NBER WORKING PAPER SERIES

TEMPERATURE AND HIGH-STAKES COGNITIVE PERFORMANCE: EVIDENCE FROM THE NATIONAL COLLEGE ENTRANCE EXAMINATION IN CHINA

Joshua S. Graff Zivin
Yingquan Song
Qu Tang
Peng Zhang

Working Paper 24821
http://www.nber.org/papers/w24821

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2018

Temperature and High-Stakes Cognitive Performance: Evidence from the National College Entrance Examination in China
Joshua S. Graff Zivin, Yingquan Song, Qu Tang, and Peng Zhang
NBER Working Paper No. 24821
July 2018
JEL No. I23,I24,Q54

ABSTRACT

We provide the first nation-wide estimates on temperature effects on high-stakes cognitive performance in a developing country using data from the National College Entrance Examination (NCEE) in China. The NCEE is one of the most important institutions in China and affects hundreds of millions of families. We find that a one-standard-deviation increase in temperature (3.29° C) decreases the total test score by 1.12% (9.62% of a standard deviation) and decreases the probability of getting into first-tier universities by 1.97% (4.38% of a standard deviation). This suggests that temperature plays an important role in high-stakes cognitive performance and has potentially far-reaching impacts for the careers and lifetime earnings of students.

Joshua S. Graff Zivin
University of California, San Diego
9500 Gilman Drive, MC 0519
La Jolla, CA 92093-0519
and NBER
jgraffzivin@ucsd.edu

Yingquan Song
Peking University
No. 5 Yiheyuan Road
Haidian District
Beijing
China
songyingquan@pku.edu.cn

Qu Tang
Jinan University
601 Huangpu Avenue West
Guangzhou China
qutang@jnu.edu.cn

Peng Zhang
School of Accounting and Finance
M507C Li Ka Shing Tower
The Hong Kong Polytechnic University
Hung Hom Kowloon, Hong Kong
peng.af.zhang@polyu.edu.hk

1 Introduction

Every day, billions of humans solve numerous mathematical and logical problems, ranging in difficulty from basic counting to complex problem-solving. Individual health and well-being depend on these processes in myriad ways. Good nutrition requires careful tracking of caloric and nutrient intake. Workers must weigh the pros and cons of micro-decisions over the course of their day and even during their commute to the office. In ways large and small, risk calculations drive nearly all aspects of decision making. Thus, any threats to cognitive performance may have sizeable impacts on the health and well-being of human-based systems. Self-evidently, the impact of external factors on cognitive performance matters more when the decisions being made are important and have potentially life changing implications.

One such threat to cognition is temperature. Large segments of the population are regularly exposed to temporarily or persistently elevated temperatures. The brain's chemistry, electrical properties and function are all temperature sensitive (Bowler and Tirri, 1974; Schiff and Somjen, 1985; Deboer, 1998; Yablonskiy et al., 2000; Hocking et al., 2001). Moreover, exposure to heat has been shown to diminish attention, memory, information retention and processing, and the performance of psycho-perceptual tasks (Hyde et al., 1997; Hocking et al., 2001; Vasmatzidis et al., 2002). The impacts of thermal stress on working memory performance are especially relevant as cognitively challenging tasks rely heavily on the working memory for multi-step processing.

In this paper, we provide the first nation-wide estimates on temperature effects on high-stakes cognitive performance in a developing country using data from the National College Entrance Examination (NCEE), or *gaokao*, in China. The NCEE offers a useful means of examining the effect of heat on cognitive performance for several reasons. It is one of the most important institutional features of admissions to post-secondary education in China and affects the lives of hundreds of millions of families. Each year, around 9 million students take the exam to compete for admission to around 2,300 colleges and universities. Unlike other countries which rely upon standardized tests along with other factors such as high-school GPA, extracurricular activities, and recommendation letters to determine college admissions, the NCEE is almost the sole determinant for college admission in China[1], making it an extremely high-stakes exam. This is especially true for those aiming for the top-tier universities, as graduates can expect, on average, to earn 40% more per month than their counterparts from lesser universities (Jia and Li, 2017). The competition is fierce. Though the overall admission rate of test takers to college or university has been around 75% in recent years (China Education Online, 2016), the admission rate for the roughly 100 first-tier universities in China is only 12% (China Education Online, 2015).

