temperatures from all June Regents exam takes up through the students' senior year in high school are computed by student.

Figure 11: Cumulative Learning Impacts of Heat Exposure during Preceding School Year

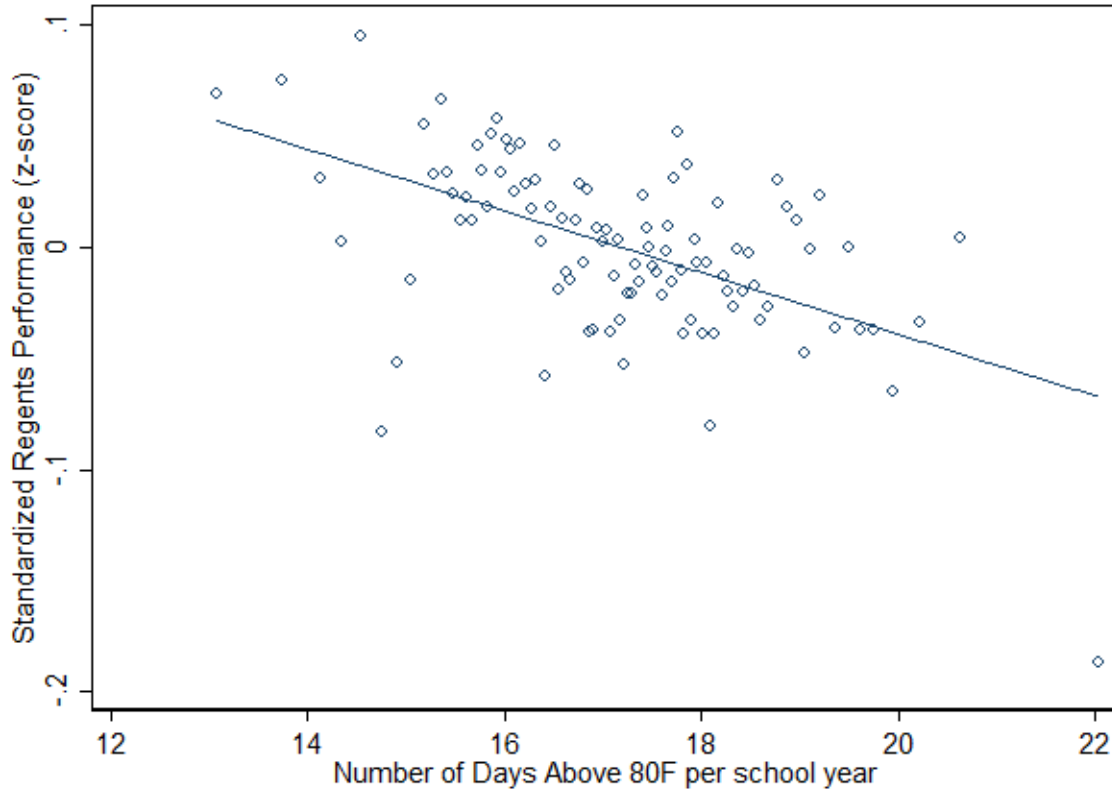


Figure 11: Temperatures measured at the school level, weighted by student population.

Notes: This figure presents a binned scatterplot of average June Regents performance by school, subject and exam date, binned by quantile of the distribution of cumulative heat stress, measured by the number of days per degree bin (showing days with maximum temperatures above 80°F), plotting residual variation after controls for exam-day temperature and precipitation, as well as school-, subject-, time of day, day of week, and day of month fixed effects are included.

Table 11: Cumulative Learning Impacts of Heat Exposure during Preceding School Year

	(1)	(2)	(3)	(4)
	Z-score	Z-score	Z-score	Z-score
Days with max temp 70°F-80°F	0.000669 (0.00171)	0.00506 (0.00322)	0.00131 (0.00187)	0.00412 (0.00348)
Days with max temp 80°F-90°F	-0.0108*** (0.00310)	-0.0121*** (0.00329)	-0.0114*** (0.00279)	-0.00337 (0.00668)
Days with max temp above 90°F	0.0209* (0.00937)	0.00459 (0.0129)	0.00804 (0.00973)	0.0198 (0.0182)
Exam-Time Temperature (°F)	-0.00580*** (0.00152)	-0.00830*** (0.00172)	-0.00790*** (0.00165)	-0.00824*** (0.00179)
Avg Temperature (°F) during school year			0.0418*** (0.00968)	
Degree Day Controls				
Cold Day Bins		X	X	
All Degree Day Bins				X
N	22563	22563	22563	22563
r2	0.587	0.587	0.587	0.587

