

Heat Stress and Human Capital Production

Jisung Park

Harvard University Economics Department *

August 31, 2016

Abstract

How does temperature affect human capital production? Despite a longstanding literature assessing the efficacy of various inputs to schooling and recent evidence on the impact of temperature stress on human welfare, few studies have explored the role of temperature as an input to the human capital production process. Evidence from 4.6 million high school exit exams from New York City suggests that heat stress can have a significant impact on student performance, leading to diminished exam performance in the short run and reduced educational attainment in the long run. Taking an exam on a 90°F day relative to a 72°F day leads to a 0.19 standard deviation reduction in performance, equivalent to a quarter of the Black-White achievement gap, and a 12.3% higher likelihood of failing a subject. Hot days during the school year reduce subsequent exam performance by approximately 0.022 standard deviations per day above 80°F, suggesting that cumulative heat exposure can reduce the rate of learning. This study finds some evidence for protective effects of air conditioning, and strong evidence for “adaptive grading”, whereby teachers partially offset the adverse performance impacts of acute heat stress by manipulating grades near passing thresholds when students have experienced hot exam sittings. These findings have implications for estimating the social cost of carbon, designing education policy, and our understanding of the role that climatic factors play in explaining income gaps across individuals and nations.

*Sincere thanks to Larry Katz, Robert Stavins, Andrei Shleifer, Geoffrey Heal, Raj Chetty, Joe Aldy, Claudia Goldin, Edward Glaeser, Jim Stock, Martin Feldstein, Melissa Dell, Jeff Miron, Jeff Liebman, Greg Mankiw, Mandy Pallais, David Keith, Josh Goodman, David Cutler, Max Auffhammer, Sol Hsiang, Sean Reardon, Richard Sweeney, Aaron Bernstein, and numerous seminar participants at Harvard, Columbia, UC Berkeley, Seoul National, Oxford, the NYC Department of Health and Mental Hygiene, 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

Beginning with seminal contributions by Becker (1961), Mincer (1958), and Schultz (1961), a vast economic literature has documented the importance of human capital accumulation in improving standards of living – for individuals and nations alike.¹ A more recent literature has highlighted the causal impact of climatic factors on welfare-relevant outcomes including health, labor supply, and even macroeconomic output which, taken together, suggest that heat stress can directly affect economic activity through its effect on human physiology and cognition.² And yet, few studies have explored the role that temperature plays in the human capital production process, particularly in school environments. This study is the first to explore the direct impact of heat stress on student performance and student attainment using data from school settings.

To assess whether and how heat stress may affect human capital production, I use administrative data from the nation’s largest school district: New York City public schools. I focus on three related but separate research questions. First, does heat stress affect student performance in the short run? That is, do early lab-based findings – wherein cognitive performance declines rapidly with elevated temperatures – extend to school contexts, where the economic stakes are presumably higher? Second, does heat stress affect longer run human capital attainment, or does it merely add noise to the signal extraction process of high-stakes testing? As in the case of many educational interventions that have been explored previously (e.g. teacher quality, class size reduction), holistic welfare accounting hinges upon whether observed performance impacts represent temporary reductions in cognitive capacity or lead to more lasting changes in human capital attainment, with their attendant impacts on earnings and other later-life outcomes.³ Third, how much and in what ways do students and teachers adapt? Assessing the potential scope for adaptation to heat stress in school settings is important in thinking about the potential human capital impacts from future climate change, especially if one is interested in what a climate-human capital link might imply for the optimal social cost of carbon (Greenstone et al, 2013; Kahn, 2016).

My research design is based on a simple premise: that short-run variations in temperature

¹For a review of the literature on wage returns to human capital, see Card (1999). For a review of the non-pecuniary returns to human capital, see Oreopolous and Salvanes (2011). For reviews of the role that health and human capital play in macroeconomic growth and development, see Hanushek and Woessman (2008), Glewwe and Kremer (2006), and Weil (2007).

²See Dell et al (2014) for an excellent review of this emerging literature. See Barrecca et al (2016) for health impacts, Graff-Zivin and Neidell (2013) for labor supply, Advahryu and Sudarshan (2013) for labor productivity, and Dell et al (2011), Hsiang (2011) and Park and Heal (2013) for output impacts of hot weather.

³For instance, as in the case of Lavy and Ebanstein (2015), who suggest that air pollution exposure during high stakes exams leads to allocative inefficiencies in the labor market. Notable examples of later-life impacts of educational interventions include the well-documented impacts of teacher-value added (Chetty et al, 2010, 2011), school choice (Sandstrom and Bergstrom, 2005; Deming et al, 2014; Lavy, 2015), school desegregation (Billings et al, 2014), and positive grade manipulation (Diamond and Persson, 2016).

are not caused by unobserved local determinants of educational performance and attainment. This empirical strategy – in conjunction with institutional features of New York City public schools which require students to take multiple high stakes exams spread out across several days each June – allows for the identification of the causal impact of heat stress on student performance. Since all students are assigned to test dates and locations without prior knowledge of temperature (and without the ability to reschedule), temperature on the day of an exam is unlikely to be correlated with student quality. Similarly, year-to-year fluctuations in the incidence of hot days by neighborhood are unlikely to be systematically correlated with student performance, especially when comparing the same schools over time. The breadth and depth 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.

I find that heat stress exerts a causal and economically meaningful adverse impact on student performance. Taking a NY State Regents exam – a high school exit exam that determines diploma eligibility and can influence college admission – on a hot day leads to considerably reduced performance, even when controlling for individual student ability. For the average student, heat stress during an exam reduces her score by approximately 0.22 % per° F above room temperature, such that a one standard deviation increase in test-time temperature has an effect on par with 1/8 of the Black-White score gap. Put another way, a 90°F day reduces exam performance by 0.165 standard deviations relative to a more optimal 72°F day, and leads to a 6.2 percentage point lower likelihood that the student passes the exam.⁴ I take this result as confirming the existing ergonomic and physiological literature, and consistent with recent empirical work on causal impacts of heat stress on labor supply, labor productivity, and human health (Dell et al, 2010, 2011; Zivin and Neidell, 2013; Barecca et al, 2016). It suggests that ambient heat may be an important input to the human capital production function for policymakers to consider when allocating public resources, especially in contexts where cooling technology adoption is incomplete, and where high-stakes exams pose real hurdles to further schooling.

Can heat exposure disrupt learning and human capital accumulation in the long run? Leveraging quasi-random variation in cumulative heat exposure over the course of the school year, I find that repeat heat exposure 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 by a smaller magnitude; a one standard deviation increase in the number of days above 80°F reduces Regents performance by approximately 1.4 points (se=0.70), or 0.071 standard deviations. The effect is similar in size to eliminating 70% of the gains associated with a 1

⁴That is, -2.95 points, or -4.55% relative to a mean of 64.7 points out of 100, and a reduction of -10.88% relative to a mean pass rate of 57.4%.

standard deviation increase in teacher value-added for one grade – which has been shown to increase cumulative lifetime incomes of the same NYC students by approximately \$37,000 per student, or \$925,000 per classroom (Chetty et al, 2012) – though there are many reasons why the later-life impacts of better teaching may be different from those of fewer disruptions due to heat stress. 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/or teachers), and consistent with ongoing work that finds hot days in the year leading up to SAT exams to reduce student performance (Goodman and Park, 2016).

Looking at longer-run outcomes, I find that acute heat stress during high stakes exams reduces the likelihood that a student graduates from high school on time, suggesting that even small amounts of heat stress can have knock-on impacts on educational attainment. An increase in average exam-time temperature for June Regents exams reduces the likelihood of graduating on time by 0.17 percentage points, or -1.07% per standard deviation increase in average exam-time temperature. This is consistent with a world in which acute heat exposure nudges some students to achieve less schooling overall due in part to institutional rigidities similar to those documented by Dee et al (2016) and Lavy et al (2014).⁵

Finally, I find that New York City public schools have employed various strategies to adaptively respond to heat in the classroom – but in perhaps unexpected ways. Using building-level air conditioning installation data, I find limited evidence for protective effects of air conditioning. 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 – 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). It is also consistent with the fact that even in the 62% of NYC public schools which have any AC equipment, over 40% were deemed to have defective components by independent building inspectors (BCAS, 2012). These results suggest that more careful research is needed in ascertaining the true cost-benefit of installing or improving AC equipment as an input to school production, at least in the context of old urban schools.⁶

⁵In ongoing work, I explore the potential of using temperature instruments to assess the possibility that teacher discretion carries hidden information regarding student ability and learning not captured by standardized exam scores, which may provide clues as to why many teacher value-added studies find fade out in scores but persistent long-run impacts (Cascio and Staiger, 2012; Chetty et al, 2010). In addition, I explore the potential for a “dynamic complementarities effect” of short-run heat shocks during exams, whereby realized scores (whether or not they accurately reflect underlying human capital or skill) provide signals within the educational system that lead to a dynamic reallocation of effort and subsequent reduction in overall schooling attainment, as suggested by Diamond and Persson (2016).

⁶Given the magnitude of adverse heat-related impacts documented in this study, it seems likely that

Perhaps surprisingly, I find strong evidence for what one might call “adaptive grading”, whereby teachers use their discretion in grading to partially offset the adverse performance impacts of acute heat stress, selectively boosting grades around pass/fail thresholds when students have experienced hot exam sittings. Building on work by Dee et al (2016) who use data from the same district to document systematic 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-by-subject-by-date-specific bunching estimator at passing cutoffs. I find that, while approximately 5.8% of pre-reform Regents exams were manipulated on average, the extent of manipulation varies systematically with the temperature students experienced during any given exam, with hot takes exhibiting approximately 1.5% more bunching behavior per degree F. This represents a hitherto undocumented channel of climate adaptation, and suggests that a possible unintended consequence of eliminating teacher discretion may have been to expose more low-performing students to climate-related human capital impacts, eliminating a protection that applied predominantly to low-achieving Black and Hispanic students.

This paper 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, 2013; Advahryu and Sudarshan, 2014; Cachone et al, 2013), violent crime (Anderson, 1987; Hsiang et al, 2013), and local economic output (Hsiang, 2010; Dell et al, 2011; Park and Heal, 2013; Burke et al, 2015), 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 that documents the efficacy and welfare implications of various inputs to schooling, including teacher value added (Chetty et al, 2010, 2011) and reductions in class size (Angrist and Lavy, 1997; Kreuger and Whitmore, 2001; Chetty et al, 2011), as well as school choice and desegregation (Sandstrom and Bergstrom, 2005; Deming et al, 2014; Billings et al, 2014).

While more careful research is needed to verify whether similar mechanisms are at play in developing country contexts, the findings presented here suggest that the interplay between climate and human capital may be an additional contributing factor to the long-debated correlation between hotter climates and slower growth (Mankiw, Romer, and Weil, 1990; Gallup, Sachs, Mellinger, 1999; Acemoglu, Johnson, and Robinson, 2000; Rodrik et al, 2004; Dell, Jones, and Olken, 2012; Burke et al, 2015). Moreover, to the extent that future climate change will likely result in greater added heat exposure for the poor both within and across countries (Hallegatte et al, 2015), these findings lend further support to the notion that

improving the built infrastructure of many public schools would provide a net benefit to student welfare, though considerable institutional hurdles and/or principal-agent problems may prevent socially optimal levels of AC adoption in schools.

climate change may have distributionally regressive impacts .

The rest of this paper is organized as follows. Section 2 provides a brief overview of the relevant ergonomic and economic literature on heat and human welfare. Section 3 presents a simple conceptual model of human capital production under temperature stress. In section 4, I provide a description of the institutional setting, including the main source of identifying variation, as well as a description of the data sources and key summary statistics. In Section 5, I present the main results for the short-run impact of heat stress on exam performance. Section 6 explores potential long-run impacts of heat exposure using information on cumulative heat exposure during preceding school years as well as data on high school graduation and dropout status. Section 7 provides analyses aimed at exploring the ways in which students and teachers adapt to heat stress. 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 human physiology in ways that can be detrimental to physical and cognitive performance; second, that most individuals demonstrate a revealed preference for mild temperatures close to room temperature – which is commonly taken to be between 65°F and 74 °F, or 18 °C and 23 °C – an amenity for which they are willing to pay non-trivial amounts when markets allow; 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, labor supply, and labor productivity, but not in educational performance or attainment in situ.

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 associated with increases in blood viscosity and blood cholesterol levels (Deschenes and Moretti, 2009), which can eventually cause increased morbidity in the form of heat exhaustion and stroke, the latter most acutely for the elderly (Graff Zivin and Schrader, 2015). Combined with humidity and intense physical exertion such as exercise or manual labor, heat can lead to acute cardiovascular or respiratory failure even for able-bodied adults.⁷

⁷Deschenes et al (2015) estimate that an additional day with temperatures above 90°F cause a 0.11 percent increase in county-level mortality rates in the United States.

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 the total heat released by the human body, 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).⁸

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

2.2 A Revealed Preference for Avoiding Extreme Temperatures

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 of us experience non-trivial direct disutility from being exposed to temperature extremes.

The willingness to avoid acute heat stress is perhaps most directly evident in energy markets: in particular, markets for summer-time electricity and cooling appliances. 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 (Mendelsohn and Mansur, and Morrison, 2008). On the extensive margin, and conditional on sufficient income levels, residential air conditioning ownership is closely linked to average climate. Even 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.⁹

The preference for avoiding heat exposure is evidenced also by data on time-use decisions of Americans. Using ATUS data, Graff-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, instead choosing to spend more time indoors and engaging in leisure activities, which are presumably less strenuous. Similarly, non-working

⁸Heat 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; Anderson et al, 1996). For instance, Kenrick and MacFarlane (1984) 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.

⁹The fact that air conditioning penetration varies substantially across countries (Park and Heal, 2013) and across households within countries (Gertler et al, 2015) 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 Gertler et al (2015).

individuals are shown to spend significantly less time engaging in outdoor leisure activities on hot days. These and other related findings suggest that the marginal disutility of effort is likely to be increasing in temperature as well, at least in ranges above room temperature.

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.¹⁰

2.3 Heat Stress and Task Performance

To what extent do these physiological and psychological channels lead to economically meaningful impacts on human performance?

Beginning with the early experiments of Mackworth (1947), 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 an inverted U-shaped relationship between temperature and human task performance in highly controlled environments.¹¹ Whether in the context of guiding a steering wheel, running 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.¹²

A more recent econometric literature finds a strong suggestion of 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, 2012; Barecca et al, 2016), labor supply (Graff Zivin and Neidell, 2014), labor productivity (Adhvaryu and Sudarshan, 2013; Cachone et al, 2013), violent crime (Anderson, 1987; Hsiang et al, 2013), and even local output and GDP (Dell et al, 2012; Park and Heal, 2013; Hsiang and Deryugina, 2015; Park, 2016). 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 et al, 2015).

Based on these findings, it seems clear that temperature stress – in particular heat stress – can have direct impacts on economic outcomes arising from the relationship between temperature and human physiology. The magnitude of the short-run effect on performance seems to converge to around 1% to 2% per degree Celsius above room temperature across a range of studies and output metrics in the case of labor-related outcomes (Dell et al, 2014; Heal and Park, 2016).

¹⁰An 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.

¹¹See Seppanen et al (2006) for a meta-review of the literature on temperature and task performance.

¹²There 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 (Seppanen et al, 2006).

Despite the emerging literature on the economics of extreme heat stress, the role that temperature plays in human capital development remains poorly understood. While the fact that acute heat stress can affect performance in laboratory settings is well-established, whether or not – and just as importantly, how – it might affect human capital attainment in the context of actual schools has not yet been explored.¹³ This study seeks to provide a first step toward understanding whether and to what extent temperature exerts a meaningful influence in the human capital production process, using evidence from high stakes exams in public schools.

There is some early evidence that the lab-based findings of adverse cognitive impacts from heat stress also occur in home environments. Zivin et al (2015) use NLSY survey data which includes cognitive language and math assessments administered to several thousand students at home, and find evidence for impacts of hot days on math performance but not verbal performance. However, systematic 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, 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).¹⁴ In contrast, there seems to be substantially more evidence regarding the impact of air pollution on student performance. Currie et al (2012) find substantial causal impacts of local air quality on student attendance. Lavy and Ebanstein (2015) find impacts of particulate matter during Israeli high school exit exams on student exam performance, which they trace out to impacts on later life earnings.¹⁵ Goodman (2014) shows that snowfall can result in disruptions to learning by increasing absenteeism selectively across different student groups.

¹³This 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>.

¹⁴To 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.

¹⁵Peet (2015) uses weather variation, including but not limited to extreme temperature, as an instrument 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.

3 A Model of Human Capital Accumulation under Temperature Stress

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 given optimizing agents in a Mincerian framework.

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.¹⁶ 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. For 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.

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, 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 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).

¹⁶ e_i may also include direct costs of schooling such as the cost of books and tuition.

¹⁷For instance, temperatures of 90 degrees F and 80 degrees F will correspond to $T(90) > T(80) > 0$, whereas 72 degrees 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.

Similarly, one might expect any given exam score to be influenced by this short-run cognitive impact of temperature stress if temperature in the classroom is elevated during an exam. Let $s_{it}(h_{it}) = (1 - \beta_T T) \cdot h_{it} + \epsilon_t$ 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 thus 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 The Impact of Temperature on Student Performance and Schooling Investment without Adaptive Responses

Suppose heat adversely affects cognition, has a weakly negative direct effect on utility, or increases the marginal cost of additional effort, all of which are suggested by the existing literature. That is, suppose $U'(T) \leq 0$, and/or $U_T \leq 0$, and/or $U_{eT} \leq 0$. Then it is easy to see intuitively that elevated temperatures in school will reduce student welfare, unless avoidance behavior is cost-less and complete.

We can write student i 's indirect utility function as:

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

The optimal schooling investment is given by the familiar first order condition, $\frac{dU}{de} = 0$, or:

$$U_c w(1 - \beta_T T) + U_e = 0 \quad (4)$$

The intuition is that the student will invest in schooling to the point where marginal pecuniary benefits equal marginal pecuniary costs, given the prevailing extent of heat stress T .

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.

Note that, in the absence of adaptive technologies or margins of effective avoidance behavior, the welfare impact of increased heat stress is necessarily negative, as long as one of $U_T < 0$, or $U_{eT} < 0$ or $a'(T) < 0$ are true as strict inequalities.²⁰

Proposition 1: If $a'(T) \leq 0$, and/or $U_T \leq 0$, and/or $U_{eT} \leq 0$, then $\frac{\Delta U}{\Delta T} \leq 0$. That is, if heat stress causes cognitive decline or direct disutility or an increase in the marginal cost of effort, then we would expect any additional heat stress in the classroom to have a non-positive impact on overall welfare.

By a parallel logic, in the absence of completely protective adaptive technologies or margins of effective avoidance behavior, the effect of acute heat exposure on any given exam score will be strictly negative in expectation: $E(\frac{\Delta s_{it}}{\Delta T_t}) < 0$.

Proposition 2: If $a'(T) \leq 0$, and/or $U_T \leq 0$, and/or $U_{eT} \leq 0$, then $E(\frac{\Delta s_{it}}{\Delta T_t}) < 0$: i.e. heat stress during an exam will reduce student performance.

What about long-run human capital accumulation? While it is possible in principle that students may respond to added heat stress and the cognitive decline/direct disutility it entails by increasing effort, struggling to offset the negative impact on their schooling achievement, one can show that such compensatory effort will occur only in the presence of very strong income effects.

Appealing to the implicit function theorem, one can differentiate the first order condition above with respect to temperature to arrive at the following comparative static:

$$\frac{\Delta e_i}{\Delta T} = \frac{w^2(1 - \beta_T T)eU_{cc} - w(1 - \beta_T T)U_{cT} + wU_c + weU_{ec} + U_{eT}}{(w(1 - \beta_T T))^2 U_{cc} + w(1 - \beta_T T)U_{ce} + w(1 - \beta_T T)U_{ec} + U_{ee}} \quad (5)$$

which, given reasonable assumptions about preferences is strictly non-positive.

Proposition 3: The total change in schooling effort investment due to elevated classroom temperatures is necessarily non-positive, $\frac{\Delta e_i}{\Delta T} \leq 0$, implying that the impact of heat stress on human capital accumulation is strictly non-positive: $\frac{\Delta h}{\Delta T} \leq 0$.

If heat exposure at school reduces the effectiveness of any incremental unit of effort or time in accumulating knowledge and skills, while not increasing the marginal disutility of effort (or opportunity cost of schooling time), then the substitution effect pushes students to allocate less effort (and/or time) in schooling. As long as the reduction in income that arises due to added heat stress (and the direct loss of learning that it entails) is not large enough to

²⁰This is because schooling effort involves direct pecuniary costs $U_e < 0$, and $\frac{\delta c}{\delta T} \leq 0$, due to the adverse impact of heat stress on the marginal productivity of schooling effort (and $U_c > 0$ by assumption).

push the student into dire poverty – i.e. very high marginal utility of consumption such that the income effect dominates – then one would expect the net effect of added heat stress on schooling investments to be negative: that is, to reduce the optimal schooling effort (and/or time) investment by students. In short, unless there are very strong income effects which dominate direct disutility and the substitution effect of making consumption more expensive from an effort standpoint due to reduced productivity, we would not expect added heat stress to increase the amount of optimal schooling effort investment by students.

3.3 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. Generally speaking, individuals may in principle reschedule strenuous activities during cooler times of day, as many in sub-tropical 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. Some such adaptive responses, in particular air conditioning, have been documented to be highly effective in blunting the acute effects of heat-related impacts (for instance, in the context of human health).²¹

In the case of students in school, however, it is unclear how much adaptive behavior is actually feasible given common constraints on student activities.

Before the advent of air conditioning, the most relevant responses to heat stress in the classroom were likely to reschedule classes, send students home on summer vacation, or simply double down and bear any potentially adverse pedagogical consequences.

Even now, a typical elementary, middle, or high school student cannot, for instance, install an air conditioner in her classroom, even if she can afford it financially. Nor, in most cases, can her parents, however altruistic in their motives or 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 deleterious learning impacts.²²

At the same time, there may exist margins of adaptation that are unique to school environments. To the extent that teachers and administrators have some discretion in grading

²¹Barecca et al (2016) find that air conditioning was highly effective at reducing the acute impacts of heat stress on mortality in the United States. Isen et al (2015) find evidence suggestive of a protective impact from air conditioning in the context of in-utero health.