Several other features of the NCEE make it particularly well-suited for measuring the causal effects of temperature on cognitive performance. First, the date of the NCEE is fixed, on June 7th and 8th, making self-selection on test dates impossible. Second, because the NCEE is held only once a year, the cost of retaking the exam is quite high, essentially requiring students to repeat an additional year of high school. Third, during our sample period of 2005-2011, students were required to take the exam in the same county as with their household registration (*hukou*). Therefore, self-selection on exam locations is heavily regulated. Finally, air conditioning is not available at testing facilities[2], thereby eliminating a potentially endogenous adaptation strategy,

[1] Less than 0.1% of students can gain admission to college without taking the NCEE (Bai et al., 2014). They usually take the exams administered by the university itself, or they are waived from having to take the NCEE because of special talent, such as the winners of National High-School Olympic Competitions.

[2] In regions where air conditioning is available, its use is prohibited during the test period to ensure fair competition with regions in which AC is not available. More details can be found at: http://news.sina.com.cn/c/2007-06-07/152711978182s.shtml and http://news.sina.com.cn/c/2014-06-05/070830296473.shtml.

and providing a better simulacrum of the conditions under which cognitive tasks are performed throughout the developing world where air conditioning penetration is quite low.

To examine the impact of temperature on the NCEE performance, we obtained a unique dataset that covers the universe of students who were admitted into college between 2005-2011 across China, yielding more than 14 million observations. The dataset reports the exam scores (ranging from 0 to 750) and exam counties for each student. We then match this dataset with daily weather data on temperature, precipitation, relative humidity, wind speed, sunshine duration, pressure, and visibility from more than 800 weather stations spread across the entire country.

We find both economically and statistically significant negative effects of temperature on test scores. In particular, a one-standard-deviation increase in temperature (3.29 °C) decreases total test scores by 1.12%, or approximately ten percent of one standard deviation in test performance. The effects are roughly linear in the temperature range found in China during early June – mean temperature during the exam period is 23.21 ℃. Given the significantly negative effect of temperature on exam scores, we then turn our attention to the effects of temperature on college admissions.[3] We find that a one-standard-deviation increase in temperature decreases the probability of getting into first-tier universities by approximately 2 percent. Together, these results indicate that temperature plays an important role in high-stakes cognitive performance and has potentially far-reaching impacts for the careers and lifetime earnings of students.

This paper builds upon a growing economics literature that examines the impacts of temperature on cognitive performance in a developed country context in which air conditioning is ubiquitous and the ramifications from underperforming on a test are significantly less

[3] In China, only students whose scores are above a pre-specified cutoff are eligible to apply for first-tier universities. Approximately 75% of students are admitted into first-tier universities if their scores are above the cutoff (Jia and Li, 2017). Since we do not have data on college admission, we proxy top-tier university admissions based on obtaining a score higher than the cutoff.

consequential (Park, 2017; Graff Zivin et al., 2018).[4] It also complements recent work by Garg et al. (2016), which finds that increases in annual temperature exposure in India can impair test performance largely through impacts on agricultural yields and nutrition. Our study also has implications for the study of standardized test performance more generally. While these tests are often viewed as gold standard for assessing the academic competence of students (Koretz and Deibert, 1996; Robelen, 2002; US Legal, 2014), recent studies have shown that the time the test is given as well as local air pollution can impact performance (see Sievertsen et al. (2015) and Ebenstein et al. (2016), respectively). Temperature appears to be another important factor to add to this list.