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

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

Table 11: Dependent variable is standardized Regents performance (z-score by subject)
Notes: The dependent variable is standardized Regents performance (z-score by subject), collapsed at the school, date-and-time, and year level. All regressions include school, subject, and time of day fixed effects which are suppressed in the output. Robust standard errors, in parentheses, are clustered at the borough by year level. Cumulative degree day variables are assigned by closest weather station, which is at the Borough level, and summed beginning on the first day of the preceding fall semester up through the first day of June Regents exams that year.

Figure 12 Climate and Standardized Math Achievement Across Countries, Controlling for Per Capita Income (2012)

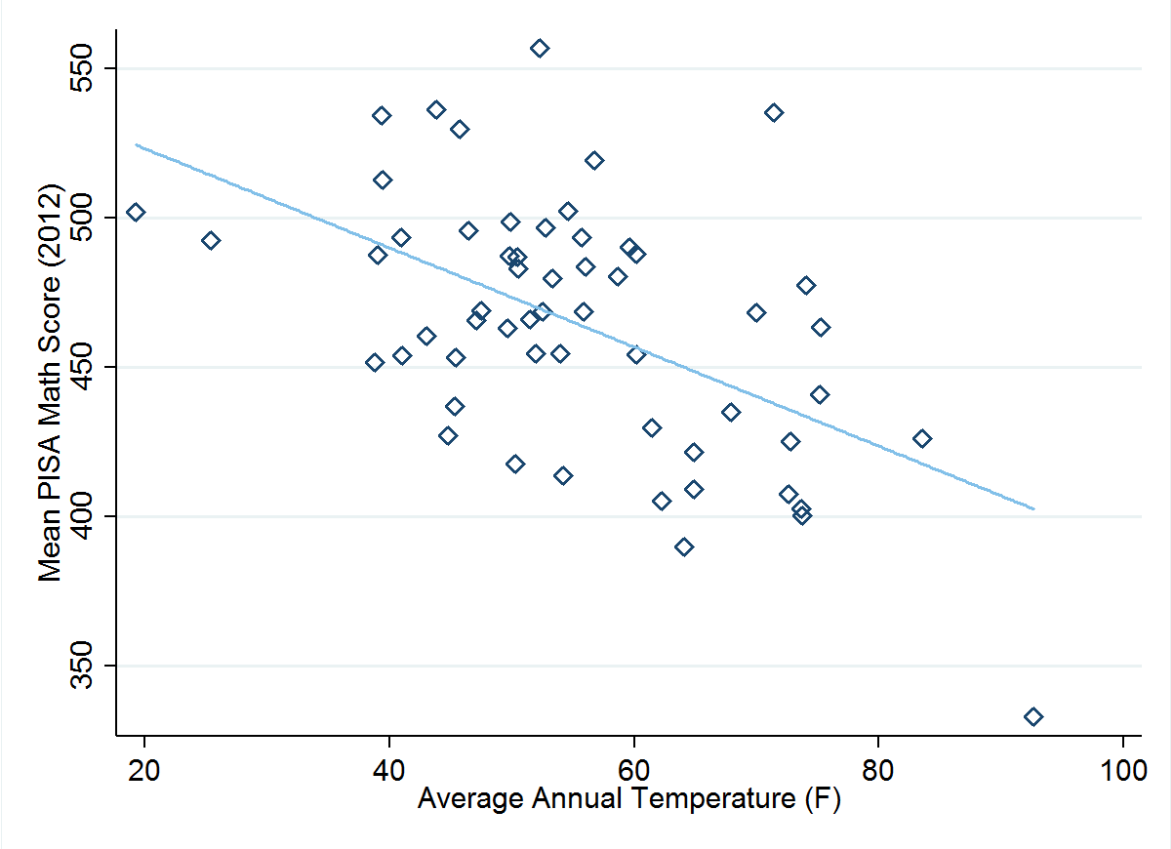


Figure 12: Mean PISA math scores and average annual temperature at the country level

Notes: This figure presents a 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.