²²For instance, in the case of New York City public schools, air conditioners must meet efficiency standards and be obtained from and installed by a specific vendor chosen by the city, in addition to receiving city approval with regard to a variety of safety regulations, contractual obligations and energy considerations. In some cases, school “sustainability” policies prohibit administrators from investing in new infrastructure unless it can be demonstrated that it has a net neutral impact on carbon emissions, a bar that new air conditioning cannot clear unless electricity is obtained completely from renewable sources.

exams or applying institutional rules regarding graduation, it is possible that teachers can use their discretion to 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. Dee 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 7.

Suppose students can engage in avoidance behaviors specific to heat stress 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. The student's value function becomes:

$$V_i(e_i^*, T, k_i^*) = \max_{e_i, k_i} U_i(w(1 - \frac{\beta_T T}{1 + k_i}) \cdot e_i, e_i, T, k_i) \quad (6)$$

where e_i^* and k_i^* denote optimal investments in schooling effort and avoidance behavior/adaptive capital respectively. The first order conditions are:

$$\frac{dU}{de} = 0 \Leftrightarrow w(1 - \frac{\beta_T T}{1 + k_i})U_c = -U_e \quad (7)$$

and

$$\frac{dU}{dk} = 0 \Leftrightarrow \frac{\beta_T T w e_i}{(1 + k_i)^2} U_c = -U_k. \quad (8)$$

Substituting $\frac{-U_e}{w(1 - \frac{\beta_T T}{1 + k_i})}$ for U_c , we get:

$$\frac{-\beta_T T e_i U_e}{(1 + k_i)^2 (1 - \frac{\beta_T T}{1 + k_i})} U_c = -U_k. \quad (9)$$

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 responsiveness or punitiveness of parents, teachers, and school administrators. This 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.

The central takeaway is 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. That is, the availability of adaptive technologies in principle does not imply their full utilization in response to environmental stressors in practice. The intuition appeals to the same insight that arises from most models of pollution control where, so long as the cost of abatement is non-zero, the socially optimal level of pollution is not zero (i.e. optimal mitigation is not infinite).

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 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 by the poor (Lavy, Ebanstein, and Roth, 2014; Currie et al, 2015). In 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 .

3.4 Empirical Predictions

The main empirical predictions of the model are as follows:

1. We expect acute heat stress to reduce exam performance, 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, so long as the cost of these technologies is non-zero.
2. Heat stress in school may reduce educational attainment in the long run, 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.

3. We expect agents – students, teachers, and parents – to adapt along the most cost-effective of available margins, mitigating the realized effect of heat stress on educational performance and attainment, but that unless adaptation technology is completely costless, the extent of adaptation will be incomplete.

4 Institutional Context, Data, and Summary Statistics

This section describes the institutional context for this study as well as the data used to assess the impact of heat stress on human capital production. New York City public schools provide a suitable setting for studying the effect of temperature on human capital production for several reasons: namely, the centralized institutional setting (in particular vis-a-vis high school exit exams), its variable climate, and its diverse student body, which combined allow credible causal identification of heat-related performance and attainment impacts on a population that is representative of many other urban public school districts.

4.1 New York City Public Schools

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

It is ethnically and socioeconomically diverse, with a student body that is 40% Latino, 31% African American, and 14% Asian, with 78% of students qualifying for federally subsidized school lunch as of 2014. While the median student is relatively low-performing (e.g. with high school achievement that does not meet “college proficiency” standards) and low-income, a substantial minority come from wealthy backgrounds and attend high-achieving magnet schools including Stuyvesant Academy and Bronx Science, which consistently rank among the nation’s best. This results in a wide representation of student types by demographic and achievement categories in the sample used in this study, which comprises high school sophomores, juniors, and seniors who take Regents exams from all public middle and high schools.²³

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%). Once again, system-wide averages mask remarkable discrepancies in achievement across neighborhoods.

²³Approximately 19% of NYC students attend private schools – in particular, residents of the Upper East Side of Manhattan (70-80%) – and are thus not included in our sample.

Schools in the predominantly Black or Hispanic neighborhoods of Brooklyn and the Bronx have four-year graduation rates as low as 35% per year (Figure 1).

4.2 New York State Regents Exams

Students in NYC public schools take a series of standardized high stakes exams – “Regents exams” – which represent the primary measure of student performance used in this study. Administered by the New York State Education Department (NYSED), and more commonly known as high school exit exams, they comprise a series of standardized subject assessments which are taken within schools primarily at the conclusion of the school year.

Regents exams are high stakes exams for the average NYC student. Students are required to meet minimal proficiency status – usually a scale score of 65 out of 100 – on Regents examinations in five “core” subject areas to graduate from high school: English, Mathematics, Science, U.S. History and Government, and Global History and Geography.²⁴ In addition, many public universities such as City University of New York (CUNY) apply strict Regents score cutoffs in college admissions decisions.

The vast majority of students take their Regents exams during a 5 to 10 day window in mid-to-late June each year.²⁵ The test dates and times for each of these Regents exams are determined over a year in advance by the NY State education authority (NYSED), and synchronized across schools in the NYC public school system to prevent cheating. Each exam is approximately 3 hours long, with a morning session that begins at 9:15am and an afternoon session beginning at 1:15pm.²⁶ Figure 32 in the Appendix shows a typical June exam schedule, and Figure 33 in the Appendix provides some typical Regents exam questions, both of which are taken from the NYC DOE website. Students who fail their exams (or are deemed unready by their teachers to progress to the next grade) are required to attend

²⁴These 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.

²⁵Most students will take at least one Regents exam each year of high school, usually in the spring of 10th, 11th, and 12th grade, and usually immediately after they have taken the corresponding class. Apart from the fact that most students take English their junior year, and Living Environment and Global History prior to other “advanced” sciences and US History respectively, there are to my knowledge no clear regularities in the timing of various subject exams throughout students’ high school careers. Some advanced students may take Regents subject exams during middle school.

²⁶The 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.

summer school, which occurs in July and August.

Because daytime high temperatures in June frequently exceed 90°F (32°C), it is not uncommon for students to be exposed to disruptive temperatures on exam day. It is also possible for students to be exposed to heat stress during class time in the semesters leading up to the exams, since the school year commences in early September – when average daytime highs are around 76°F, and frequently veer above 90°F – and ends during the last week of June. The record daytime high temperatures for June and September in Manhattan are 101°F and 102 °F respectively.

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 each year in advance of the exams. 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.²⁷

Exams were, however, locally graded by teachers in the students’ home schools, at least until grading reforms were implemented in 2011. As has been documented by Dee et al (2016), a substantial portion of NYC Regents exams featured bunching at passing cutoffs, clear evidence for discretionary grade manipulation by teachers. 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, which means, among other things, that they may provide information about compensating behavior that is not available in low-stakes laboratory studies. Second, they are offered at a time of year when temperatures are likely to be hot but not always so, due to the considerable variability in day-to-day temperatures in June. Because they occur at the end of the school year, they are also more likely than periodic assessments to reflect cumulative impacts of heat stress that may have accrued over the course of the school year. Third, they are taken by a diverse mix of students, as opposed to by high- (or low-) performing subgroups alone as is the case of, for instance, AP or SAT exams, which allows for rich heterogeneity analysis and more likely permitting out-of-sample validity. Fourth, Regents exams were centrally administered and compulsory for all public school students during the study period, meaning there is relatively little possibility of anticipatory alteration of exam timing based on weather forecasts, or for bias due to selection into taking the exam or by selective grading criteria. Fifth, location and timing information for each student can be matched to ambient temperature data for that

²⁷In 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), though there are many reasons to believe this may not necessarily be the case in practice, leading us to prefer a model that incorporates year fixed effects when possible.

day, as well as to air conditioning data for the school in which the students were enrolled.

4.3 The Climate (and Micro-climates) of New York City

To what extent does heat stress matter for students in NYC public schools, and what are the factors that the econometrician must take into account in attempting to assess the impact of heat on school outcomes using observational data (as opposed to a controlled lab experiment)?

June temperatures in NYC vary considerably – both day to day and year to year, providing a substantial amount of quasi-experimental variation. During the test dates in the sample, daytime high temperatures range between 65°F and 98°F. Figure 2 plots the day to day variation in daytime high temperatures across a typical June for the five boroughs. Figure 3 shows the variation in outdoor temperature by school across two test dates within the sample period.²⁸ Temperature varies diurnally as well, especially during transitional months such as June, which must be taken into account since exams are taken both in the morning as well as during the afternoon.

Despite limited geographic extent, New York City also exhibits substantial urban heat island effects due to its highly urbanized and variegated local environment. Local climatic factors such as urban heat islands may be of particular importance in studying the potential welfare impacts of climate change, given the density of schools in urban contexts (Rosenzweig et al, 2006). On a given summer day, experienced 2 meter surface temperature can vary by as much as 20 °F within the 5 boroughs of NYC.²⁹

4.4 Weather Data - Spatial and Temporal Corrections

Unless one is able to observe the experienced temperature of students in the actual test-taking environment, readings from local weather monitors will provide at best very crude indicators of the temperature variable of interest.

To best approximate experienced temperature, I take daily station level data from NOAA, which provides daily min, max, and mean temperatures, precipitation (in millimeters) and dew point information from a national network of several thousand weather stations over the period 1950-2014, and perform a series of corrections. I take only official US weather service stations in the vicinity of the five boroughs of New York City and which were available for the entirety of the sample period (1998-2011). I adjust for diurnal temperature variation by fitting a quartic hourly temperature model using data from June 1998-2011, creating temperature

²⁸Summer temperatures can get substantially hotter. While not explored in detail in this paper, this might matter for distributional equity, as summer school is primarily remedial, an institutional that is not limited to NYC public schools.

²⁹Local environmental factors including surface albedo, tree cover, proximity to bodies of water/ocean, and exhaust from other air conditioned buildings all contribute to substantial differences in micro-climates within the city proper. Figure 29 is taken from (USGS(2014)) and shows 2m surface temperature for August 21, 2001, estimated using satellite infrared and a local climate model.

variables corresponding to midpoint of the morning and afternoon exam sessions: 10:30am and 2:30pm respectively.

Combining this data with a local climate model and satellite reanalysis output provided by Gaffin et al (2012), which provides near surface temperature at 30m by 30m resolution for a representative summer day (8/21/2001), I generate spatially and temporally corrected temperature readings by latitude and longitude, date, and time (morning vs afternoon).³⁰

These variables are matched geographically using street addresses for each of the 957 school buildings in my sample.

4.4.1 Measuring Cumulative (School-year) Heat Exposure

To measure the cumulative temperature stress experienced by most students while they are in school and engaged in learning, I measure the number of hot days during the school year for the years 1997-98 through 2010-2011, using start and end dates for each of the school years in our sample.³¹ These weather variables are constructed by summing extreme heat day indicator variables over the course of a school year leading up to the first day of exams:

$$DD\pi_T^\circ = \sum_{d=SS_{T-1}}^{365} DD\pi_{T-1,d}^\circ + \sum_{d=1}^{SE_T} DD\pi_{T,d}^\circ \quad (10)$$

Here, d is day of the year T , SS_{T-1} denotes the school year start date of the prior calendar year, and SE_T represents the effective school year end date in that calendar year, which I take as the first day of Regents exams that year. The indicator variables thus denote whether a particular day had maximum temperatures in a temperature range π° , and are divided into 10°F increments over the school year (e.g. 80°F-90°F, 90°F-100°F).

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

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. The average number of extremely hot days with temperatures above 90°F is 2.5, with a standard deviation of 1.3. Most of these

³⁰A key assumption in the spatial correction is that the spatial variability in temperature during summer months (June through August) is stable across time. In effect, this does not influence causal identification because all identification strategies feature some form of location fixed effect. I verify that the spatial correction reduces standard errors and increases point estimate slightly, by presumably reducing measurement error. As a sanity check, I also create a variable that simply assigns nearest USGS official station reading by school and date-time. The main results are not sensitive to using this uncorrected temperature measure, though standard errors are somewhat larger.

³¹School year start and end dates, which vary each year but tend to be from early September to late June, are taken from the NYC DOE website.

days occur during the months of September, October, and June. Summer school students are, on average, subject to an additional 9 days above 90°F. Despite documented warming for the US and the world as a whole over the past several decades, temperatures in the NYC area seem to have remained relatively stable over the study period (tests for stationarity and trend-stationarity do not suggest time trends in these extreme heat day variables).

4.5 Student-Level Administrative Data (NYC DOE)

I obtain student-level information from the New York City Department of Education (NYC DOE). The data is de-identified to protect students' privacy, and includes the universe of all public school students who took one or more Regents exams over the period 1999 to 2014. Each year includes approximately 75,000 students, approximately 1.2 million in total over the period 1999-2014, with students represented by unique 9-digit ID numbers.

The data includes exam scores recorded at the student-, subject-, year-, and term-level. Each cell includes information about the school in which the student was enrolled during the year in which she took the examination observed. The data also includes information on 4-, 5-, and 6-year graduation status by student, for those students who entered high school after the beginning of the 2004 school year, which includes approximately 450,000 of the 1.2 million students for whom exam scores are available.

Information on student ethnicity, gender, and federally subsidized school lunch eligibility (a measure of student SES) is available for students in the 2004 and above entering cohorts, as well as for students who are observed in a separate data set that comprises standardized state English (ELA) and math assessments for grades 3 through 8. Combining these two sources, demographic information, in addition to information on previous student ability as measured by previous standardized ELA and math performance, is available for approximately 90% of the students in the sample.

Throughout the study period, students typically took Regents exams at the school in which they were enrolled, according to official NYC DOE policies, unless they required special accommodations which were not available at their home school.³²

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, which provide the date and time of each NY Regents exam taken by NYC public school students over the past two decades. Due to inconsistencies in the way exam subjects and terms are coded in later years, however, I drop exam records for years 2012 through 2014, and use only the records of exams taken in the years 1999 to 2011. I retain data from these latter

³²To the extent that some students took their exams at different locations than their home school, we would expect additional measurement error in the spatial dimension of the temperature variable, though not in the temporal dimension, since exam dates are uniform across the city (and state).

years when replicating the Dee et al (2016) bunching estimators and evaluating the efficacy of the NYC grading reforms of 2011-2012, since the analysis does not require accurate exam dates to conduct. The exact matching process, in addition to the rationale for limiting the sample, is described in the Appendix.

4.6 Summary Statistics

The final working dataset consists of 4,509,102 exam records for 999,582 students from 947 different middle and high schools. The sample comprises data from 91 different exam sessions (date by time) pertaining to the core Regents subjects (11 in total) over the 13 year period from 1998-1999 to 2010-2011. Less than 0.2% of students are marked as having been absent on the day of the exam, corroborating the high-stakes, compulsory nature of these exams.

Tables 6, 7, 8 and 9 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, and a standard deviation of 17.9, though there is considerable heterogeneity by borough as well as student type. African American and Hispanic students tend to perform substantially worse than Whites and Asians, with average scores of 61.2 and 61.5 and 72.9 and 74.7 respectively, or between 0.65 and 0.75 standard deviations worse on average. The within-school differences are not surprisingly smaller, with Blacks and Hispanics performing on average 4.9 and 4.4 points below Whites respectively. Girls tend to perform slightly better than boys, as do students who are not eligible for federally subsidized school lunches (higher SES).

I calculate the average combined z-score of each student for whom previous standardized ELA and math exam records are available. This measure is used in the regression analyses that follow as a supplementary control in lieu of student fixed effects. Combined z-scores are constructed by computing standardized z-scores by subject and year, and computing the annual average by student. Figure 31 in the Appendix motivates the use of previous ELA/math z-scores as an alternative measure of student ability (in lieu of a student fixed effect), plotting average Regents performance on previous combined ELA and math z-score by student and suggesting a tight correlation.

The average student takes 6.6 June Regents exams over the course of her high school career, and is observed in the data for between 2 and 3 years, though many under-achieving students are observed for more than 5 years, as they continue to retake exams upon failing. NYC students tend to do better 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.

Graduation rates have been improving slightly over time, but remain lower than the national average. Not surprisingly, graduation status and Regents performance are highly

correlated, even when controlling for ethnicity, gender, and previous z-scores, suggesting that reduced performance on these high stakes exams can affect long-run educational attainment even if underlying human capital has not changed.

4.7 Grade Manipulation by NYC Teachers

Dee et al (2016) document systematic evidence for grade manipulation by NYC teachers on NY State Regents exams, using a similar dataset from 2003 to 2012. They suggest most of the manipulating behavior occurred at or around passing margin of 65 and while varied in magnitude across schools and student types, was a near-universal “cultural” phenomenon within the NYC schools system. Reforms put in place in 2012 effectively eliminated such behavior, though no studies to date have attempted to link this behavior to climatic factors such as exam-time temperature.

Figure 5 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 a score of 65. Dee et al (2016) use a sixth-order polynomial bunching estimator to suggest that approximately 6% of all Regents scores prior to 2012 and 40-50% within the manipulable region were manipulated. I replicate and expand on these findings below, using a similar bunching estimator at the school-by-subject-and-take level, described in greater detail in section 7.

How does this affect our estimation strategy? As discussed in more detail below, such upward grade manipulation likely results in downward bias of heat-related performance impacts. To the extent that the degree of manipulation is uncorrelated with exam time temperature (and likelihood of heat exposure is uncorrelated with likelihood of scoring within the manipulable zone), we would expect this behavior to bias the estimate downward, since all students will experience upward score revisions ‘indiscriminately’.

To the extent that teachers *selectively* manipulated grades in response to experienced exam-time temperatures, the score distributions will be endogenous to temperature, and the estimated impact of beta will again be biased downward.³³

An interesting possibility is that teachers engage in *more positive* grade manipulation for exams that experienced more heat. That is, it could be the case that teachers are aware of the potential knock-on impacts of failing a Regents exam despite having the knowledge or skills

³³The 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.

to deserve a passing grade (and/or are motivated by a sense of fairness), and thus provide a buffer to heat-related educational impacts by manipulating grades selectively based on environmental conditions. I explore the possibility of this form “adaptive grading” in section 7.2.

Empirical Strategy and Primary Results

The following sections present the primary empirical strategy and results of the paper. First, I describe the strategy for identifying causal impacts of acute heat exposure on contemporaneous student performance and present the primary results from this “short-run” analysis, focusing on the effect of exam-time ambient temperature on Regents exam performance. I then present an analysis of potential long-run impacts of heat stress on student attainment, using both short-run and cumulative heat stress as sources of identifying variation. Finally, I explore whether and how students and teachers adapt to heat stress in school settings, using building-level air conditioning installation data, as well as a version of the bunching estimator developed by Chetty et al (2011) and applied to Regents exams by Dee et al (2016).

5 Short-Run Impacts: Does Heat Stress Affect Student Performance?

5.1 Regents Scores

Figure 10 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, controlling for average differences across subjects, average differences across years, and exam-day precipitation and humidity. The plot strongly suggests that exams taken on hot days exhibit lower scores.

While temperature is likely not endogenous to student behavior, nor is it likely that there is differential selection into various day-to-day temperature “treatments” based on unobserved student characteristics, it is in theory possible that hot temperature and unobserved determinants of student performance are correlated. 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. Similarly, if certain exams tend to be scheduled more often toward the end of the exam period (Thursday as opposed to Monday), and student type and subject are correlated, we may be concerned about bias arising from this correlation.

To further isolate the causal impact of temperature 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. Specifically, I estimate a baseline model

that includes student, year, and subject fixed effects, as well as controls for time of day, day of week, and day of month:

$$Y_{ijsty} = \beta_0 + \beta_1 T_{jsty} + \beta_2 X_{jsty} + Time_{sty} + DOW_{sty} + DOM_{sty} + \gamma_i + \eta_s + \theta_y + \epsilon_{ijsty} \quad (11)$$

where Y_{ijsty} denotes exam score (0-100) for student i , taking subject exam s in school j on date t in year y . T_{jsty} is the temperature experienced by school j during the exam (subject s on date t , year y). X_{jsty} is a school-by-date-specific vector of weather controls, which include precipitation and dewpoint. $Time_{sty}$, DOW_{sty} , and DOM_{sty} represent fixed effects for time of day (Time=1 denotes afternoon exam), day of week, and day of month respectively. The terms γ_i , η_s and θ_y denote student, subject, and year fixed effects respectively. Student fixed effects ensure that we are comparing the performance of the same student across different exam sittings, some of which may be taken on hot days, others not, leveraging the fact that the average student takes approximately 7 June Regents exams over the course of their high school career. Subject fixed effects control for persistent differences in average scores across subjects (for instance, NYC students tend to perform markedly worse on Global History and Government than they do on US History and Government, see Table 8). Year fixed effects control for possible spurious correlation between secular performance improvements and global warming.

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 a lower bound estimate of the causal impact of temperature on exam performance (β_T from the model presented in section 3), subject to the attenuation bias due to measurement error in weather variables discussed above.

Table 10 presents the results from running variations of equation (11) with and without fixed effects for time of day, day of week, and day of month. As suggested by the first row of columns (1)-(3), exam-time heat stress seems to exert a significant causal impact on student performance. The estimates are robust to allowing for arbitrary autocorrelation of error terms within boroughs and test dates, with standard errors clustered at the borough-by-date-and-time level, which is the level of exogenous temperature shock.

Taking an exam amid heat stress reduces Regents scale scores by approximately 0.145 points (se=0.035) per degree F, which amounts to a 0.22% decline relative to a sample mean value of 64.8 points. This translates into approximately -0.94 points (-1.45% decline) per standard deviation of test-day temperature, or -2.95 points (-4.55% decline) if a student takes an exam on a 90°F day as opposed to a more optimal 72°F day. The effect of a

90°F day is thus comparable in magnitude to roughly 1/4 of the Black-White score gap (3/4 of the within-school Black-White score gap). It is comparable to though slightly smaller than the effect sizes found in laboratory studies that explore cognitive performance impacts, which, according to a recent meta-review (Seppanen et al, 2008) clustered around -0.6% per degree F, perhaps because the stakes are higher and students are thus more willing to absorb the disutility cost of additional effort under heat stress compared to subjects enrolled in laboratory experiments, or simply due to attenuation bias from measurement error.

These effect sizes are comparable also to those found by Zivin et al (2015) on NLSY home math assessments, and the effects on Israeli high school exit exams of a standard deviation increase in pm2.5 and CO pollution found by Lavy and Ebanstein (2015). 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.