The NCEE is one of the important institutions in China that affects the lives of hundreds of millions of families in ways large and small (Bai et al., 2014; Chen and Kesten, 2016; Jia and Li, 2017; Cai et al., forthcoming). Since high temperatures impair the NCEE test performance, hotter regions may be unfairly penalized by the current system. We believe there exist at least three policy responses to remedy this injustice. First, the time of the NCEE might be shifted from June to cooler months, such as March, April, or May. In fact, the time of the NCEE was shifted once from July to June in 2003, to avoid the adverse effects of hot weather on students, but our results suggest that this shift was insufficient to fully address that concern. Second, AC could be installed and used in the exam rooms to help protect against the harmful effects of heat and level the playing field across regions which vary considerably in the average summertime temperatures. Finally, the college admission authorities could simply adjust test scores under the current system based on our

[4] While the relationship between long-run temperature exposure and changes in test scores have also been examined, these changes in test scores (controlling for weather during the test) reflect the impacts of weather on learning, as opposed to performance. Graff Zivin et al. (2018) find no such effects on learning, while recent work by Goodman et al. (2018) find evidence of very small effects that are completely offset by access to air conditioning.

estimates to ensure that environmental factors are purged from the assessment of intellectual capabilities.

The remainder of the paper is organized as follows. In Section 2, we introduce the background on the NCEE. In Section 3, we describe the data and the empirical strategy. We present our results in Section 4 and conclude in Section 5.

2 Empirical Background

The NCEE is a prerequisite for entrance into almost all higher education institutions at the undergraduate level in China. It is held annually, and is generally taken by students in their last year of high school. The NCEE has undergone continuous reform since 1978. It was once uniformly designed by the Ministry of Education (MOE) such that all the students across the country took exactly the same examination. In the early 2000s, the MOE launched the "unified examination, provincial proposition" reform (Zhu and Lou, 2011). Provinces and municipalities were allowed to customize their own exams independently, while the MOE continued to provide a national exam that could be used by provinces not employing independent exams. In 2011, 16 out of 31 provinces created customized exams while the others adopted national exam versions.

The most common examination format across provinces during our study period (from 2005 to 2011) is the two-day exam, which takes place annually on June 7th and 8th, and is scored on a 0-750 scale based on the "3+X" subjects system.[5] In the "3+X" subjects system, "3" refers to the three compulsory subjects: Chinese, Mathematics, and a foreign language usually English (each

[5] Six provinces take a three-day exam on Jun 7th, 8th and 9th, including Shanghai, Jiangsu and Guangdong from 2005 to 2011, Hainan and Shandong from 2007 to 2011, and Zhejiang from 2009 to 2011. Four provinces use a scale rather than 0-750 marks, including Hainan 0-900 from 2005 to 2011, Guangdong 0-900 from 2005 to 2006, Jiangsu 0-440 in 2008, and Shanghai 0-630 from 2005 to 2011. We normalize the scale to 750 marks for these four provinces.

accounting for 150/750 of the total score) and "X" refers to the combination of science subjects (biology, chemistry, and physics) for students on the science track, or the combination of art subjects (geography, history, and political science) for students on the art track (accounting for 300/750 of the total score).[6]

The NCEE is an extremely high-stakes exam. It is almost the sole determinant for higher education admission in China. Every year, around 9 million students in China take the exam to compete for admission to approximately 2,300 colleges and universities. These colleges are divided into two hierarchical categories: regular colleges and universities that are degree-granting and academically oriented; and advanced vocational colleges that certify students based on the attainment of practical and occupational skills. Though the overall admission rate of exam takers to both forms of higher education ranges from 57% to 72% during our study period, the admission rate for former category is only around 30%.[7] The regular colleges and universities can further be classified into three tiers according to the recruitment process – Tier 1 universities, generally regarded as elite or key universities, recruit before Tier 2 and Tier 3 universities and require a much higher cut-off score for admission.[8] Admission rates for Tier 1 universities in recent years has hovered around 12% and was even lower in earlier years (China Education Online, 2015). The higher the NCEE score the greater the chance that a student can attend an elite university, which is highly correlated with future life opportunities and earning potential (Jia and Li, 2017).