It is worth noting that precipitation seems to have a consistently positive impact on performance. This may be due to the strong correlation between precipitation and air quality, as rain can act to cleanse the air of particulate matter and other pollutants, which have been shown to affect cognitive function and student performance in similar contexts (Lavy and Ebanstein, 2015; Currie et al, 2012).³⁴

The results are robust to a model that replaces student fixed effects with school-by-year fixed effects, and controls for student ability by using average combined z-scores from previous standardized ELA and Math exams (3rd through 8th grade), in addition to a full suite of observable demographic controls including ethnicity, gender, and federally subsidized school lunch eligibility. In this case, students who move schools are assigned the modal school ID – that is, the school in which they spend the most years.³⁵ These results are presented in Table 30 (Appendix).

5.2 Pass Rates and College Proficiency

What, if any, are the educational consequences for students of the heat-related performance impacts described above? 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.

As mentioned previously, students must score a 65 or above on any given Regents subject

³⁴In ongoing work, I assess the potential impact of particulate matter and ozone on student performance, controlling for temperature.

³⁵If 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.

exam to pass the subject and thus have it count toward receiving a HS diploma.³⁶ 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 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.³⁷

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 previously, 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:

$$p_{ijsty} = \beta_0 + \beta_1 T_{jsty} + \beta_2 X_{jsty} + Time_{sty} + DOW_{sty} + DOM_{sty} + \gamma_i + \eta_s + \theta_y + \epsilon_{ijsty} \quad (12)$$

$$c_{ijsty} = \beta_0 + \beta_1 T_{jsty} + \beta_2 X_{jsty} + Time_{sty} + DOW_{sty} + DOM_{sty} + \gamma_i + \eta_s + \theta_y + \epsilon_{ijsty} \quad (13)$$

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 12 and 13 report the results from running variations of equations 12 and 13 that

³⁶Until 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.

³⁷The 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.

include subject, year, student, time of day, day of week, and day of month 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 the well-documented grade manipulation around the passing threshold. Taken together, 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.

Using these point estimates, and the observed score distributions, it is possible to estimate a lower bound on the number of students who were pushed below the passing threshold due to heat stress. Suppose heat stress affects all students equally – that is, the impact of heat stress is uniform across the potential score distribution. Assume also, for the time being, that there is no grade manipulation by teachers. Then one would in principle be able to calculate the number of students who fall below the pass threshold due to heat stress by integrating the fraction of students who would have scored between 65 and $65 + \Delta T \cdot \beta_T$ in the hypothetical “without heat stress” distribution. If the impact of heat is a spread-preserving shift of the entire score distribution of $-\Delta T \cdot \beta_T$, then one can estimate the number of students who are pushed below the pass threshold due to temperature by integrating the students who score between $65 - \Delta T \cdot \beta_T$ and 65 in the observed distribution. To the extent that grade manipulation pushes some students who would have scored between $65 - \Delta T \cdot \beta_T$ and 65 above the passing threshold, performing this calculation using the observed “manipulation-inclusive” distribution will provide a lower bound. By this method, I calculate that, between 1998 and 2011, at least 510,000 additional exams received a failing grade as opposed to a passing grade due to heat exposure, which implies that at least 90,000 students were affected in this way.³⁸ These estimates, though crude, are suggestive of the potential magnitude of heat-related disruptions to educational attainment.

Considering that NYC and many other public school systems administer high stakes exams with numerical (cardinal) score cutoffs, and that, once students fail a particular exam, they must enroll in summer school and wait until the ensuing August to retake it, it is possible

³⁸To arrive at this figure, I integrate the number of students in the “exposed to heat” portion of the test score distribution multiplied by the likelihood that the amount of heat experienced (distance from the optimal temperature, which I take, conservatively, to be 72°F) results in a score that is below passing. I take the average deviation from 72°F experienced by the average student across all takes and students in the study period as the measure of ΔT .

that even short-run heat exposure during a few exam days can have lasting consequences for final schooling attainment, a possibility explored in section 6.

Table 23 presents the results from running equation 11 by demographic sub-group (and with the full suite of interaction terms between demographic identifiers and temperature): namely, ethnicity, gender, and eligibility for federally subsidized school lunch. The estimates suggest that black students may be marginally more sensitive to an incremental increase in heat stress, but the results do not appear to be significantly different. Running the analysis using pass and proficiency dummies as the dependent variable provide some limited evidence that black students suffer larger aggregate consequences of heat-related performance impacts, given the higher likelihood that they score at or near pass/fail cutoffs.

6 Long-Run Impacts: Does Heat Stress Affect Human Capital Attainment?

The previous analyses suggest that short-run heat stress exerts a causal and statistically significant impact on student performance in high stakes, real-world school settings.

This begs the question of whether short periods of heat exposure can lead to lasting impacts on human capital attainment and their attendant effects on later-life labor market and other outcomes (e.g. Chetty et al, 2011), or whether they represent transient shocks to scores – added noise in the signal extraction process of schooling (Lavy and Ebanstein, 2015). To the extent that the actual learning that these exams are intended to assess occurs during the days and months prior to the exam itself, one may wish to assess whether cumulative heat exposure during class time reduces the amount and rate of human capital accumulation.

I explore these potential long-run impacts of heat exposure on human capital attainment using two approaches:

First, building on the model intuition from section 3 – that heat may reduce the effective amount of pedagogical engagement during class time – and leveraging year to year and cross-sectional variation cumulative heat exposure within NYC public schools, I assess whether cumulative heat exposure over the course of a school year effects end of year exam performance.

Second, noting the common institutional and resource constraints faced by most students in public education system, I assess whether acute heat stress during high stakes exams affects the likelihood that a student graduates from high school. That is, whether short-run heat stress can have “knock-on effects” on educational attainment, where “knock-on effects” are defined as impacts on realized schooling attainment that arise from temporary shocks to cognition which presumably do not reduce the level of human capital directly.³⁹

³⁹ As I describe in greater detail below, in an idealized, frictionless world, a score shock should be reversible

6.1 Cumulative Learning Impacts of Hot Days during the School Year

In addition to the effects documented above, the model in section 3.2 predicts that cumulative heat exposure during schooling may affect human capital production as well, in part by reducing the effectiveness of any given hour or day of educational engagement. I take results from section 5 as highly suggestive that heat stress affects students’ cognitive ability and pedagogical engagement, which suggests also that heat exposure during class time may have a disruptive effect on learning.

Figure 17 presents a binned scatterplot of Regents score on the number of days above 80°F, controlling for exam-day temperature and precipitation, as well as school-, subject-, time of day, day of week, and day of month 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 are 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 they are 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. This is for a couple of reasons. Because I do not observe student-specific measures of cumulative heat exposure over the preceding school year, keeping student-level exam observations will likely introduce additional measurement error, since cumulative heat exposure during preceding school years is measured at the school level and some students may have been present for more days than others or live in neighborhoods that are more prone to heat stress than others. 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. Recall that the main source of identifying variation for short-run impacts is day-to-day (and AM vs PM) variation in exam-time temperature within student-year cells, which is best approximated in this case by variation across subjects within school-year cells.

with enough retakes, and so one should not observe such “knock-on impacts” of random acute heat exposure. The presence of “knock on impacts”, if credibly identified, may reflect a combination of institutional rigidities or dynamic complementarities in learning.

I thus estimate variations of the following model:

$$y_{jsty} = \beta_0 + \beta_1 T_{jsty} + \beta_2 X_{jsty} + \Sigma_d^D \beta_2^d DD_{jT}^d + \chi_j + \eta_s + Z_{jsty} \\ + Time_{sty} + DOW_{sty} + DOM_{sty} + \beta_3 Year_y + \beta_4 Year_y^2 + \beta_5 Year_y^3 + \epsilon_{jsty} \quad (14)$$

where y_{jsty} denotes the average Regents score (0-100) 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 ; χ_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. $Time_{sty}$, DOW_{sty} , and DOM_{sty} represent fixed effects for time of day (Time=1 denotes afternoon exam), day of week, and day of month respectively; and $Year_y \dots Year_y^3$ 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. Demographic controls at the school-take level capture possible correlation between demographic composition of the students taking particular subjects and temperature, due to potential correlation between the timing of certain subjects and temperature that are not accounted for by time of day, day of week, and day of month fixed effects.

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.⁴⁰ I use a number of bin classifications for hot days, motivated by the existing literature (e.g. Deschenes et al, 2015; Hsiang and Deryugina, 2015) as well as the analyses presented in the previous section, which find negative impacts of heat stress beginning around 70°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 suggests to be 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. 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, the latter of which is taken from weather station readings in Central Park and assigned uniformly across all schools in the city.

⁴⁰School 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.

I run variations of equation 14 that use different weather controls to allow for a flexible characterization of the “reference category” against which we can interpret the impact of hot days. Table 18 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 18, 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 in section 5. Focusing on column (3), which controls for average daily max temperatures during the school year, and controlling for the contemporaneous effect, we can see that days between 70 and 80°F and 80° and 90°F have a negative impact of of -0.10 (se=0.058) and -0.23 points (se=0.09) points respectively, which represent -0.16% and -0.37% reductions relative to a sample mean of 62.3 points. 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.⁴¹

These estimates suggest that a one standard deviation (3.91 days) increase in the number of days with maximum temperatures above 80°F can reduce learning by approximately 1.4%, or 0.071 standard deviations, as measured by end-of-year exam performance. These impacts are on par with the learning impacts of a 0.7 standard deviation reduction in average teacher value-added (Chetty et al, 2011), or 3/4 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 sum, 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 (2011) which examine the same population of NYC students and find significant impacts of improved learning on later-life outcomes.

⁴¹Cold 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.

6.2 Knock-on Impacts of Acute Heat Stress on Graduation Status

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.⁴² 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 reflect permanently reduced human capital: $\Delta s_{it} \neq \Delta h_{it}$.

In an idealized, friction-less world with fully flexible institutions, unlucky students who experience a hot exam sitting would 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 a) limit the effective number of retakes possible, b) impose non-trivial time and effort costs to retaking an exam by, for instance, requiring students to attend remedial courses, or c) reduce the pecuniary return to education once a subject has been failed the first time – for instance, employers may treat a student who graduated from high school in five years differently from one who graduated “on-time” for a variety of reasons – it is possible that even short-run heat exposure can have knock-on effects on educational attainment. 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).

So far, studies have found evidence for knock-on effects of temporary score shocks in the context of teacher manipulation (Dee et al, 2016; Diamond and Persson, 2016) and air pollution (Lavy and Ebenstein, 2015), 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 zone⁴³. Using administrative records from Swedish middle schools, Diamond and Persson (2016) also find substantial knock-on effects of upward score manipulations on subsequent performance, graduation likelihood, and even later life income. Lavy and Ebenstein (2015)

⁴²In 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. In practice, this seems unlikely, and I assume that the physiological impact of temporary heat stress does not reduce human capital directly.

⁴³They 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.

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 systematically lower wages later in life.⁴⁴ 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.

6.2.1 Empirical Estimation of Knock-on impacts

How does one estimate the potential causal impact of acute heat stress on longer-run outcomes? In this section, I explore the possibility that short-run heat stress may affect educational attainment of NYC public school students by using information on graduation and drop-out status.

Figure 14 presents a binned scatterplot of 4-year graduation status on average exam-time temperature by student during core subject June Regents exams, controlling for the number of exams taken by student as well as student-level observables (ethnicity, gender, average previous z-scores, federally subsidized school lunch eligibility) and average precipitation at the school level, and motivates the analysis that follows. 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.⁴⁵

This poses additional challenges to causal identification. Computing a measure of average heat exposure across multiple exam sittings by student results in a mechanical correlation between average experienced temperature and the number of takes such that students with a higher number of exam takes are more likely to be assigned average temperature values that are closer to the climatic mean, and students with a lower number of takes are more likely to be assigned tail-end extreme values. That 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 figure 34 in the Appendix).

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 take-count

⁴⁴Lavy and Ebanstein find that a one standard deviation increase in the fraction of “heavily polluted” exam days is associated with a 2.19 and 2.70 percentage point decline in the probability of receiving a Bagrut matriculation certificate for PM2.5 and CO respectively

⁴⁵Graduation 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.

and year-count cells: that is, compare long-run outcomes for students who are observed for the same number of years and take the same number of June Regents exams, controlling for all of the higher-level observable factors as in section 6.

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

$$g_{ijcn} = \alpha_0 + \alpha_1 \overline{T_{ij}} + \alpha_2 X_{ij} + \chi_j + \theta_c + Z_i + takes_n + \epsilon_{ijc} \quad (15)$$

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 . X_{ij} is a vector of weather controls 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 $takes_n$ denotes number of takes 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.

School fixed effects account for potential bias arising from the fact that, if low- (high-) performing schools tend to be located in areas of New York that are more- (less-) likely to experience heat stress, then graduation rates and average exam-time heat exposure may be correlated due to these cross-sectional differences. Time-trends in graduation rates allow for the possibility that heat exposure and graduation rates are correlated due to secular trends in both variables – that is, secular changes in graduation rates may in principle be correlated with shifts in the climate distribution, 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 . Time of day, day of week, and day of month variables account for possible correlation between average exam-time temperature and other unobserved dimensions of student quality that happen to be systematically correlated with exam schedules (for instance, if more difficult subjects taken by higher ability students tend to be scheduled during earlier parts of the testing calendar, or during cooler morning periods).⁴⁶

⁴⁶The 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.

Table 14 presents the results from running variations of equation 15 with and without school and cohort fixed effects. Standard errors are clustered at the school by date and time level, based on the intuition that this is 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 0.13(se=0.04) to 0.17(se=0.04) percentage point decline in the likelihood of graduating on time. This corresponds to a -0.80 to -1.07 percentage point decline in the likelihood of on-time graduation from a one standard deviation in average exam-time temperature (+4.3°F), or 1.17% to 1.57% decline relative to a mean value of 68%.

6.2.2 Shock Persistence, Fade Out, and the Economic Consequences of Teacher Discretion

How persistent are these knock on impacts? And might the persistence of these plausibly random, temperature-driven score shocks shed light on the apparent puzzle of test score fade out, wherein successful educational interventions exhibit positive score effects that “fade out” over time but exhibit persistent long-run impacts on later-life outcomes (Cascio and Staiger, 2012; Chetty et al, 2010)?

Recent studies find evidence for highly persistent impacts from upward grade manipulation by teachers, including on later school performance, graduation status, number of credits completed, and even later-life labor market outcomes (Dee et al, 2016; Diamond and Persson, 2016). Studies that use such “exogenous” score shocks due to grade manipulation by teachers are careful to note the possibility that selective grade manipulation by teachers may not be entirely random; that is, that they may incorporate “hidden” information regarding student quality or underlying human capital that teachers perceive but test scores do not pick up. If, in contrast, score shocks due to heat stress are indeed truly random, it is in theory possible to examine this hypothesis by comparing the persistence and magnitude of longer-run impacts between students who receive higher (lower) scores due to selective grade manipulation by teachers with students who receive higher (lower) scores due to random fluctuations in test day temperature. If the latter effect is smaller or less persistent, this would be consistent with a world in which teacher discretion carries additional information regarding a student’s underlying ability, the amount of learning he/she actually achieved, or other skills that the teacher imparted that are not reflected in the score-based assessment, and that some of the documented long-run impacts of grade-manipulation and/or other educational interventions whose success is evaluated on the bases of standardized exam scores may be a function of this omitted information.⁴⁷

⁴⁷In future work, I hope to examine this hypothesis in greater detail, by exploiting the two different sources of Regents score variation documented in this study and Dee et al (2016), and comparing the respective long-run impacts.

Table 16 presents results from running equation 15 with dummy variables for whether a student graduated in 4, 5, or 6 years, as well as whether she is ever recorded in my data set as having graduated from high school on the left-hand side. Whereas average exam-time temperature during the first four years of high school has a negative impact on 4- and 5-year graduation rates, it is not significantly correlated with 6-year graduation rates, or a dummy for whether or not the student ever graduated. One interpretation of this could be that students who experience random score shocks may, with enough time and enough retakes, be able to receive scores reflective of true underlying human capital. That is, unlike *selective* score shocks caused by teacher discretion, *truly random* score shocks exhibit relatively limited long-run impacts, consistent with a world in which test scores provide noisy signals of human capital, and that teachers are often able to identify student-specific components of human capital that are not accurately reflected in cardinal test score values. These results, though suggestive, do not capture later-life outcomes beyond high school, and should be taken if anything as motivating evidence for further research.

7 Adaptation: (How) Do Agents Respond to Heat Stress?

If heat stress affects an economic agent’s welfare directly, and the effect is at least partially observable by that agent (or her principal, in a principal-agent arrangement), theory would predict the agent (and/or the principal) to adaptively respond. Thus, if heat stress affects student performance in school settings, we would expect students, parents, and teachers to respond to mitigate this impact, presumably along the most cost-effective margins.

However, adaptation to climatic stressors often takes place in a constrained context, for instance due to pre-existing built infrastructure (Hallegatte, 2009), implying almost by default that optimization in response to temperature stress will be second-best at least in the short run.⁴⁸

In the case of heat stress in schools, central air conditioning as part of a well-designed combined HVAC (heating, ventilation, and air conditioning) system would seem to be the first-best option, given ergonomic assessments and existing evidence on the effectiveness of air conditioning at mitigating adverse health impacts from heat stress. In the case of NYC public schools specifically, and possibly other resource-constrained urban schools, there may be physical, financial, or institutional constraints that make the first-best option unfeasible. For instance, the typical NYC public school building was constructed in 1932, meaning that the vast majority of schools were built before centralized air conditioning systems began to be included as the default for new construction. This would make air conditioning additionally costly in the case of retrofits on existing schools, and possibly add new sources of student

⁴⁸Some have suggested similar mechanisms to be responsible for the well-documented “energy efficiency gap” (Jaffe and Stavins, 1994; Allcott and Greenstone, 2012).

distraction (e.g. noise, poor air quality) if air conditioning is added piecemeal, without an integrated revamping of the building’s ventilation system.⁴⁹

In this section, I explore the potential means and realized efficacy of two possible adaptation strategies – air conditioning and “adaptive grading” – assuming that agents are at least partially aware of the adverse impacts of heat stress. Specifically, I use building-level air conditioning (AC) installation data to assess the effectiveness of AC at mitigating the short-run impacts of heat stress on exam performance. I then build on the results of Dee et al (2016) to examine the possibility of such “adaptive grading,” whereby teachers, in the presence of institutional and physical constraints, exercise *selective discretion* in grade manipulation to blunt some of the adverse welfare impacts of heat stress by positively manipulating the grades of students who experience hot exam sessions and score just below the passing cutoff.

7.1 Air Conditioning

7.1.1 Air Conditioning in NYC Public Schools

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

According to the 2012 BCAS, 62% of all NYC public school buildings were reported as having any kind of air conditioning equipment on its premises, including window units, which means that fully 38% of school buildings (comprising over 40% of NYC public schools) did not have any form of air conditioning equipment available. Of the 62% that were reported as having air conditioning, 42% (302 out of 719) were cited as having defective components, according to the third-party auditors conducting the BCAS assessments.

What do we expect the effect of having a typical air conditioning unit to be? 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

⁴⁹Institutional features such as Regents exams with pre-determined and harmonized dates and locations across the entire city and state may pose additional constraints in adapting to heat stress since, in principle, a potential solution to short-run performance impacts of heat stress would be to reschedule exams on cooler days.

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 (Niu, 2004), suggesting that AC may have unintended side effects on cognition and task performance. It is therefore unclear ex ante whether the net effect of air conditioning on student performance in the NYC public school context should be unambiguously positive.

In addition, 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.

I estimate a variant of equation 11 that replaces student fixed effects with school fixed effects, separately for sub-groups of students who took exams in schools with and without non-defective central air conditioning, as well as for sub-groups in schools with and without any air conditioning at all. The results from these regressions are reported in Table 24. Column (1) reproduces the main effect on the entire sample. Columns (2), (3) and (4), (5) present results for sub-groups with and without non-defective central AC, and with and without any AC respectively. The point estimates are smaller for sub-samples with AC units, -0.073 (se=0.041) and -0.068 (se=0.036), relative to sub-samples without AC: -0.091 (se=0.042) and -0.096 (se=0.045). The differences however are not statistically significant. These results suggest that air conditioning may offer a protective effect against the adverse impacts of heat stress in school, but that more careful analysis is needed to ascertain the scope of possible effectiveness, as well as relative costs and benefits inclusive of forward-looking climate change projections.

7.2 Adaptive Grading

Dee et al (2016) document systematic grade manipulation in NYC Regents exams. They suggest the most plausible explanation to be the goodwill of teachers who seek to offset the impact of “a bad test day”, but do not expound upon the factors may give rise to such bad test days. As they put it: *“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).”*

This begs a natural question: 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?

As mentioned previously, 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.⁵⁰

It seems possible, however, that teachers may be able to observe or at least intuit the disruptive impacts of elevated classroom temperatures on test day (recall that Regents exams are taken in home schools). 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”.

To assess the potential presence and magnitude of adaptive grading, I first estimate a version of Dee et al’s bunching estimator by school, subject, month, and year – in effect, 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, 11 subjects overall. Following Dee et al, I fit a polynomial to these fractions by subject, excluding data near the proficiency cutoffs with a set of indicator variables, using the following regression:

$$F_{ks} = \sum_{i=0}^q \psi_{ismyj} (Score)^i + \sum_{i \in -M_{cs}, +M_{cs}} \lambda_{ismyj} \mathbf{1}[Score = i] + \epsilon_{ksmyj} \quad (16)$$

where F_{ks} denotes the fraction of observations with score k for subject s , q is the order of the polynomial, and $-M_{cs}, +M_{cs}$ represent manipulable ranges below and above the passing thresholds. The subscripts s, m, y and j denote subject (e.g. ELA), 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, 2010), 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

⁵⁰Indeed, running equation 11 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 significantly reduced: $\beta = -0.076$ ($se=0.016$) as opposed to -0.141 ($se=0.035$) in the full sample.

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 (see Dee et al, 2016, for details).

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.⁵¹

This generates a set of predicted fractions by score by 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 1998-2011, and using the subject-specific bunching estimator above, I find that 5.8% of all June Regents exams exhibited bunching at or near the passing score cutoffs.

I then calculate observed fractions for each score from 0 to 100 by school, month, year, and subject, and generate a measure of bunching that integrates the differences between observed and predicted fractions: that is, summing the excess mass of test results that are located to the right of the cutoff (above the predicted curve) and the gaps between predicted and observed fractions of test results to the left of the cutoff (below the predicted curve):

$$\zeta_{smj} = \frac{1}{2} \sum_{i \in +M_{ck}} (F_{ks} - \hat{F}_{ksmj}) + \frac{1}{2} |\sum_{i \in -M_{ck}} (F_{ks} - \hat{F}_{ksmj})| \quad (17)$$

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.