[6] There are also a small number of specialized tracks, which include sports, art (music, painting, dancing), military, and pedagogical. These constitute less than 10% of all track specializations in our data. All students choose their track of study prior to the start of their second year of high school.

[7] See annual statistical data from the MOE: http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_1651/index.html

[8] The cut-off score for each tier is the minimum qualifying score for students to apply to universities of the tier, and varies annually across provinces and subject tracks. It is determined by the Provincial Admission Offices based on each year's admission quota and the distribution of student scores within the province (Chen and Kesten, 2017).

3 Empirical Strategy

3.1 Data

We obtain the NCEE data from the China Institute for Educational Finance Research at Peking University, which reports the total score and ID for the universe of students enrolled into college during 2005-2011.[9] This dataset includes observations for roughly 2 million students each year. The student ID contains a six-digit code for county of residence, which we use to match with weather data. The ID also reports the specific track, allowing us to explore heterogeneity across the science and art tracks. Unfortunately, we do not have data on scores by specific subject nor information on where each student enrolls. Data on the cut-off scores that determine eligibility to apply to first-tier universities for each province-year-track are obtained from a website specialized for the exam: gaokao.com.

The weather data are obtained from the China Meteorological Data Service Center, which is an affiliate of the National Meteorological Information Center of China. The data report daily maximum, minimum and average temperatures, precipitation, relative humidity, wind speed, sunshine duration, and atmospheric pressure for more than 800 weather stations in China. Data on visibility are obtained from the National Oceanic and Atmospheric Administration of the U.S. We extract weather data during the exam time and then convert from station to county using the inverse-distance weighting (IDW) method (Deschênes and Greenstone, 2007, 2011). The basic algorithm calculates weather for a given county based on weighted averages of all weather station

[9] Note that the sample only includes students admitted into some type of college or university since test score data are not available for those students that are not admitted to a school of higher education. Whether our results generalize to those students that do not attend college is an open question that will depend on the relative temperature sensitivity of this population.

observations within a 200 km radius of the county centroid, where the weights are the inverse distance between the weather station and the centroid.

Panel A of Table 1 presents the summary statistics of the exam score, which ranges from 0 to 750. There are more than 14 million observations in total, with a mean of 518.96, and a standard deviation of roughly 60 marks. Approximately 26% students were in the art track, and 64% students were in the science track, with the remaining 10% corresponding to students in specialized tracks.

In panel B, we define a dummy variable which is equal to one if a student's score is above or equal to the cutoff for the first-tier universities, as a proxy for admission into a first-tier university. Approximately 75% of students with a score above the cutoff are admitted into first-tier universities (Jia and Li, 2017). That corresponds to approximately 30% of the students in our sample.[10] Admission rates for the science track are higher than the art track using this proxy, a results that is consistent with actual admissions patterns at top-tier universities..

We report the summary statistics of weather variables in panel C. The average mean temperature during the exam period is 23.21 °C. The histogram of average temperature during the 2-day exam period is plotted in Figure 1. This figure reveals a great heterogeneity, with temperatures ranging from 10 °C to 30 °C, and a peak around 25 °C. To measure the non-linear effects of temperature, we construct two measures. The first is degree days (DD), which is a piece-wise linear function that measures the number of degrees above and below a threshold. Following Graff Zivin et al., (2018), we deploy a threshold of 20 °C. As can be seen in panel C of Table 1, the average degree days above or equal 20 °C (DD>=20) is 3.59, and the average below 20 °C (DD<20) is 0.38, consistent with the skewed distribution of temperature seen in Figure 1. The

[10] This rate is higher than the 10% admission rate for the entire population of high school graduates since our sample only includes students who enrolled into an institution of higher learning.

second measure we deploy to capture non-linear effects is a series of indicators of 2 °C bins (Deschênes and Greenstone, 2011; Barreca et al., 2016; Graff Zivin and Neidell, 2014; Graff Zivin et al., 2018), with the lowest bin including all temperatures below 12 °C and the highest bin including all temperatures above 30 °C due to data sparseness at the extremities of the distribution. Figure A1 in the online appendix plots the percentage of days that fall into each bin.