Figure 20 presents a binned percentile plot 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 – bunching around the passing cutoffs – and the ambient temperature during the exam being graded. Recall that Regents exams were, up until 2011, graded by teachers in students’ home schools. To the best of my knowledge, they were graded at the home school either in the evening following the exam or on a pre-specified day at the end of each month-specific exam period (for instance, the last Friday of the exam period), which means that it is possible that teachers remember which exams were subject to more heat stress even within a given exam period. Figure 21 presents a similar binned percentile

⁵¹As 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.

plot, but adding school fixed effects to allow for arbitrary differences in the average amount of grade manipulation across schools. The clear positive association between bunching estimate and temperature remains.

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:

$$\zeta_{smj} = \delta_0 + \delta_1 T_{smj} + \delta_2 X_{smj} + \chi_j + \epsilon_{smj} \quad (18)$$

$$\zeta_{smj} = \delta_0 + \delta_1 T_{smj} + \delta_2 X_{smj} + \chi_j + \eta_s + \theta_y + \epsilon_{smj} \quad (19)$$

$$\zeta_{smj} = \delta_0 + \delta_1 T_{smj} + \delta_2 X_{smj} + \chi_j + \eta_s + \delta_3 Year_y + \delta_4 Year_y^2 + \delta_5 Year_y^3 + \epsilon_{smj} \quad (20)$$

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 $Year_y...Year_y^3$ 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.⁵²

According to the estimates presented in Table 22, 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. While these results are highly suggestive of adaptive grading, it is of course not possible to infer teachers' intentions. It could be the case that teachers have an intuitive sense of whether a particular student scored below his or her "true ability", irrespective of whether or not this was due to temperature or other exam-time conditions, and that they respond by manipulating grades in the case of students on the passing margin.

8 Discussion and Conclusion

This paper explores the impact of heat stress on human capital production. 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 student performance on high-stakes exams as well as reducing

⁵²It 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.

the amount of learning achieved over the course of the school year. The research design exploits quasi-random variation in local temperature to identify the causal impact of hot days on short run exam performance and long run educational attainment. The breadth and depth of the data set allows not only for credible causal estimation of the adverse impacts of heat stress, but also an assessment of possible adaptive responses by students and teachers, which may be especially important in thinking about the potential human capital impacts of future climate change.

Taking a high school exit exam on a 90° F day results in a 3% (0.17 standard deviation) reduction in exam performance relative to a more optimal 72°F day controlling for student specific unobservables, which amounts to roughly 1/4 of the Black-White test score gap. Given existing institutional rigidities – namely, cardinal pass/fail thresholds and graduation requirements that depend on exam scores in a formulaic way – these short-run impacts on student performance can have non-trivial knock-on effects on student attainment, leading to a 10.88% reduction in the probability of passing that particular subject exam, and, for the average New York City student, a roughly 2% lower likelihood of that student graduating on time.

I find evidence supportive of the notion that heat stress can disrupt learning and thus reduce the rate of human capital accumulation, which could have important implications for both climate and education policy. Cumulative heat exposure over the course of the preceding school year, measured by the number of days during the school year where temperatures exceed 80°F, is associated with significant reductions in end of year exam performance, controlling for the exam-day effects of heat stress noted above. A year with five additional 80°F+ days is associated with a 2.1% (-0.11 standard deviation) reduction in learning on average, as measured by the same end of year regents exams, 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 the gains from having a one standard deviation higher value-added teacher for that school year, though more careful research is needed to examine whether they result in comparable effects on later-life outcomes.

Using building level air-conditioning installation data, I find that air conditioning may have a mild but incomplete protective effect. 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 (evidence of grade manipulation) is highly correlated with exam time temperature at the school and during the test session in which any particular set of subject exams were administered, suggesting that teachers may have been engaging in a form of “adaptive grading” to offset some of the adverse welfare consequences of momentary heat stress during a high stakes exam, which presumably affects students scores but does not reduce human capital per se.

These results have several implications. First, they imply that heat stress should be in-

cluded among the long list of inputs to the human capital production function that economists have studied to date. This is especially true in light of the fact that climate change will lead to greater additional heat exposure for poorer populations both across and within countries given well documented relationships between income and hot current climate (Acemoglu and Dell, 2010; Park et al, 2015) as well as between income and air conditioning ownership (Gertler et al, 2016).

From the perspective of climate policy, this study lends further support to the notion that direct human impacts of heat stress should be included in climate damage functions and resulting social cost of carbon estimates (Tol, 2009; Burke et al, 2016; Heal and Park, 2016). These results also seem to underscore the importance of taking the interaction between weather shocks and existing institutions into account in assessing the realized welfare impacts of climate change.⁵³

From a methodological perspective, these findings support the use of weather instruments as a means of isolating truly random variation in test scores, which can in principle be used to inform the issue of score fade out (Cascio and Staiger, 2012; Chetty et al, 2010).

In terms of our understanding of the role of climate in economic development, this study lends further evidence that the longstanding correlation between geography and economic growth (Gallup, Sachs, Mellinger, 1999; Acemoglu, Johnson, Robinson, 2000; Rodrik et al, 2004) may have a causal as opposed to merely correlational component.

The findings from this study suggest several fruitful avenues for future research, and raises a number of policy-relevant questions. If heat exposure can affect both short- and long-run student performance, how much of the residual racial (Black-White, Hispanic-White) test-score gap be a function of climatic factors?⁵⁴ If heat stress during high stakes standardized testing affects exam performance differentially across regions, could standardized exams such as the SAT or GRE that are curved relative to a national population exhibit general equilibrium impacts? Finally, do the effects documented here generalize to developing country contexts such as India or Sub-Saharan Africa, where present and future heat burdens are likely to be more severe?

⁵³Such institutional factors may be especially important in thinking about the welfare consequences of heat in the human capital production process if they exacerbate underlying principal-agent problems. For instance, school principals may not fully internalize the welfare costs of elevated school temperatures (lack of proper cooling equipment) if the link between principal behavior and student outcomes is imperfectly observable. The result may be a socially inefficient investment in this particular form of productive capital.

⁵⁴Blacks and Hispanics are more likely to reside in hotter states and to attend school in highly urbanized districts, which may be subject to additional urban heat island effects.

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. Work. Pap* (2015).
4. Alexander, Karl L., Doris R. Entwisle, and Linda S. Olson. "Schools, achievement, and inequality: A seasonal perspective." *Educational Evaluation and Policy Analysis* 23.2 (2001): 171-191.
5. Albouy, David, et al. Climate amenities, climate change, and American quality of life. No. w18925. National Bureau of Economic Research, 2013.
6. Allcott, Hunt, and Michael Greenstone. "Is there an energy efficiency gap?." *The Journal of Economic Perspectives* 26.1 (2012): 3-28.
7. 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.
8. Anderson, Craig A., William E. Deuser, and Kristina M. DeNeve. "Hot temperatures, hostile affect, hostile cognition, and arousal: Tests of a general model of affective aggression." *Personality and Social Psychology Bulletin* 21.5 (1995): 434-448.
9. Angrist, Joshua D., and Victor Lavy. Using Maimonides' rule to estimate the effect of class size on student achievement. No. w5888. National Bureau of Economic Research, 1997.
10. Anthoff, David, Cameron Hepburn, and Richard SJ Tol. "Equity weighting and the marginal damage costs of climate change." *Ecological Economics* 68.3 (2009): 836-849.
11. Anderson, Craig A. "Temperature and aggression: effects on quarterly, yearly, and city rates of violent and nonviolent crime." *Journal of personality and social psychology* 52.6 (1987): 1161.
12. 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.

13. Ardon-Dryer, Karin, Y-W. Huang, and Daniel J. Cziczo. "Laboratory studies of collection efficiency of sub-micrometer aerosol particles by cloud droplets on a single-droplet basis." *Atmospheric Chemistry and Physics* 15.16 (2015): 9159-9171.
14. Becker, Gary S. "Investment in human capital: A theoretical analysis." *The journal of political economy* (1962): 9-49.
15. Billings, Stephen B., David J. Deming, and Jonah Rockoff. "SCHOOL SEGREGATION, EDUCATIONAL ATTAINMENT, AND CRIME: EVIDENCE FROM THE END OF BUSING IN CHARLOTTE-MECKLENBURG." *The Quarterly Journal of Economics* 435 (2014): 476.
16. Bowler, K., and R. Tirri. The temperature characteristics of synaptic membrane ATPases from immature and adult rat brain. *Journal of neurochemistry* 23, no. 3 (1974): 611-613.
17. Burke, Marshall, and Kyle Emerick. "Adaptation to Climate Change: Evidence from US Agriculture." (2015).
18. Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. "Global non-linear effect of temperature on economic production." *Nature* (2015).
19. Burke M, Craxton M, Kolstad CD, Onda C, Allcott H, Baker E, Barrage L, Carson R, Gillingham K, Graff-Zivin J, Greenstone M. Opportunities for advances in climate change economics. *Science*. 2016 Apr 15;352(6283):292-3.
20. Cachon, Gerard, Santiago Gallino, and Marcelo Olivares. "Severe weather and automobile assembly productivity." *Columbia Business School Research Paper* 12/37 (2012).
21. Cascio, Elizabeth U., and Douglas O. Staiger. Knowledge, tests, and fadeout in educational interventions. No. w18038. National Bureau of Economic Research, 2012.
22. Card, David, and Jesse Rothstein. "Racial segregation and the black-white test score gap." *Journal of Public Economics* 91.11 (2007): 2158-2184.
23. Chay, Kenneth Y., and Michael Greenstone. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy* 113.2 (2005).
24. 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.

25. Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *THE AMERICAN ECONOMIC REVIEW* 104.9 (2014): 2593-2632.
26. 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.
27. 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.
28. Cragg, Michael I., and Matthew E. Kahn. "Climate consumption and climate pricing from 1940 to 1990." *Regional Science and Urban Economics* 29.4 (1999): 519-539.
29. 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.
30. Currie, Janet. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* (2009): 87-122.
31. 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.
32. 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.
33. 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.
34. Deryugina, Tatyana, and Solomon M. Hsiang. "Does the environment still matter? Daily temperature and income in the United States." (2014).
35. Deschnes, 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.

36. 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.
37. 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.
38. Fitzpatrick, C., McKinnon, R. D., Blair, C. B., & Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children?. *Learning and Instruction*, 30, 25-31.
39. Fogel, R. W. (2004). Health, Nutrition, and Economic Growth*. *Economic Development and Cultural Change*, 52(3), 643-658.
40. Gaffin, S. R., Rosenzweig, C., Khanbilvardi, R., Parshall, L., Mahani, S., Glickman, H., . . . & Hillel, D. (2008). Variations in New York City's urban heat island strength over time and space. *Theoretical and applied climatology*, 94(1-2), 1-11.
41. Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger. "Geography and economic development." *International regional science review* 22.2 (1999): 179-232.
42. Glewwe, Paul, and Michael Kremer. "Schools, teachers, and education outcomes in developing countries." *Handbook of the Economics of Education* 2 (2006): 945-1017.
43. Goldin, Claudia Dale, and Lawrence F. Katz. *The race between education and technology*. Harvard University Press, 2009.
44. Goodman, Joshua. Flaking out: Student absences and snow days as disruptions of instructional time. No. w20221. National Bureau of Economic Research, 2014.
45. Goodman, Joshua, and Jisung Park. *Hot Takes: Temperature, Student Performance, and the Racial Achievement Gap*. Unpublished Manuscript. 2016
46. 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.
47. Greenstone, Michael, and B. Kelsey Jack. "Envirodevonomics: A research agenda for an emerging field." *Journal of Economic Literature* 53.1 (2015): 5-42.
48. Grether, W. (1973). "Human performance at elevated environmental temperatures." *Aerospace Medicine* 44(7): 747-755.

49. 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.
50. Hallegatte, Stephane. "Strategies to adapt to an uncertain climate change." *Global Environmental Change* 19.2 (2009): 240-247.
51. 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.
52. 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.
53. Hogrefe, C., Biswas, J., Lynn, B., Civerolo, K., Ku, J. Y., Rosenthal, J., & Kinney, P. L. (2004). Simulating regional-scale ozone climatology over the eastern United States: model evaluation results. *Atmospheric Environment*, 38(17), 2627-2638.
54. 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.
55. Hsiang, Solomon M., Marshall Burke, and Edward Miguel. "Quantifying the influence of climate on human conflict." *Science* 341.6151 (2013): 1235367.
56. Isen, A., Rossin-Slater, M., & Walker, R. (2015). Heat and long-run human capital formation. Unpublished Manuscript.
57. Jaffe, Adam B., and Robert N. Stavins. "The energy-efficiency gap What does it mean?." *Energy policy* 22.10 (1994): 804-810.
58. 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.
59. Krueger, Alan B., and Diane M. Whitmore. "The effect of attending a small class in the early grades on collegetest taking and middle school test results: Evidence from Project STAR." *The Economic Journal* 111.468 (2001): 1-28.
60. Lavy, Victor, Avraham Ebenstein, and Sefi Roth. The impact of short term exposure to ambient air pollution on cognitive performance and human capital formation. No. w20648. National Bureau of Economic Research, 2014.

61. Lavy, Victor, Avraham Ebenstein, and Sefi Roth. The Long Run Human Capital and Economic Consequences of High-Stakes Examinations. No. w20647. National Bureau of Economic Research, 2014.
62. Lavy, Victor. "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics." *The American economic review* 99.5 (2009): 1979-2011.
63. 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.
64. 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.
65. Mincer, Jacob. "Investment in human capital and personal income distribution." *The journal of political economy* (1958): 281-302.
66. J. Mincer. (1974) *Schooling, Experience, and Earnings*. New York: Columbia University Press for the National Bureau of Economic Research, pp. 7-22.
67. Niu, Jianlei. "Some significant environmental issues in high-rise residential building design in urban areas." *Energy and buildings* 36.12 (2004): 1259-1263.
68. Matthew E. Kahn *The Climate Change Adaptation Literature*. *Rev Environ Econ Policy* 2016.
69. Mackworth, Norman H. "Effects of heat on wireless operators." *British journal of industrial medicine* 3.3 (1946): 143.
70. Maddala, Gangadharrao Soundaryarao, and Kajal Lahiri. *Introduction to econometrics*. Vol. 2. New York: Macmillan, 1992.
71. Melillo, Jerry M., Terese TC Richmond, and G. W. Yohe. "Climate change impacts in the United States." *Third National Climate Assessment* (2014).
72. 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.
73. Park, Jisung, Stephane Hallegatte, Mook Bangalore, and Evan Sandhoefner. "Households and heat stress: estimating the distributional consequences of climate change." *World Bank Policy Research Working Paper* 7479 (2015).

74. Raichle, Marcus E., and Mark A. Mintun. Brain work and brain imaging. *Annual Review of Neuroscience* 29 (2006): 449-476.
75. Roback, J (1982) Wages, Rents, and the Quality of Life *Journal of Political Economy* Vol. 90 No. 6 pp1257-1278.
76. 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.
77. Rosen, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition *Journal of Political Economy* 82: 34-55.
78. 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).
79. Sandström, F. Mikael, and Fredrik Bergström. "School vouchers in practice: Competition will not hurt you." *Journal of Public Economics* 89.2 (2005): 351-380.
80. 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.
81. Schultz, Theodore W. "Investment in human capital." *The American economic review* (1961): 1-17.
82. Schiff, Steven J., and George G. Somjen. The effects of temperature on synaptic transmission in hippocampal tissue slices. *Brain Research* 345, no. 2 (1985): 279-284.
83. Seppanen, O., et al. (2006). "Effect of temperature on task performance in office environment." Lawrence Berkeley National Laboratory.
84. 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.
85. Sinha, Paramita, and Maureen L. Cropper. Household Location Decisions and the Value of Climate Amenities. No. w21826. National Bureau of Economic Research, 2015.
86. Stock, James H., and Mark W. Watson. "Introduction to econometrics." (2007).

87. Sudarshan, Anant, and Meenu Tewari. The economic impacts of temperature on industrial productivity: Evidence from indian manufacturing. Working Paper, 2013.
88. Vasmatazidis, Ioannis, Robert E. Schlegel, and Peter A. Hancock. An investigation of heat stress effects on time-sharing performance. *Ergonomics* 45, no. 3 (2002): 218-239.
89. Weil, David N. "Accounting for the Effect of Health on Economic Growth." *The Quarterly Journal of Economics* (2007): 1265-1306.
90. Zafar, Basit. 'How Do College Students Form Expectations?'. *Journal of Labor Economics* 29.2 (2011): 301-348
91. Zivin, Joshua Graff, and Matthew Neidell. "The Impact of Pollution on Worker Productivity." *The American Economic Review* 102.7 (2012): 3652-3673.
92. 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.
93. 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 4-year High School Graduation Rates by Borough of New York City

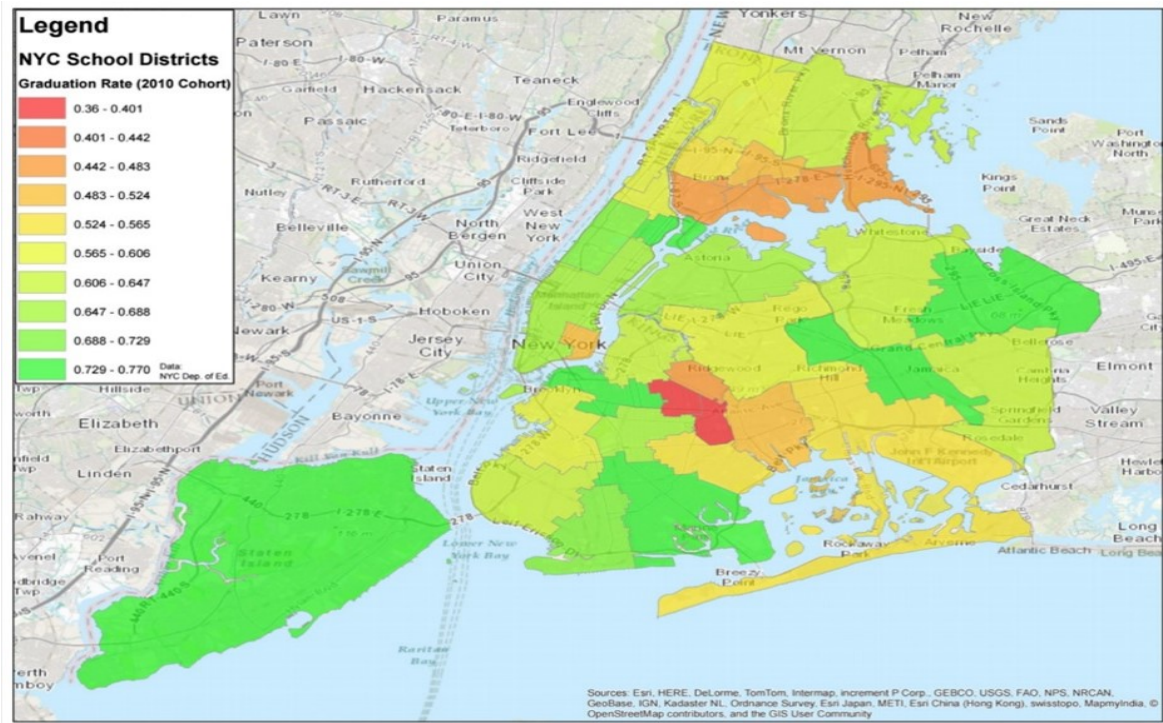


Figure 1: Average 4-year graduation rates by district

Notes: This figure presents average 4 year graduation rates for the 2010 entering cohort by sub-district within the New York City Public Schools system.

Figure 2 Short-Run Identifying Variation in Temperature

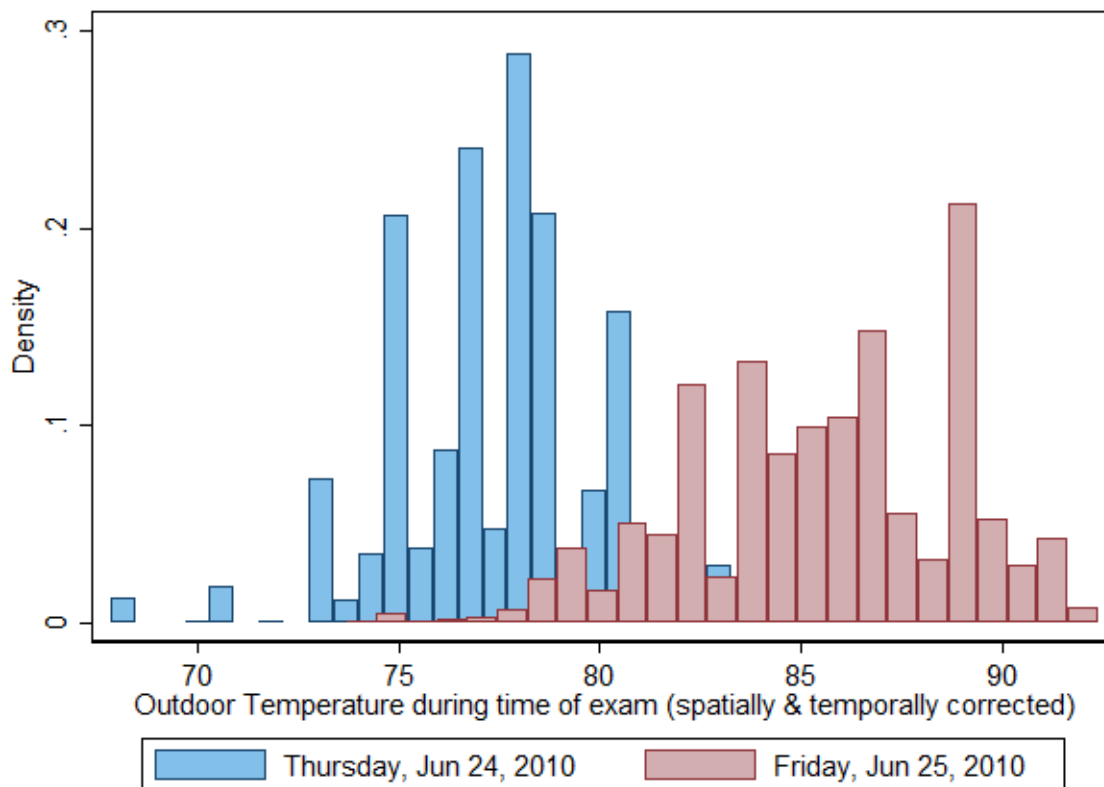


Figure 2: Temperatures measured at the school level, weighted by student population.