3.2 Econometric Model

In order to assess the effect of temperature on students' performance, we estimate the following equation:

$$Y_{ict} = \alpha_0 + \beta_1 T_{ct} + \beta_2 \boldsymbol{W}_{ct} + \gamma_c + \eta_t + \varepsilon_{ict},$$

where i denotes an individual student, c denotes the county in which the exam was taken, and t denotes the year the exam was taken. We have two measures for Y_{ict}. The first is the logarithm of the exam score. The logarithm specification was chosen to facilitate interpretation, since point estimates correspond to the semi-elasticity of exam scores with respect to temperature. As we will show later, our results are also robust to specifying exam scores in levels. The second is a dummy variable which is equal to one if a student's score is equal to or higher than the cutoff for first-tier universities and zero otherwise. Both specifications are estimated using OLS, although our results for admission to elite universities remain unchanged when we use a logit specification (as shown in Table 6). We use T_{ct} to denote the average of daily mean temperature (the average between daily maximum and minimum temperatures) on June 7th and 8th. We do not include temperature in each day separately because of the strong serial correlation in temperature across days. To explore the non-linearity of temperature, we use degree day measures and a series of 2 °C bins as described earlier.

The variable $\boldsymbol{W}_{ct}$ denotes a vector of weather variables, including precipitation, relative humidity, wind speed, sunshine duration, atmospheric pressure, and visibility. As with our temperature variable, all of these are averaged across the two-day exam period. We use γ_c to denote county fixed effects, which controls for any county-specific time-invariant characteristics, such as geography or cultural and demographic features that are stable over our study period. We use η_t to represent year fixed effects, to control for any nation-wide policy or economic shocks that could differ by year but affect test takers equally across all counties. The error terms ε_{ict} are clustered by county to allow for serial and spatial correlation within each county.

In the end, our identifying variation is based on county deviations from the mean after we adjust for common shocks for the whole country in a given year. One potential concern with this strategy is that exam difficulty varies by province-year, but we cannot include year-by-province fixed effects because they absorb most of our variation in weather (Fisher et al., 2012). Since the absence of these fixed effects are only a concern if exam difficulty across provinces is correlated with temperature, we test this directly. Column (1) of Table 2 reports the average weather for provinces that use their own exams, and column (2) reports the average weather for provinces using national exams. Temperatures in provinces that use their own exams are 1.90 °C higher than temperatures in provinces using national exams (column 3). This difference disappears once the comparison is conditional on the county and year fixed effects that are included in our model specification (see column 4).

The coefficient of interest is β_1. Under our linear measure of temperature, this coefficient measures the percentage change in total score (or the probability change of admission to first-tier universities) when temperature during the exam increases by 1 °C. When we use degree days, the coefficient of DD>=20 (DD<20) measures the percentage change in total score (or the probability change of admission to first-tier universities) if temperature increases (decreases) by 1 °C conditional on temperature being above (below) 20 °C. The non-linear binned approach has a slightly different interpretation. Here the coefficient of each bin measures the percentage change in total score (or the probability change of admission to first-tier universities) when temperature falls into that bin rather than the reference bin of 18-20 ℃ which was chosen following Graff Zivin et al. (2018).