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 described in section 4, plotting observations at the student-exam level.

Figure 3 Variation in Daily Temperatures for a Typical June in New York City

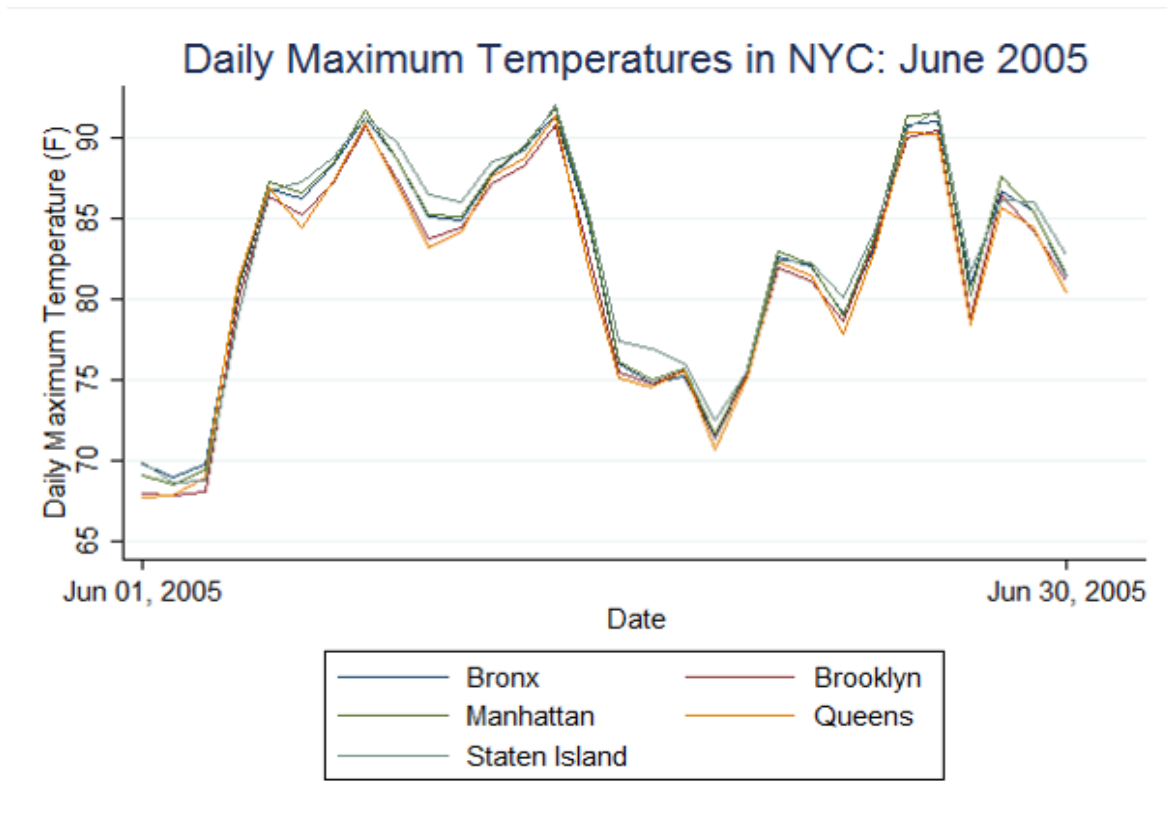


Figure 3: Temperatures measured at the nearest weather station (borough) level.

Notes: This figure illustrates day to day variation in temperature during a typical June, which is the month during which Regents exams are taken by most NYC public school students, across the five boroughs of New York City. Temperature readings are taken from USGS weather stations, one from each of the five boroughs/counties.

Figure 4 Variation in Cumulative Heat Exposure by Borough and Year in New York City

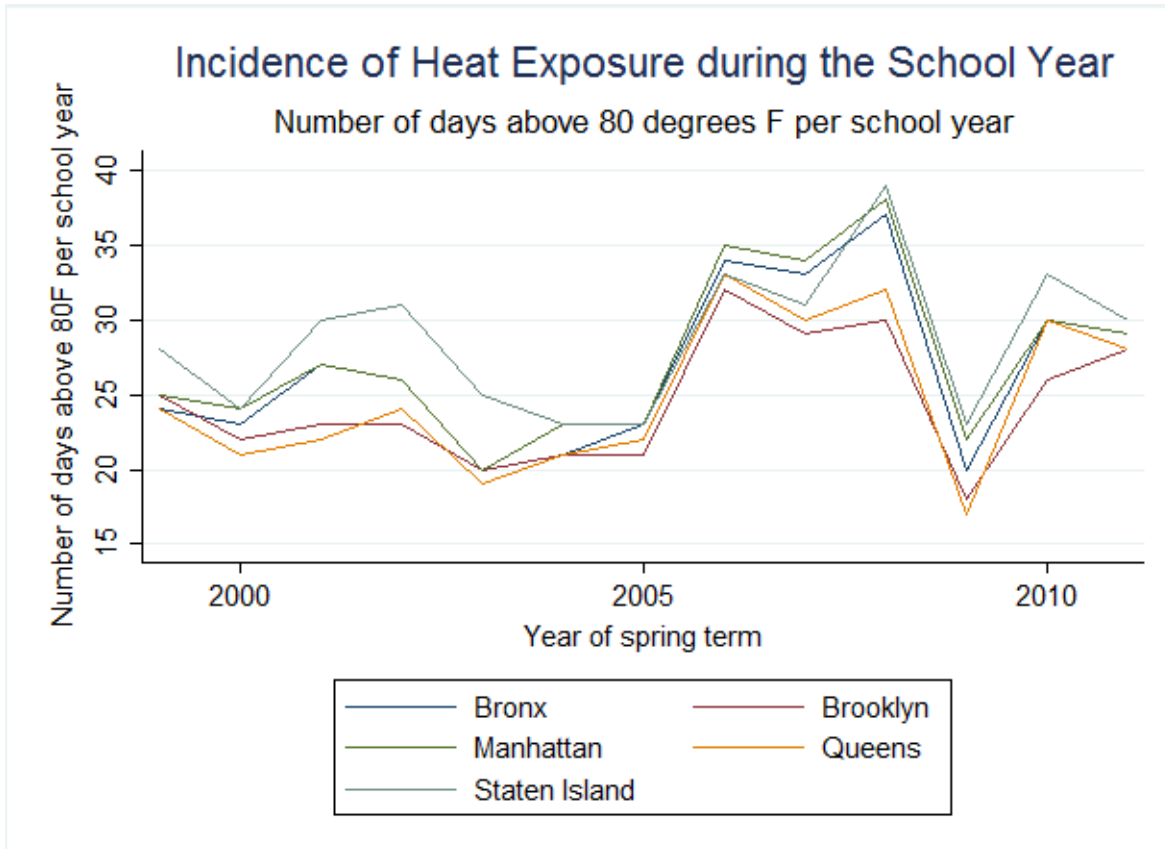


Figure 4: Temperatures measured at the nearest weather station (borough) level.

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.

Figure 5 Documenting Grade Manipulation: Bunching at Passing and Proficiency Thresholds (NY State Regents Exams taken in NYC public schools 1999-2011)

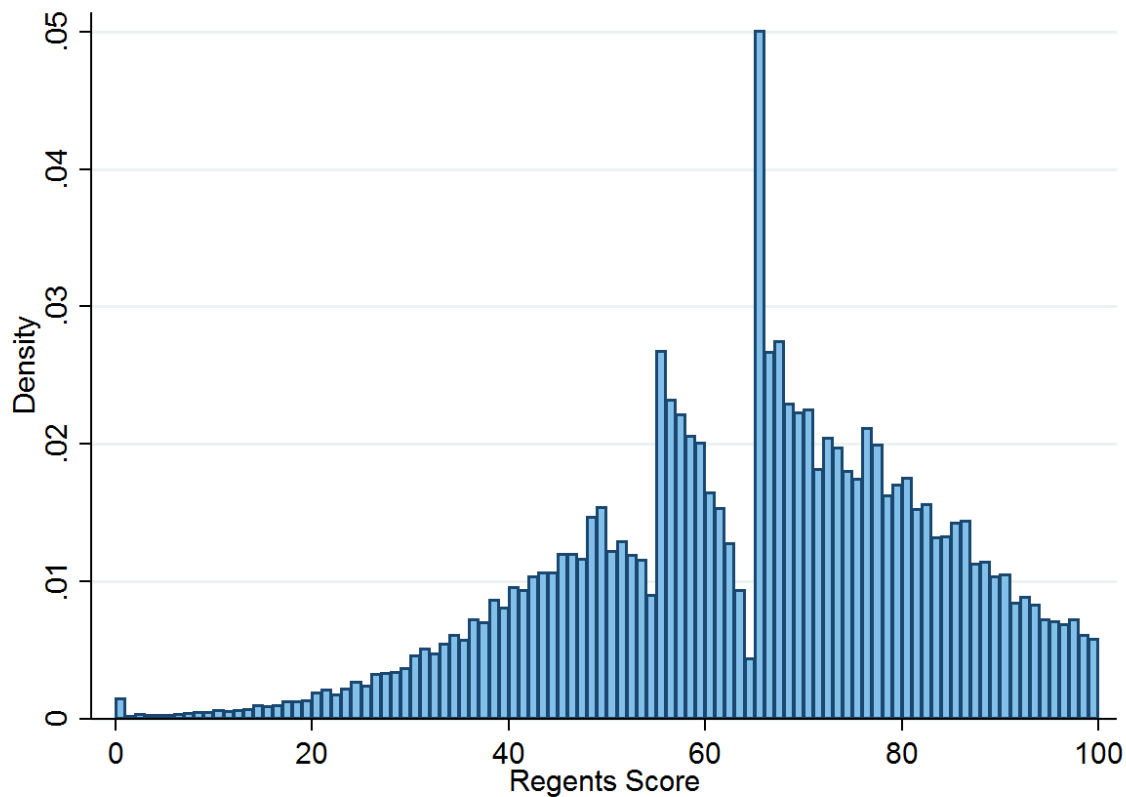


Figure 5: All Regents exams in core subjects prior to NYC grading reforms in 2011-2012.

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.

Table 6 Summary Statistics (A)

	Regents Score	Pass Rate	Proficiency Rate	Previous Ability
Borough				
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)

Figure 6: Summary statistics for key outcome variables

Notes: This figure presents summary statistics for key outcome variables by borough of NYC, with standard deviations in parentheses. 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.

Table 7 Summary Statistics (B)

	Regents Score	Pass Rate	Proficiency Rate	Previous Ability
Ethnicity				
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)
Missing	63.06 (17.28)	0.52 (0.50)	0.27 (0.44)	0.00 (0.06)
Total	64.86 (17.92)	0.57 (0.49)	0.32 (0.47)	0.16 (1.42)

Figure 7: Summary statistics for key outcome variables

Notes: This figure presents summary statistics for key outcome variables by demographic sub-group. Missing observations correspond to the 14% of students for whom Regents scores and scrambled student id's are available but demographic information is not.

Table 8 Average June Regents Scores by Subject

Subject	Mean	Standard Deviation
Algebra 2/Trigonometry	62.78	21.20
Chemistry	64.91	15.14
ELA	64.24	19.17
Earth Science	62.68	17.56
Geometry (Math 2)	65.26	16.52
Global History & Government	63.06	19.89
Integrated Algebra	64.03	15.05
Living Environment	66.20	15.40
Math A	63.38	17.24
Physics	71.68	16.82
US History & Government	67.58	19.87
Total	64.86	17.92

Figure 8: Average June Regents Scores by subject.

Notes: This figure presents Average June Regents Scores by core Regents subject, for the subset of 4.5 million exams (990 thousand students) for whom reliable exam-date matching is possible. For details regarding the matching process, see the Appendix.

Table 9 Average June Regents Scores by Year

Year	Mean	Standard Deviation
1999	62.34	16.30
2000	64.13	16.24
2001	65.71	16.89
2002	62.96	17.74
2003	64.21	18.25
2004	66.63	17.59
2005	65.18	18.33
2006	65.14	18.91
2007	64.05	18.57
2008	65.52	18.23
2009	63.88	18.01
2010	64.44	18.03
2011	66.43	17.07
Total	64.86	17.92

Figure 9: Average June Regents Scores by year.

Notes: This figure presents Average June Regents Scores by year for the subset of 4.5 million exams (990 thousand students) for whom reliable exam-date matching is possible. For details regarding the matching process, see the Appendix.

Figure 10 Short-Run Performance Impacts of Heat Stress on Exam Performance

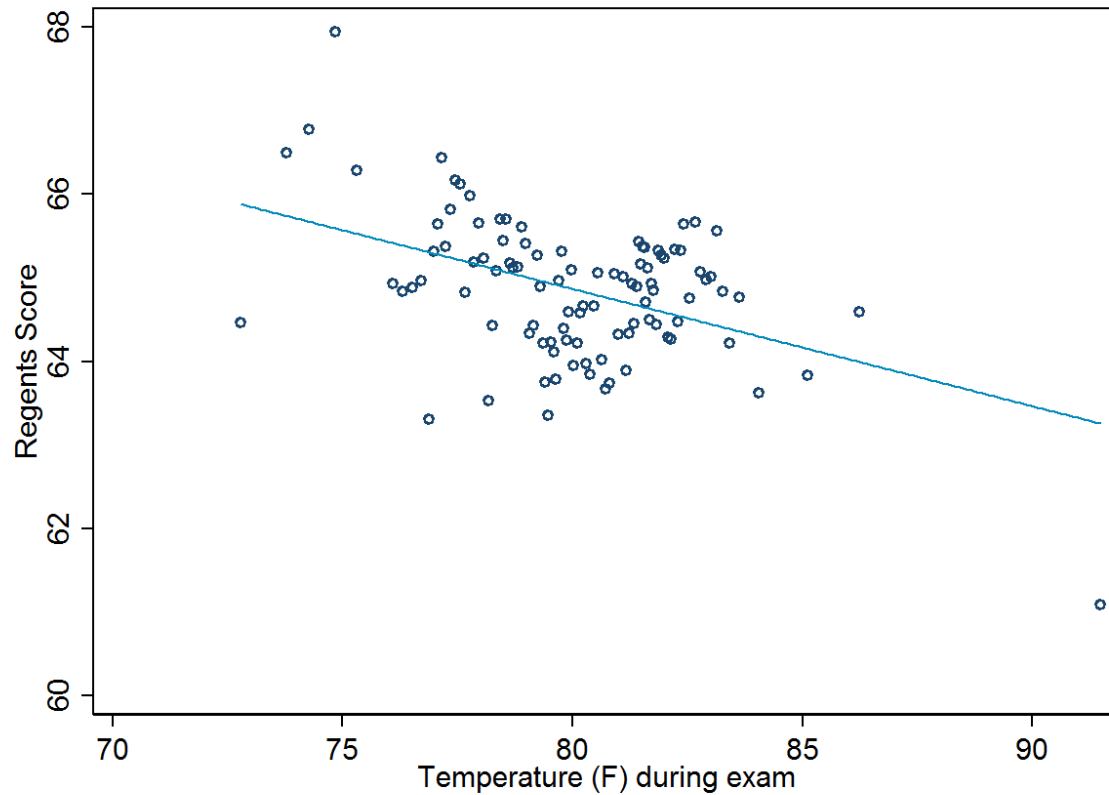


Figure 10: Temperatures measured at the school level, weighted by student population.

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, average differences across years, and exam-day precipitation and humidity (N=4,509,095). Each dot represents approximately 45,000 exam observations.

Table 11 Short-Run Performance Impacts of Heat Stress on Exam Performance:
Regents Scores for NYC Public High School Students

Dep. Var	Score (1)	Score (2)	Score (3)
Temperature (F)	-0.138***	-0.130***	-0.109***
	0.035	0.034	0.037
Precipitation (10 mm)	0.041**	0.039**	0.045**
	0.017	0.018	0.019
Time (Afternoon)		-0.526**	-0.728***
		0.223	0.254
Day of Week			0.434***
			0.086
Day of Month			-0.116*
			0.068
Fixed Effects			
Student FE	X	X	X
Subject FE	X	X	X
Year FE	X	X	X
Time of Day FE		X	X
DOW, DOM FE			X
Observations	4,347,719	4,347,719	4,347,719
R-squared	0.709	0.709	0.709
*** p<0.01, ** p<0.05, * p<0.1			

Figure 11: Short run performance impacts of heat stress during Regents exams

Notes: The dependent variable is Regents score (0-100), with observations at student-exam-date-and-time-level. All regressions include controls for daily precipitation and dewpoint. Student, school, day of week, day of month fixed effects suppressed in output. Demographic controls include ethnicity, gender, and subsidized school lunch status. For student FE regressions, 161,383 singleton observations are dropped. June Regents exams in 11 core subjects only, for years 1998-2011. Temperature is temporally corrected to account for diurnal fluctuations (AM vs PM sessions, see Appendix), and spatially corrected to account for urban heat island effects (see Appendix). Robust standard errors are clustered at school by date and time level.

Table 12 Short-Run Performance Impacts of Heat Stress on Exam Performance:
Regents Exam Pass Rates for NYC Public High School Students

Dep. Var	Pass (1)	Pass (2)	Pass (3)
Temperature (F)	-0.0031*** 0.001	-0.0029*** 0.001	-0.0022*** 0.001
Precipitation (10 mm)	0.0008** 0.000	0.0008** 0.000	0.0010** 0.000
Time (Afternoon)		-0.012** 0.005	-0.0188*** 0.006
Day of Week			0.0083*** 0.002
Day of Month			-0.0033** 0.002
Fixed Effects			
Student FE	X	X	X
Subject FE	X	X	X
Year FE	X	X	X
Time of Day FE		X	X
DOW, DOM FE			X
Observations	4,347,719	4,347,719	4,347,719
R-squared	0.55	0.55	0.551
*** p<0.01, ** p<0.05, * p<0.1			

Figure 12: Short run pass rates

Notes: The dependent variable is a dummy variable denoting whether an exam score was at or above the relevant pass threshold (0=fail, 1=pass), with observations at student-exam-date-and-time-level. All regressions include controls for daily precipitation and dewpoint. Student, school, day of week, day of month fixed effects suppressed in output. Demographic controls include ethnicity, gender, and subsidized school lunch status. For student FE regressions, 161,383 singleton observations are dropped. June Regents exams in 11 core subjects only, for years 1998-2011. Temperature is temporally corrected to account for diurnal fluctuations (AM vs PM sessions, see Appendix), and spatially corrected to account for urban heat island effects (see Appendix). Robust standard errors are clustered at school by date and time level.

Table 13 Short-Run Performance Impacts of Heat Stress on Exam Performance:
College Proficiency Status on Regents Exams for NYC Public High School Students

Dep. Var	Proficient (1)	Proficient (2)	Proficient (3)
Temperature (F)	-0.0031*** 0.001	-0.0029*** 0.001	-0.0024*** 0.001
Precipitation (10 mm)	0.0013*** 0.000	0.0012*** 0.000	0.0014*** 0.000
Time (Afternoon)		-0.0096** 0.004	-0.0144*** 0.005
Day of Week			0.0084*** 0.002
Day of Month			-0.0027** 0.001
Fixed Effects			
Student FE	X	X	X
Subject FE	X	X	X
Year FE	X	X	X
Time of Day FE		X	X
DOW, DOM FE			X
Observations	4,347,719	4,347,719	4,347,719
R-squared	0.611	0.612	0.612
*** p<0.01, ** p<0.05, * p<0.1			

Figure 13: Short run proficiency status

Notes: The dependent variable is a dummy variable denoting whether an exam score was at or above the relevant college proficiency threshold (0=fail, 1=pass), with observations at student-exam-date-and-time-level. All regressions include controls for daily precipitation and dewpoint. Student, school, day of week, day of month fixed effects suppressed in output. Demographic controls include ethnicity, gender, and subsidized school lunch status. For student FE regressions, 161,383 singleton observations are dropped. June Regents exams in 11 core subjects only, for years 1998-2011. Temperature is temporally corrected to account for diurnal fluctuations (AM vs PM sessions, see Appendix), and spatially corrected to account for urban heat island effects (see Appendix). Robust standard errors are clustered at school by date and time level.

Figure 14 4-year Graduation Status and Average Exam-Time Outdoor Temperature by student for NYC Public High School Students

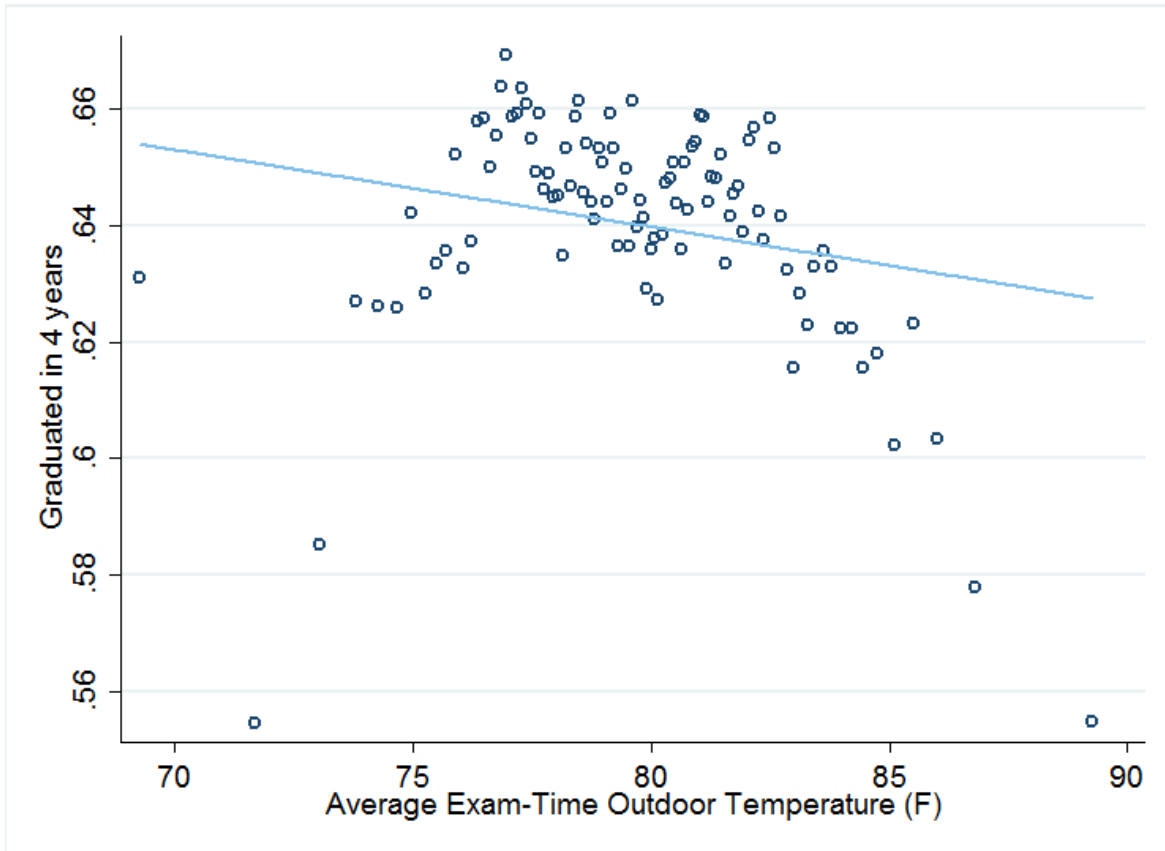


Figure 14: 4-Year Graduation likelihood and average exam-time temperature during first 4 years of high school

Notes: This figure presents a binned scatterplot of 4-year graduation status by percentile of the observed average exam-time temperature distribution, where exam-time temperature for all June Regents core subject exam sessions during the first four years of high school are averaged by student, plotting residual variation after controls for the total number of exams taken by student as well as student-level observables (ethnicity, gender, average previous z-scores, federally subsidized school lunch eligibility) and average precipitation at the school level are included.

Table 15 Knockon Impacts - 4 Year Graduation Rates of NYC Public High School Students

Dep. Var	Graduated (4-year) (1)	Graduated (4-year) (2)	Graduated (4-year) (3)
Average Exam-Time Temperature (June exams)	-0.0013*** 0.000	-0.0014*** 0.000	-0.0017*** 0.000
Average Exam-Day Precipitation (June exams)	0.0010** 0.001	0.0059*** 0.001	0.0036*** 0.000
Previous Z-scores (ELA+Math)	0.0945*** 0.001	0.0903*** 0.001	0.0833*** 0.001
Asian (omitted)	(omitted)	(omitted)	(omitted)
Black	-0.0776*** 0.003	-0.0708*** 0.002	-0.0613*** 0.002
Hispanic	-0.0898*** 0.003	-0.0841*** 0.003	-0.0835*** 0.002
Native American	-0.0176*** 0.003	-0.0180*** 0.003	-0.0307*** 0.002
White	-0.1497*** 0.010	-0.0152* 0.009	0.003 0.009
Female	0.0793*** 0.001	0.0758*** 0.001	0.0689*** 0.001
Subsidized School Lunch Eligible	-0.0444*** 0.003	-0.0683*** 0.003	-0.0616*** 0.002
Fixed Effects			
Number of Takes FE	X	X	X
Cohort FE		X	X
School FE			X
Observations	398,736	398,736	398,692
R-squared	0.238	0.262	0.319
*** p<0.01, ** p<0.05, * p<0.1			

Figure 15: Knockon Impacts of Short-Run Heat Stress during High Stakes Exams - 4 Year Graduation Rates

Notes: The dependent variable is a dummy for whether student graduated on time (4 years), with observations measured at the student level. Robust standard errors are clustered at school by date and time level. All regressions include controls for daily precipitation and dewpoint. Number of takes, 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 during the students' first four years of high school are computed by student.

Table 16 Knockon Impacts of Short-Run Heat Stress during High Stakes Exams - 4, 5, 6 Year Graduation Rates

Dep. Var	Graduated (4-year)	Graduated (5-year)	Graduated (6-year)	Ever Graduated
	(1)	(2)	(3)	(4)
Average Test-Time Temperature (June exams)	-0.0017*** 0.0003	-0.0010** 0.0003	0.0001 0.0001	0.0001 0.0001
Average Exam-Day Precipitation (June exams)	0.0036*** 0.000	0.00285*** 0.000	0.00271*** 0.000	0.0026*** 0.000
Previous Z-scores (ELA+Math)	0.0833*** 0.001	0.07672*** 0.001	0.07055*** 0.001	0.0713*** 0.001
Asian (omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Black	-0.0613*** 0.002	-0.0352*** 0.002	-0.0233*** 0.002	-0.0245*** 0.002
Hispanic	-0.0835*** 0.002	-0.0519*** 0.002	-0.0368*** 0.002	-0.0375*** 0.002
Native American	-0.0887*** 0.010	-0.0700*** 0.009	-0.0536*** 0.009	-0.0535*** 0.009
White	0.003 0.009	0.0812*** 0.006	0.0533*** 0.006	0.0525*** 0.006
Female	0.0689*** 0.001	0.0431*** 0.001	0.0309*** 0.001	0.0316*** 0.001
Subsidized School Lunch Eligible	-0.0616*** 0.002	-0.0418*** 0.002	-0.0317*** 0.002	-0.0324*** 0.002
Fixed Effects				
Number of Takes FE	X	X	X	X
Cohort FE	X	X	X	X
School FE	X	X	X	X
Observations	398,692	435,547	444,690	448,639
R-squared	0.319	0.333	0.339	0.338
*** p<0.01, ** p<0.05, * p<0.1				

Figure 16: Knockon Impacts - 4, 5, 6 Year Graduation Rates

Notes: Dependent variables comprise dummies for whether student graduated in 4, 5, or 6 years, as well as for whether the student ever graduated according to the data available, with observations measured at the student level. Robust standard errors are clustered at school by date and time level. All regressions include controls for daily precipitation and dewpoint. Number of takes, cohort, and school fixed effects are suppressed in output. Temperature is temporally corrected to account for diurnal fluctuations (AM vs PM sessions, see Appendix), and spatially corrected to account for urban heat island effects (see

Appendix). Average temperatures from all June Regents exam takes during the students' first four years of high school are computed by student.

Figure 17: Cumulative Learning Impacts of Heat Exposure during the Preceding School Year (Regents Exam Scores for NYC Public High School Students)

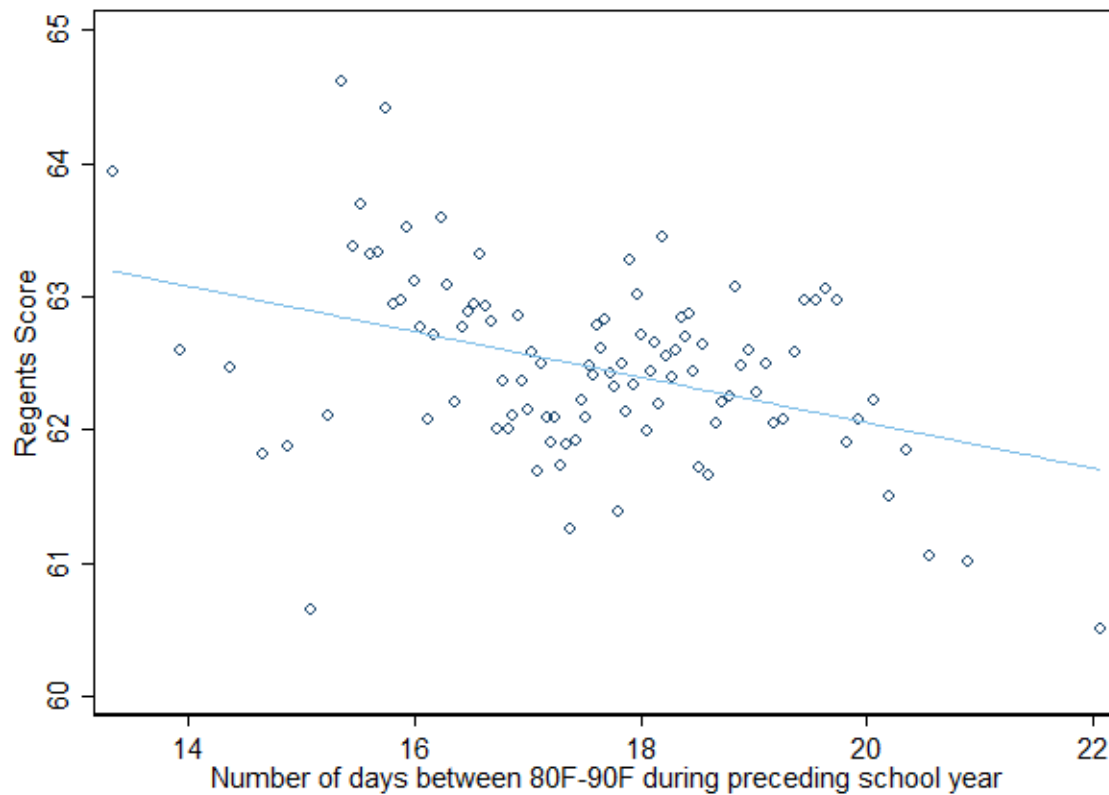


Figure 17: Temperatures measured at the school level, weighted by student population.

Notes: This figure presents a binned scatterplot of end-of-year Regents performance at the school-subject-date level, by percentile of the distribution of cumulative heat stress, measured by the number of days per degree bin (showing days with maximum temperatures between 80°F and 90°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 18 Cumulative Learning Impacts of Heat Exposure during School Year

Dep. Var	Score	Score	Score	Score
	1	2	3	4
Days during school-year with max temp 70F-80F	-0.165***	-0.059	-0.101*	-0.129*
	0.042	0.062	0.058	0.073
Days during school-year with max temp 80F-90F	-0.116*	-0.152*	-0.229**	-0.106
	0.062	0.093	0.093	0.119
Days during school-year with max temp above 90F	0.214	-0.132	0.151	-0.018
	0.172	0.220	0.180	0.319
Days during school-year with max temp below 20F		0.117		0.899
		0.454		0.635
Days during school-year with max temp 20F-30F		0.005		-0.055
		0.095		0.105
Days during school-year with max temp 30F-40F		-0.101***		-0.092***
		0.028		0.026
Days during school-year with max temp 40F-50F				0.002
				0.076
Days during school-year with max temp 50F-60F				0.082
				0.050
Exam-Time Temperature (F)	-0.099***	-0.126***	-0.135***	-0.125***
	0.030	0.037	0.037	0.039
Exam-Day Precipitation (10mm)	0.055***	0.043**	0.049**	0.046**
	0.021	0.022	0.021	0.022
Precipitation during school year (10mm)	-0.193	0.311	-0.177	0.541
	0.575	0.665	0.564	1.178
Average Temperature during school year (F)			0.69371*	
			0.378	
Fixed Effects				
School FE	X	X	X	X
Subject FE	X	X	X	X
Time of Day FE	X	X	X	X
DOW DOM FE	X	X	X	X
Demographic Controls	X	X	X	X
Observations	20,696	20,696	20,696	20,696
R-squared	0.585	0.586	0.585	0.586

*** n<0.01 ** n<0.05 * n<0.1

Figure 18: Cumulative Learning Impacts of Heat Exposure during School Year

Notes: The dependent variable is average Regents score (0-100), collapsed at the school by date-and-time by year level. School, subject, time of day, day of week, day of month fixed effects are suppressed in the output. Robust standard errors are clustered at 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 19 Grade Manipulation Pre- and Post-NYC School Reforms

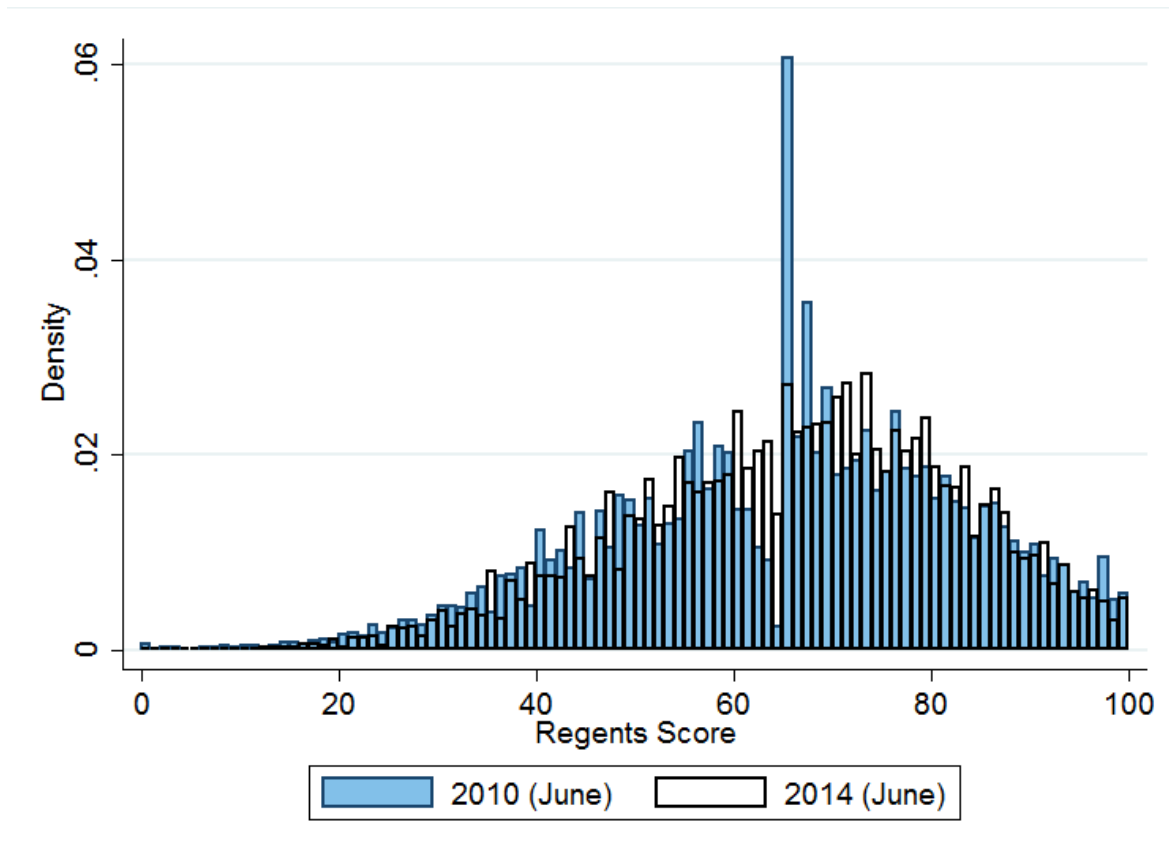


Figure 19: Grade manipulation in NYC Regents exams in June 2010 (pre-reforms) and June 2014 (post-reforms)

Figure 20 Adaptive Grading: Variation in Grade Manipulation by Exam-Time Temperature (Regents Exams 1998-2010)

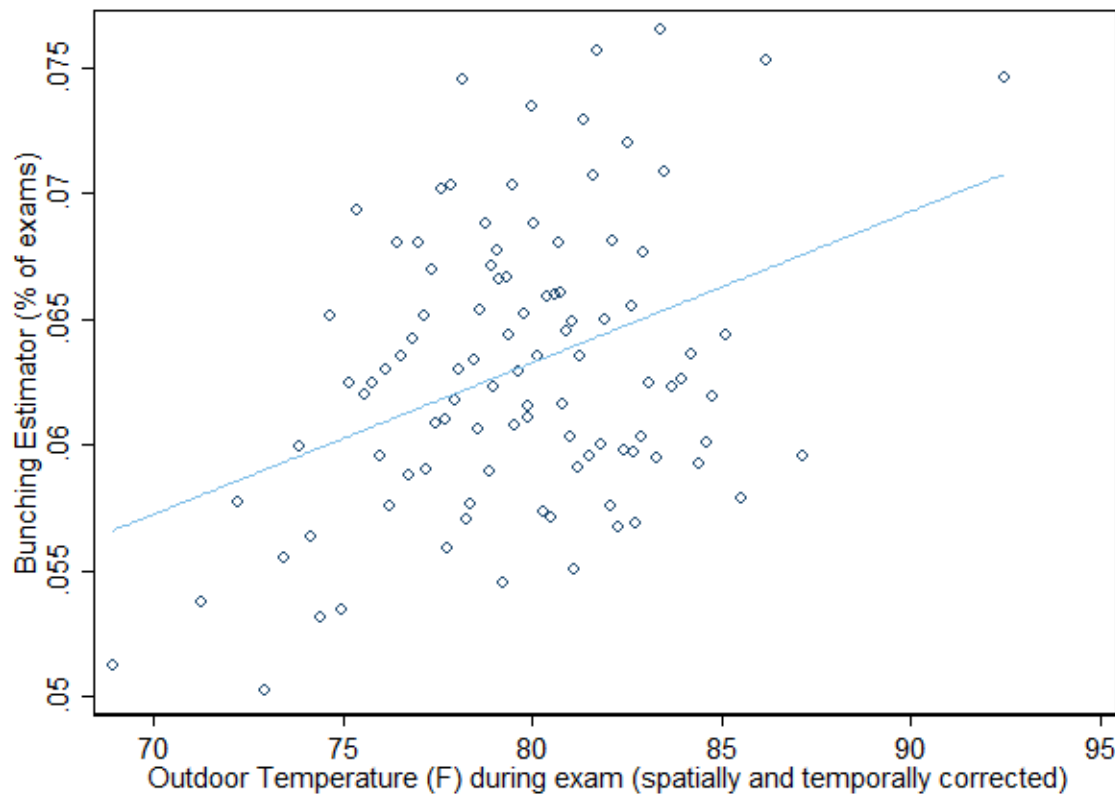


Figure 20: Grade Manipulation varies with exam-time temperature by subject, school, and take.

Notes: This figure presents a binned scatterplot of degree of bunching at the school-subject-date level by percentile of the exam-time temperature experienced by students taking that subject in that school that year (date). The bunching estimator is calculated by integrating the distance between predicted and observed score fractions of scores within the manipulable zone, and is described in greater detail in section 7.

Figure 21 Adaptive Grading: Robustness Checks

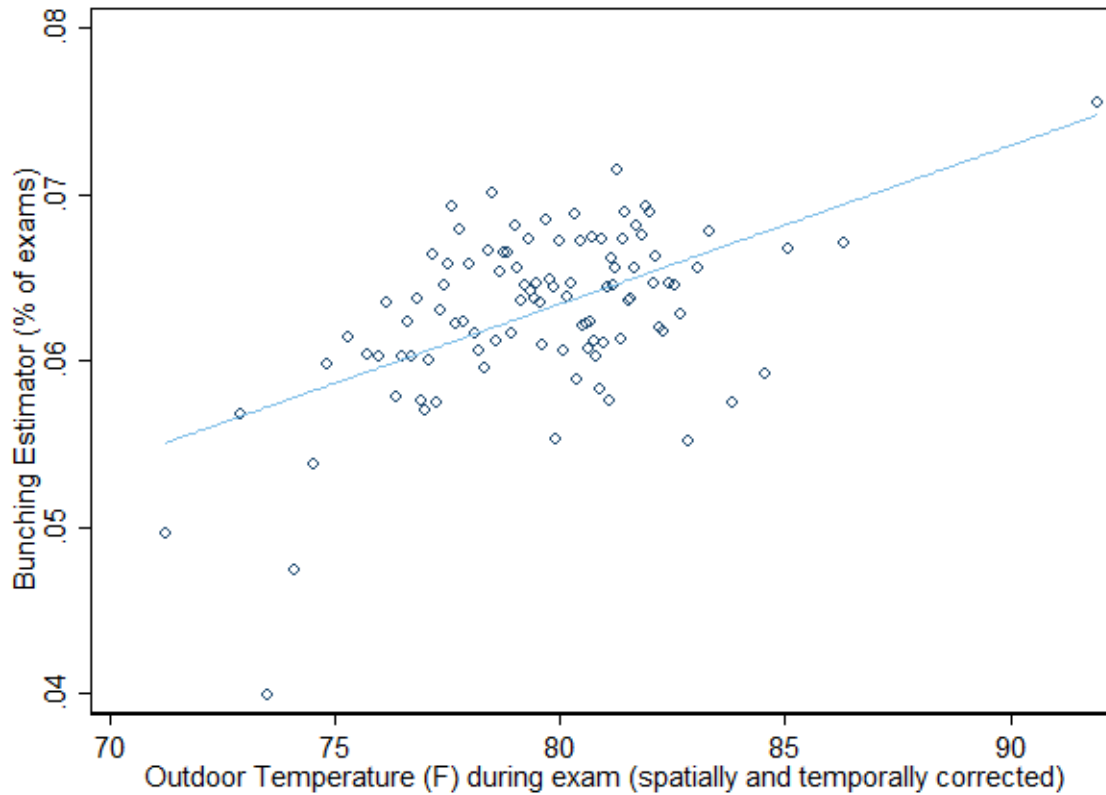


Figure 21: Grade manipulation and exam-time temperature, adding school fixed effects

Notes: Notes: This figure presents a binned scatterplot of degree of bunching at the school-subject-date level by percentile of the exam-time temperature experienced by students taking that subject in that school that year (date), plotting residual variation after controls for subject, year, and school fixed effects, and precipitation and humidity during the exam are included. The bunching estimator is calculated by integrating the distance between predicted and observed score fractions of scores within the manipulable zone, and is described in greater detail in section 7.

Table 22 Adaptive Grading: Grade Manipulation by Exam-Time Temperature

Dep. Var	Bunching Estimator (1)	Bunching Estimator (2)	Bunching Estimator (3)
Temperature (F)	0.0013***	0.0016***	0.0010***
	0.0004	0.0005	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
Observations	30,731	30,731	30,731
R-squared	0.018	0.082	0.081
*** p<0.01, ** p<0.05, * p<0.1			

Figure 22: Bunching Estimates by Exam-Time Temperature

Notes: The dependent variable is the extent of bunching by school-subject-month-year, 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 (see Appendix). All regressions include controls for exam-day precipitation and dewpoint. School and year fixed effects suppressed in output. The bunching estimator is calculated by integrating the distance between predicted and observed score fractions of scores within the manipulable zone, and is described in greater detail in section 7.

Table 23 Heterogeneity in Short-Run Impacts by Demographic Sub-Group

	All	Asians	Black	Hispanic	White	Male	Female	FSS Lunch
Dep. Var	Score	Score	Score	Score	Score	Score	Score	Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temperature (F)	-0.109***	-0.054**	-0.070*	-0.067**	-0.066**	-0.062**	-0.076**	-0.067**
	0.037	0.025	0.039	0.034	0.032	0.029	0.031	0.030
Precipitation (10 mm)	0.045**	0.034**	0.053**	0.061***	0.034*	0.043**	0.055***	0.048***
	0.019	0.016	0.022	0.021	0.019	0.018	0.018	0.018
Day of Week	0.434***	0.040	0.427***	0.330***	0.157*	0.309***	0.262***	0.309***
	0.086	0.083	0.094	0.089	0.095	0.084	0.081	0.081
Day of Month	-0.116*	-0.003	-0.088	-0.126*	-0.141**	-0.063	-0.119**	-0.090
	0.068	0.061	0.073	0.071	0.067	0.063	0.061	0.061
Time (Afternoon)	-0.728***	-0.209	-0.928***	-0.907***	-0.599**	-0.732***	-0.796***	-0.813***
	0.254	0.244	0.274	0.264	0.270	0.235	0.232	0.228
Fixed Effects								
School FE	X	X	X	X	X	X	X	X
Subject FE	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
Time of Day FE	X	X	X	X	X	X	X	X
DOW, DOM FE	X	X	X	X	X	X	X	X
Observations	4,347,719	621,460	1,332,245	1,431,342	546,287	2,075,490	1,890,026	3,633,600
R-squared	0.709	0.734	0.661	0.655	0.752	0.712	0.709	0.698
*** p<0.01, ** p<0.05, * p<0.1								

Figure 23: Heterogeneity in Short-Run Impacts by Demographic Sub-Group

Notes: The dependent variable is Regent exam score (0-100), with observations at student-subject-date-and-time-level. All regressions include controls for daily precipitation and dewpoint. Student, school, day of week, day of month fixed effects suppressed in output. Demographic controls include ethnicity, gender, subsidized school lunch status, and previous ability variable (combined math and ela z-score from 3rd-8th grade). For student FE and school FE regressions, 161383 and 7 singleton observations are dropped respectively. June Regents exams in 11 core subjects only, for years 1998-2011. Robust standard errors are clustered at school by date and time level.

Table 24 Heterogeneity in Short-Run Impacts by School-Level Air Conditioning Installation Status

	All	Non-Defective AC = Yes	Non-Defective AC = No	Any AC = Yes	Any AC = No
Dep. Var	Score	Score	Score	Score	Score
	(1)	(2)	(3)	(4)	(5)
Temperature (F)	-0.109*** 0.037	-0.073* 0.041	-0.091** 0.042	-0.068* 0.036	-0.096** 0.045
Precipitation (10 mm)	0.045** 0.019	0.043** 0.021	0.048** 0.023	0.043** 0.022	0.050** 0.023
Day of Week	0.434*** 0.086	0.429*** 0.103	0.415*** 0.100	0.421*** 0.098	0.427*** 0.103
Day of Month	-0.116* 0.068	-0.148* 0.078	-0.133* 0.078	-0.136* 0.076	-0.146* 0.080
Time (Afternoon)	-0.728*** 0.254	-0.909*** 0.299	-0.705** 0.293	-0.726** 0.289	-0.861*** 0.302
Fixed Effects					
School FE	X	X	X	X	X
Subject FE	X	X	X	X	X
Year FE	X	X	X	X	X
Time of Day FE	X	X	X	X	X
DOW, DOM FE	X	X	X	X	X
Observations	4,347,719	1,019,280	1,611,336	1,415,376	1,214,830
R-squared	0.709	0.72	0.707	0.715	0.709
*** p<0.01, ** p<0.05, * p<0.1					

Figure 24: Heterogeneity in Short-Run Impacts by School-Level Air Conditioning Installation Status

Notes: The dependent variable is Regent exam score (0-100), with observations at student-subject-date-and-time-level. All regressions include controls for daily precipitation and dewpoint. Student, school, day of week, day of month fixed effects suppressed in output. Demographic controls include ethnicity, gender, subsidized school lunch status, and previous ability variable (combined math and ela z-score from 3rd-8th grade). For student FE and school FE regressions, 161383 and 7 singleton observations are dropped respectively. June Regents exams in 11 core subjects only, for years 1998-2011. Robust standard errors are clustered at school by date and time level.

Appendix

Supporting Evidence for Model Premises

Studies using variation in housing rents and house prices across different climates find that, controlling for other observable neighborhood and home characteristics, housing units in milder climates command substantial premiums, suggesting that individuals prefer mild temperatures and are willing to pay substantial amounts to live in places with more mild days and fewer extremely hot or extremely cold days (Rosen, 1974; Roback, 1982; Albouy et al, 2012; Sinha and Cropper, 2015).⁵⁵ The implied willingness to pay for this amenity aspect of climate alone has been estimated to be upwards of 3% percent of household income in the US to avoid business-as-usual global warming by 2100, or a 3.3 °F increase in mean summer temperature.

Given the relationship between cognitive or physical activity level, the body’s metabolic rate, and realized core body temperature, increased physical exertion may affect the marginal disutility of temperature stress in a non-linear fashion. Exerting physical or mental effort, which raises the body’s metabolism, can have one of two effects on experienced comfort and physical function. If ambient temperatures are above room temperature, then increased physical or mental activity will act to raise core body temperature further, exacerbating the adverse physiological impacts of elevated temperature. On the other hand, if ambient temperature is low (cold), then additional effort can offset some of the adverse impacts of cold stress. Whereas, under hot conditions, the increased metabolism from physical exertion pushes core body temperature further above the comfortable/optimal zone, under cold conditions, the reverse may be true, as increased metabolism moves core body temperature closer to the thermoregulatory optimum. That is, physical exertion may exacerbate temperature-related discomfort under hot conditions but may mitigate thermal discomfort under cold conditions.

In some sense, these empirical findings confirm the longstanding intuition that temperature and human flourishing may be intimately linked, as articulated by the late Singaporean Prime Minister Lee Kwan Yiu: *“Air conditioning was a most important invention for us, perhaps one of the signal inventions of history. It changed the nature of civilization by making development possible in the tropics. Without air conditioning you can work only in the cool early-morning hours or at dusk. The first thing I did upon becoming prime minister was to install air conditioners in buildings where the civil service worked. This was key to public efficiency (Lee, 1999).”*

⁵⁵The intensity of these preferences may vary across demographic groups. Some physiological studies suggest that men and women have divergent “optimal” temperatures due to differences in physiology (Sinha and Cropper, 2015), and that experienced temperatures may vary somewhat by basal metabolic rate, which has been shown to depend on age, BMI, and activity level.

A model of human capital accumulation under heat stress

Definitions:

- human capital/skills of student $i = h_i = h_i(e_i)$
- composite schooling effort/time investment of student $i = e_i$
- effectiveness of effort/time = $a(T) = (1 - \beta_T T)$
- extent of temperature stress (deviation from thermoregulatory optimum) = T

$$h_i = (1 - \beta_T T)e_i \quad (21)$$

$$a'(T) < 0 \quad (22)$$

The student's utility function is given by:

$$U_i = U_i(C_i, e_i) = U_i(w(1 - \beta_T T)e_i, e_i, T) \quad (23)$$

where

$$U_c > 0; \quad U_e < 0; \quad U_T < 0 \quad (24)$$

$$U_{cc} < 0; \quad U_{ee} < 0; \quad U_{TT} < 0 \quad (25)$$

Student chooses optimal time/effort investment, e_i , in the presence of some level of average heat stress T :

$$\max_{e_i} U_i(a(1 - \beta_T T)e_i, e_i, T) \quad (26)$$

First order condition:

$$\frac{dU}{de} = 0 \Leftrightarrow U_c w(1 - \beta_T T) + U_e = 0 \quad (27)$$

Intuition: invest in schooling such that the marginal pecuniary returns equal the marginal pecuniary costs. Using the implicit function theorem, we can differentiate the FOC with respect to T :

$$\frac{d}{dT} U_c w(1 - \beta_T T) + U_e \quad (28)$$

And solve for $\frac{de}{dT}$:

$$\frac{de}{dT} = \frac{w^2(1 - \beta_T T)eU_{cc} - w(1 - \beta_T T)U_{cT} + wU_c + weU_{ec} + U_{eT}}{(w(1 - \beta_T T))^2 U_{cc} + w(1 - \beta_T T)U_{ce} + w(1 - \beta_T T)U_{ec} + U_{ee}} \quad (29)$$

Proof of Proposition 1

Proposition 1:

$$\text{If } a'(T) < 0 \text{ and } U_T \leq 0, \text{ then: } \frac{\Delta U}{\Delta T} \leq 0.$$

Proof:

Note that $\frac{\Delta C}{\Delta T} \leq 0$, $U_C > 0$, $\frac{\Delta k_i}{\Delta T} \geq 0$, $U_k < 0$, and $U_T \leq 0$. This implies that, if $\frac{\Delta e_i}{\Delta T} \geq 0$, then $\frac{\Delta U}{\Delta T} \leq 0$.

What would need be true in order for $\frac{\Delta e_i}{\Delta T}$ to be *greater* than zero? In effect, the student would have to be at a point where she is not maximizing utility prior to the increased temperature, ΔT .

Proof of Proposition 2

Proposition 1:

$$\frac{\Delta e_i}{\Delta T} \leq 0$$

Proof:

$$\frac{\Delta e_i}{\Delta T} = \frac{w^2(1 - \beta_T T)eU_{cc} - w(1 - \beta_T T)U_{cT} + wU_c + weU_{ec} + U_{eT}}{(w(1 - \beta_T T))^2 U_{cc} + w(1 - \beta_T T)U_{ce} + w(1 - \beta_T T)U_{ec} + U_{ee}} \quad (30)$$

It is reasonable to suppose that $U_C > 0$, $U_{CC} < 0$, $U_e < 0$, $U_{ee} < 0$, $U_T < 0$, and $U_{TT} < 0$.

That is, utility is concave with respect to consumption, disutility of effort is convex, and disutility of heat exposure is also convex.

Then the above statement is true if we further assume that the cross-partials U_{Ce} and U_{eC} are equal, and that both marginal utility of consumption and marginal disutility of effort are unchanging with temperature ($U_{CT} = 0$ and $U_{eT} = 0$) or that both are negative ($U_{CT} < 0$ and $U_{eT} < 0$).

Distributional Incidence of Heat-Related Welfare Impacts

Several factors suggest that the realized welfare impacts of a given amount of heat stress may be larger for low SES individuals. First, if the impact of heat exposure depends on out-of-school adaptations such as air conditioning at home, we might expect heterogeneity in the impact of a given hot weather event across socioeconomic lines, due to the fact that they will be less likely to own such equipment. This may be because of credit constraints, as documented by Gertler et al (2016), or due to lower awareness of the adverse impacts of environmental stress generally.⁵⁶

Second, children who grow up in poverty have been shown to exhibit reduced executive function, which is critical in maintaining focus in presence of distractions (Fitzpatrick et al, 2014). It is possible that low ability students are more likely to be thrown off course due to the distracting influence of environmental factors such as temperature, noise, or pollution.⁵⁷ Similarly, students with low non-cognitive skills generally may be more easily discouraged by adverse test-taking conditions while students of similar cognitive ability but higher emotional resilience may be able to maintain concentration despite the heat.

Third, low-ability or low-SES students may face greater uncertainty over the returns to schooling investment, which becomes magnified in the context of adverse schooling or test-taking conditions. Suppose that, instead of the deterministic case presented above, wage returns to schooling are stochastic: $c_i = w(1 - \beta_T T) \cdot e_i + \gamma_i$, where $\gamma_i \sim (0, \sigma_i)$, and σ_i is a decreasing function of family income. That is, low SES individuals face a more uncertain mapping between educational attainment and future wages, due perhaps to fewer family connections or simply a perception that the system is unfair. If increased heat stress means that incremental time investments in schooling impose a certain direct disutility cost but an uncertain incremental wage return, risk-averse individuals may in principle optimally allocate time and effort away from schooling.

Finally, it may simply be the case that students from low-income households face higher opportunity costs of time in school, $U_{e_i}(I_i)$, where I denotes income endowment and $\frac{d}{dT}U_{e_i} < 0$, perhaps because there is greater pressure for teenage students to work part- or full-time to ease the immediate financial burdens faced by credit constrained or hand-to-mouth-consuming household members. This would imply that any reduction in the perceived pecuniary returns to schooling may elicit a larger effort/time response by students from low-income families.

⁵⁶It is possible that the poor have a lower implied WTP for climate control, given the documented gap in WTP between rich and poor countries for other environmental amenities including clean air and water (Greenstone and Jack, 2015), though it is as yet unclear as to why.

⁵⁷For instance, Lavy et al (2015) find that the test score impact of elevated particulate pollution is up to four times higher for low ability and low SES high school students in Israel.

Sample Restrictions and Test Date and Time Matching

I keep records for only the primary (June) takes of core Regents subjects, dropping language exams, which feature small sample sizes, and RCT exam, which are for students with disabilities, and other subjects that cannot be classified based on the original data code.

Regents exams are also offered during two additional waves each year — once in August, at the conclusion of summer school, and again in January, prior to the start of spring term. These scores are included as part of the NYC DOE data set, however I focus only on the subset of June exams, for two reasons. First, both August and January waves likely feature substantial selection on unobservables. Summer school is primarily remedial for public school students in NYC, meaning that August exams are taken only by a small subset of underperforming students who fail their June exams. Similarly, January exams are often taken by students who are enrolled in classes outside the usual order or cycle, either due to advanced standing or remediation. Second, unlike the main June wave, which is spread out over the course of 5 to 10 days, August and January exams are both administered over two days each year, reducing the amount of daily temperature variation available for identification. The main results presented in sections 5, 6 and 7 are not sensitive to the inclusion of January and August exams with month fixed effects, though the impact coefficient of temperature in a linear model is, somewhat predictably, reduced slightly in size, since January exams are likely to suffer from cold stress rather than heat stress.

For most exams, I matched publicly available exam date and time information to individual exam-level observations using the year, subject, and term codes provided by in the primary NYC DOE dataset.

For exams post-2008, no term codes are available. However, dates can be matched by using information in "item name" and "year". In effect, most post-2008 exams were recorded with consistent item name entries for the period 2009-2011, but not for 2012-2014. For instance, Algebra 1 was listed as "Algebra Reg" (2009, 2010, 2011) or "Algebra Reg June" and/or "Algebra Reg Jun" (2012, 2013, 2014) or "Common Core Algebra Reg Jun" (2014 only).

By looking at exam frequencies by item name and year, it is possible to assess whether the exams do indeed correspond to, say, June Regents exams in subject X, since June exams consistently have overwhelmingly large numbers of test takers. Since each subject is offered only once each wave (Jan, June, Aug) this seems like a safe matching strategy.

MATH B is basically non-existent in the list of subject codes, and is therefore dropped, bringing the number of subjects to 11.

Measuring Contemporaneous (Exam-time) Heat Exposure

Schools were assigned temperature values using a NASA Landsat satellite image taken of NYC on Aug 14, 2002 at 10:30 am. The specific Landsat temperature estimation image used here was prepared by Hunter College of NYC in 2004 (Gaffin et al, 2004). The Landsat temperature image was superimposed and transformed to fit onto a GIS rendering of the NYC metro area. The Landsat image estimates temperatures using a 10-color scale corresponding to different temperature ranges. ArcGIS was used to identify each of these distinct colors and then assign the median temperature value from the corresponding temperature ranges to these colors. The image was then rasterized using ArcGIS. The resulting raster has a grid cell resolution of 0.004 mi x 0.005 mi (or 6.4 meters x 8 meters). The temperature raster and a shapefile of NYC public school points were then spatially joined by location, resulting in each individual school being assigned the value of the raster cell in which it's corresponding shapefile point was located. The same process was used to assign temperature deviation values to each school.

Unfortunately, we were unable to obtain a GIS shapefile where the schools are represented by their actual spatial extent. We only have a shapefile in which schools are represented by a single point located at the school address. As a result, each school was given the value of the raster cell in which it's shapefile point fell in.

There are still many factors influencing realized temperature that cannot be fully accounted for. Given urban heat island effects and the occupational density of many school buildings (which often results in multiple floors), we would expect indoor air temperatures to be significantly warmer than outdoor temperatures in the case of schools without adequate AC. Moreover, the presence of AC heat exhaust spillovers would mean that a non-air conditioned building would bear the local externality of surrounding buildings that do have AC, a likely phenomenon given the nearly universal rates of commercial air conditioning in New York City (NYC Department of Health and Mental Hygiene, 2015).

Administration of NY State Regents Exams

The order of administration is as follows:

- NYSED creates subject-specific exam questions.
- NYSED sets exam dates a year in advance of the exam sitting.
- Psychometricians provide a scaling conversion chart ahead of time, based on their “structural” and “experimental” assessment of how the particular exam compares in terms of difficulty level. That is, they have a set of known and uniform difficulty anchor questions, and then test out the relative difficulty of other questions using supposedly representative students ahead of time.
- Students take exams.
- Teachers are instructed on how to grade, given detailed scoring guidelines (for instance, a multiple choice key and a detailed answer key), then asked to convert raw scores to scaled scores. Note that these scaling rubrics are not endogenous to temperature; they are determined ahead of time. Most exams are graded by multiple teachers.
- Prior to 2012, Teachers come together in a grading committee, grade the exams. They often graded their own students. Prior to 2012, NYC policy was that exams with scores between 60-64 needed to be regraded a second time. These policies were discontinued as of 2012.

Air Conditioning Data: NYC Building Condition Assessment Surveys

Information on building-level air conditioning equipment comes from the New York City School Construction Authority (SCA).

While the universe of survey results were not publicly available, building-specific 2012 BCAS reports were available in pdf format on the school's website for a large subset of schools (94%). I obtained survey records for 1,159 school buildings by downloading each pdf, scraping information pertaining to air conditioning equipment, all of which are recorded in the same format. The 1,159 records constitute 94.5% of the 1,226 unique school addresses in the New York City public school system as of 2012. Many NYC public schools are housed in buildings that contain multiple schools, especially in the case of middle and high schools. As such there are approximately 1,550 schools located at 1,226 unique addresses in our study sample.

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

I use this information to construct an indicator variable for whether or not a school building had non-defective central AC installed in any given year, using the earliest year of installation as the first year in which any AC was available for each school.

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

Additional Figures and Tables

Figure 25 Spatial Corrections to Account for Urban Heat Island Effects

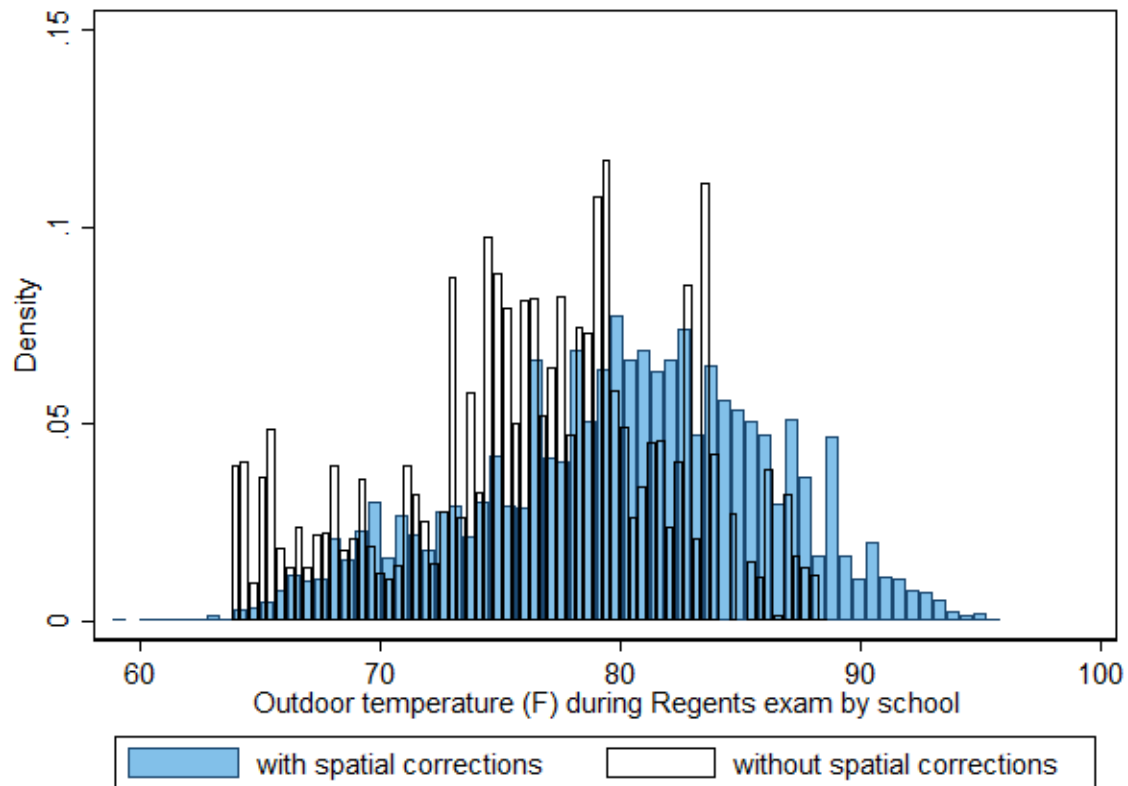


Figure 25: Temperatures measured at the school level, weighted by student population.

Notes: Temperatures for all June exam dates during the sample period (1998-2011), weighted by student population.

Figure 26 Household Income by Zip Code

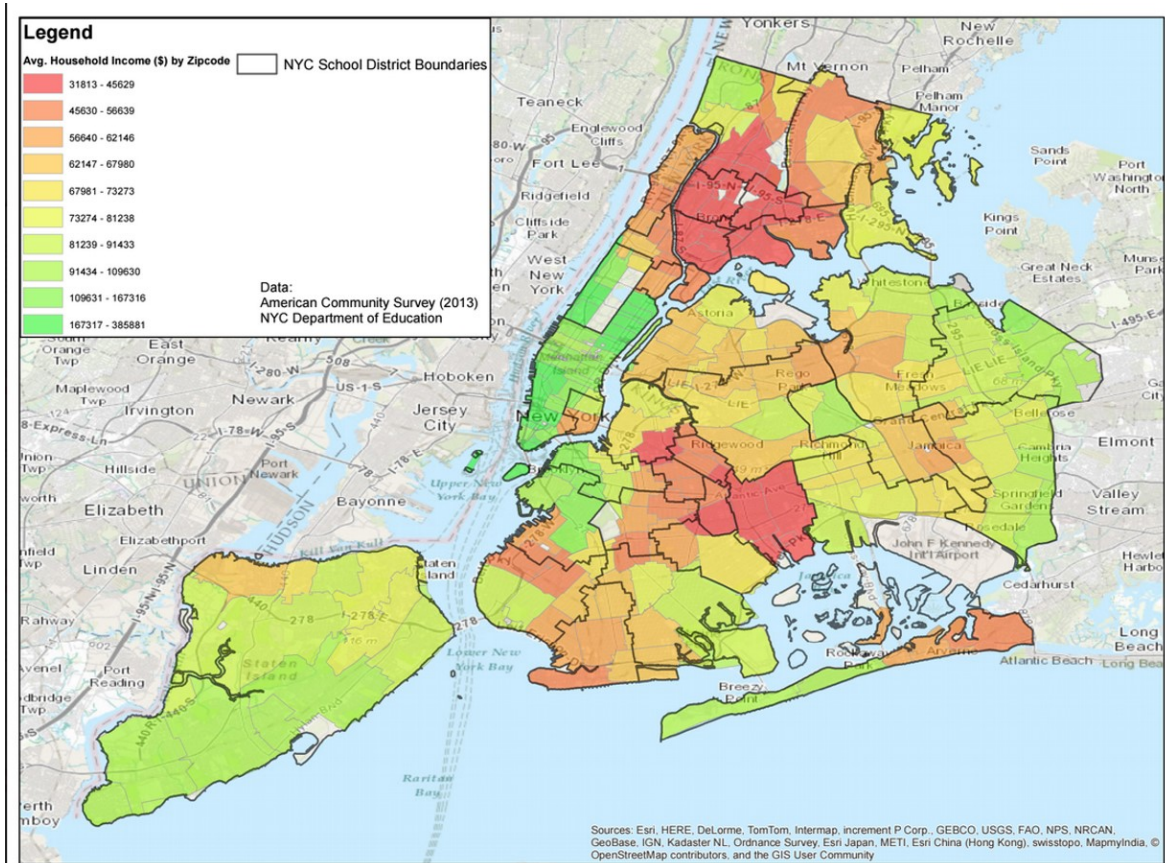


Figure 26: Average household income by zip code

Notes: This figure presents average household income in 2010 by zip code, with New York City Public School districts super-imposed.

Figure 27 Percent Black by District

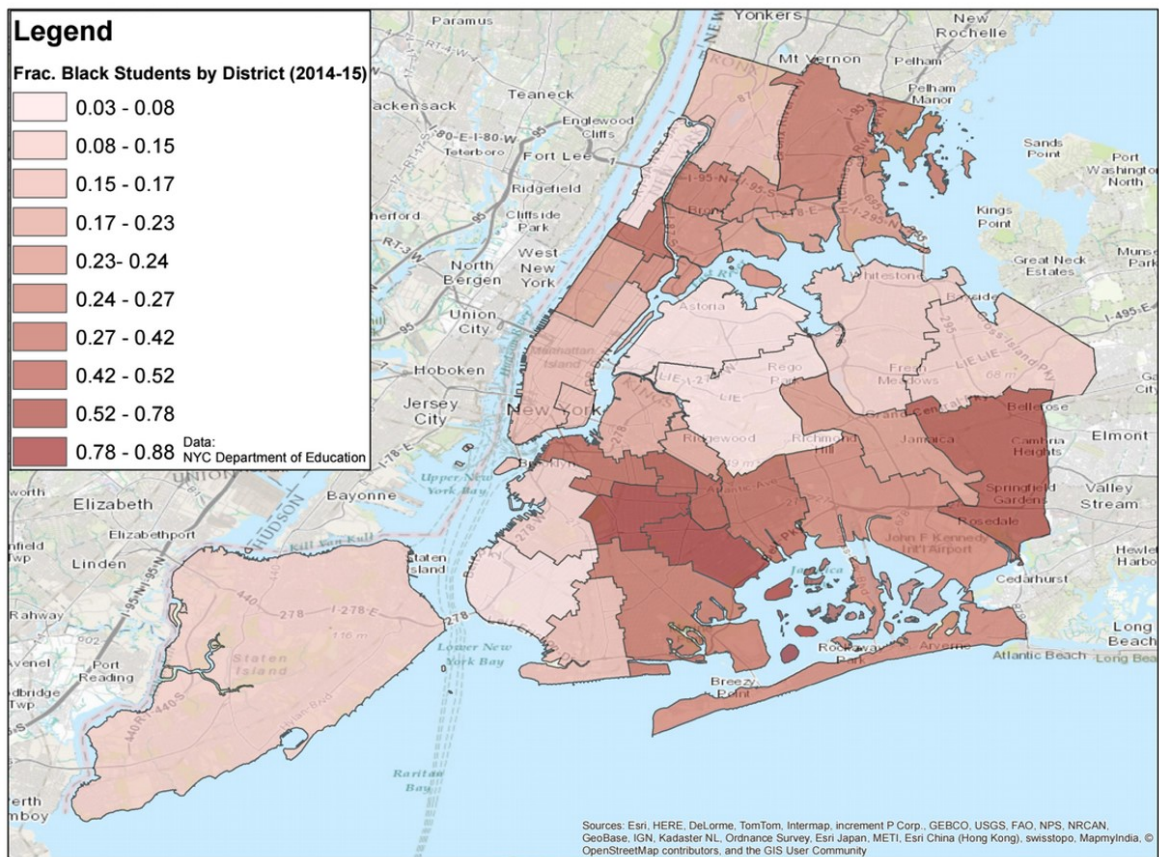


Figure 27: Black students by district

Notes: This figure presents the average percentage of black students in 2014-2015 by sub-district within the New York City Public Schools system.

Figure 28 Percent Hispanic by District

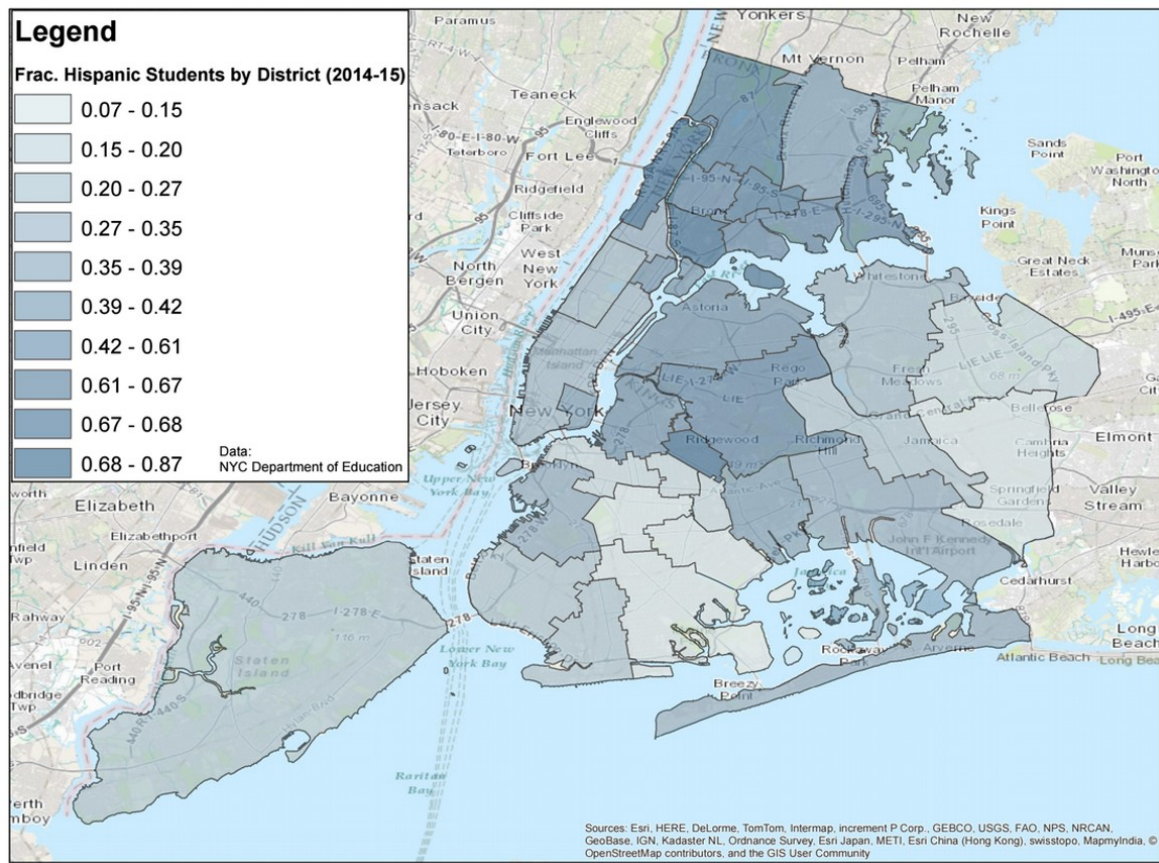


Figure 28: Hispanic students by district

Notes: This figure presents the average percentage of Hispanic students in 2014-2015 by sub-district within the New York City Public Schools system.

Figure 29 Urban Heat Island Effects in New York City: Satellite Reanalysis Data

Landsat Surface Temperature August 14 2002 10:30am

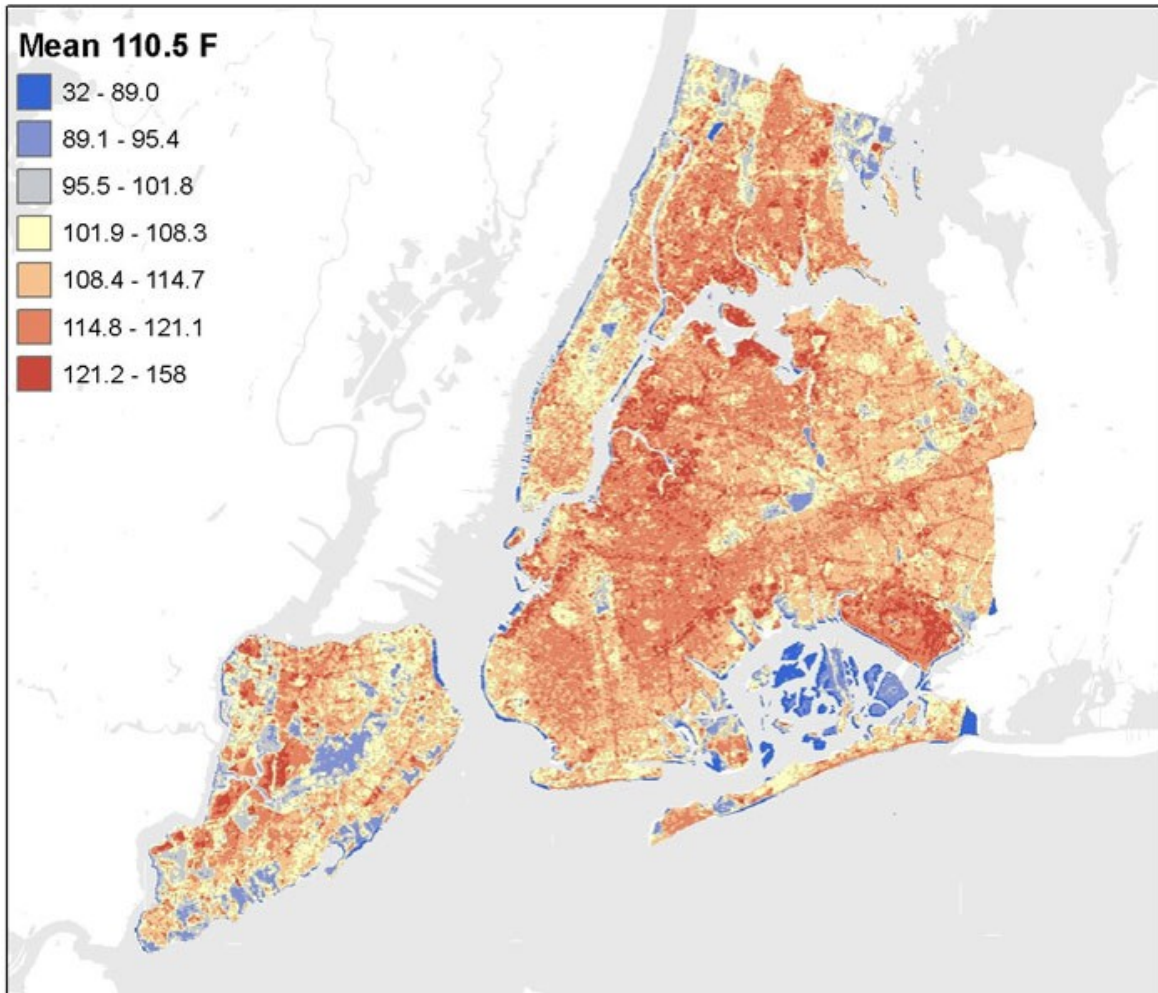


Figure 29: Surface temperature (2m) from a combined landsat/local atmospheric model reanalysis.

Figure 30 Short-Run Impact: Robustness Checks

Dep. Var	Score (1)	Score (2)	Score (3)
Temperature (F)	-0.156***	-0.149***	-0.141***
	0.042	0.040	0.047
Precipitation (10 mm)	0.064***	0.063***	0.061***
	0.018	0.018	0.019
Previous Z-scores (ELA+Math)	5.900***	5.900***	5.903***
	0.072	0.072	0.071
Asian	(omitted)	(omitted)	(omitted)
Black	-3.892***	-3.893***	-3.906***
	0.207	0.207	0.206
Hispanic	-3.855***	-3.856***	-3.868***
	0.185	0.185	0.184
White	-2.056***	-2.055***	-2.065***
	0.150	0.150	0.149
Female	-4.037***	-4.056***	-3.857***
	0.641	0.639	0.643
Subsidized School Lunch	-0.893***	-0.894***	-0.900***
	0.115	0.116	0.115
Time (Afternoon)		-0.374	-0.510*
		0.241	0.263
Day of Week			0.449***
			0.096
Day of Month			-0.055
			0.077
Fixed Effects			
School FE	X	X	X
Subject FE	X	X	X
Year FE	X	X	X
Time of Day FE		X	X
DOW, DOM FE			X
Observations	4,509,095	4,509,095	4,509,095
R-squared	0.403	0.403	0.404

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

Figure 30: Robustness checks for short-run impact: school fixed effects with full demographic controls including previous ability as measured by previous z-scores.

Notes: The dependent variable is Regents score (0-100), with observations at student-exam-date-and-time-level. All regressions include controls for daily precipitation and dewpoint. School, day of week, day of month fixed effects suppressed in output. Demographic controls include ethnicity, gender, and subsidized school lunch status. Robust standard errors are clustered at school by date and time level.

Figure 31 Previous Standardized Exam Scores and Regents Performance

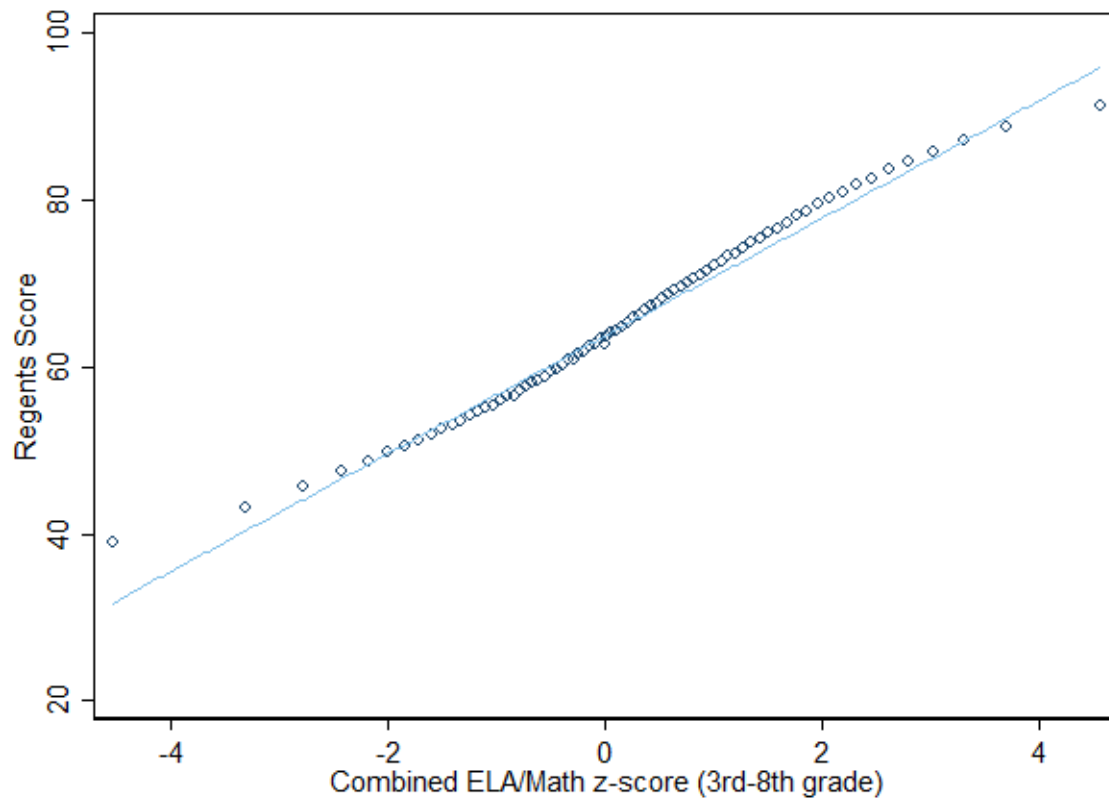


Figure 31: Combined ELA/math z-scores as a measure of student ability.

Notes: This figure presents a binned scatterplot of Regents scores and previous z-scores (ELA and math z-scores by grade and year, averaged by student) by percentile of the combined z-score distribution.

Figure 32 Sample Regents Exam Schedule (June 2016)

<p style="text-align: center;">The University of the State of New York THE STATE EDUCATION DEPARTMENT Office of State Assessment Albany, New York 12234</p> <p style="text-align: center;">EXAMINATION SCHEDULE: JUNE 2016</p> <p style="text-align: center;"><i>Students must verify with their schools the exact times that they are to report for their State examinations.</i></p>								
June 1 WEDNESDAY	June 14 TUESDAY	June 15 WEDNESDAY	June 16 THURSDAY	June 17 FRIDAY	June 20 [*] MONDAY	June 21 TUESDAY	June 22 WEDNESDAY	June 23 THURSDAY
9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	9:15 a.m.	RATING DAY
Algebra II (Common Core) ♦	RE in Global History & Geography	Living Environment	Algebra I (Common Core)	Physical Setting/ Earth Science Algebra 2/ Trigonometry	RCT in Mathematics* <small>* Suggested date for administering locally developed tests aligned to the Checkpoint A and Checkpoint B learning standards for languages other than English (LOTE).</small>	Physical Setting/ Chemistry RCT in Global Studies*	RCT in Writing	
1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	1:15 p.m.	<u>Uniform Admission Deadlines</u> Morning Examinations: 10:00 a.m. Afternoon Examinations: 2:00 p.m.
SPECIAL ADMINISTRATION: Integrated Algebra	RE in English Language Arts (Common Core)	RE in U.S. History & Government	Comprehensive English	Geometry (Common Core)	RCT in U.S. History & Government*	Physical Setting/ Physics RCT in Reading*	RCT in Science*	

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

Figure 32: Sample Regents Exam Schedule

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

ALGEBRA 2/TRIGONOMETRY

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

Student Name: _____

School Name: _____

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

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

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

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

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

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

Figure 33: Sample Regents subject exam cover sheet and questions

Figure 34 Mean-Reversion in Average Exam-time Temperature as the number of Observed Exam Takes Increases

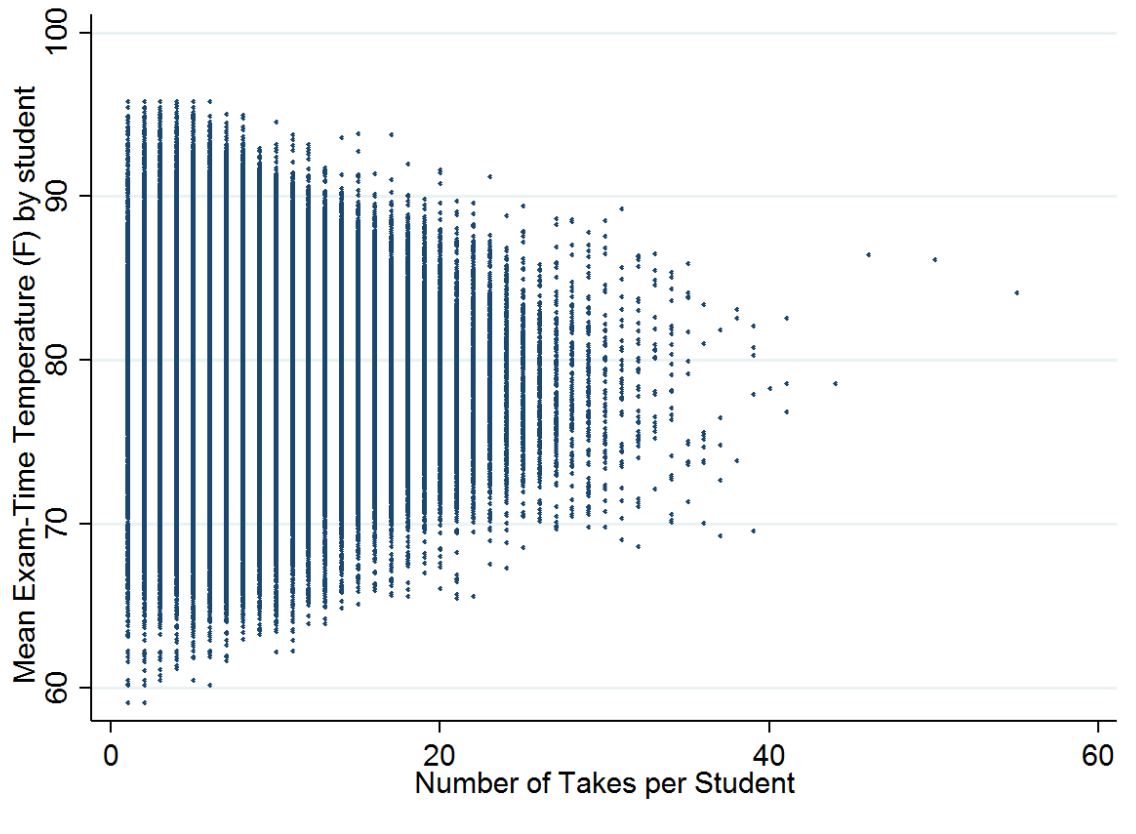


Figure 34: Variability in average experienced (June) exam-time temperature by student falls with number of takes, suggesting mechanical correlation between average exam-time temperature and graduation status if number of takes and unobserved student characteristics are correlated.