4 Results

4.1 Main Results

Table 3 presents the main regression results, where outcomes are defined as the logarithm of the total test score. The total test score is the summation of scores from three compulsory subjects, including Chinese, mathematics, and foreign language (typically English) with 150 marks each plus scores from one combined subject with 300 marks comprising politics, history, and geography for the art track and physics, chemistry, and biology for the science track. Unfortunately, the data does not report the score for each specific subject. We report results for all students in columns (1) and (2), only students in the art track in columns (3) and (4) and only those in the science track in columns (5) and (6).

In columns (1), (3), and (5), temperature is measured using the average of daily mean temperature during June 7^{th} and 8^{th}. All the estimates are negative and statistically significant at the 1% significance level. The coefficient of temperature in column (1) suggests that a 1 ℃ increase in temperature decreases the total test score by 0.34%, or 1.76 marks evaluated at the mean level (mean=518.96). To better place these figures in context, it is helpful to situate them relative to the weather variability in our dataset. A one-standard-deviation increase in temperature (3.29 ℃) decreases total test scores by 1.12%, or 9.62% of a standard deviation (standard deviation=60.40).

In columns (2), (4), and (6), we relax the assumption of linearity by specifying temperature in terms of degree days as described above. As can be seen in column (2), the effect of DD>=20 is significantly negative, and the magnitude is quite close to the linear effect in column (1). This near identical result is largely an artefact of exam timing. June temperatures in China tend to be quite high, with a mean temperature of 22.78 ℃ , which lies above the degree days threshold. In contrast, the effect of DD<20 is significantly positive, providing additional support for our linear specification.

When we run subsample analyses for each track separately (see columns (3) – (6)), we find that the negative effect of temperature is much larger for students in the art track than those in the science track. For example, a 1 ℃ increase in temperature decreases the score for the art track by 0.36%, but only by 0.18% for the science track. Evaluated at the mean level, this is equivalent to 1.85 marks for the art track (mean=512.66) and 0.94 marks for the science track (mean=521.20). One possible explanation for this difference is sample composition. The art track is disproportionately female relative to the science track and recent research suggests that female test performance in China may be more stress-dependent (Cai et al., forthcoming).

In addition to temperature, we also include precipitation, relative humidity, wind speed, sunshine duration, pressure, and visibility in the regression model. We find a significantly positive effect of wind speed, consistent with the notion that higher wind speeds reduce perceived temperature – the effect of so-called wind chill.[11] The effect of sunshine duration is also significantly positive, as many studies find that sunshine induces good mood and happiness

[11] http://www.nws.noaa.gov/om/cold/wind_chill.shtml.

(Schwarz and Clore, 1983; Guven, 2012) and further increases labor productivity (Osward et al., 2015). The effect of precipitation, humidity, pressure, and visibility are either weakly significant or statistically insignificant.

In Table 4, we turn our attention to the effects of temperature on admissions to elite universities. Our estimate in column (1) suggests that a 1 ℃ increase in temperature decreases the probability of being admitted to first-tier universities by 0.60%, or 1.33% of a standard deviation (standard deviation=0.45). When we measure temperature using degree days in column (2), we find that a 1 ℃ increase in temperature above 20 ℃ decreases the admission probability by 0.75%. As with the linear results, we also find that the effect is larger for those students in the art track. Interestingly, the impacts of other weather variables differ under this specification, with the coefficients on precipitation, humidity, and pressure all statistically significant, small (relative to their mean values) and in the expected direction. The coefficient on visibility, a proxy measure for pollution, is also negative and statistically significant and reasonably large (a result we will explore more directly below).

Figure 2 plots the coefficients (in blue) as well as 95% confidence intervals (in grey) under our non-parametric binned approach when the dependent variable is the log of exam score. As noted earlier, the 18-20 ℃ bin is omitted as the reference group, so all other estimates are relative to it. We find that the coefficient decreases monotonically for all bins hotter than 12-14 ℃. The magnitude here is also comparable to column (1) in panel A of Table 3. For example, the estimated coefficient for the above 30 ℃ bin is -0.0310. Since the difference between bins above 30 ℃ and 18-20 ℃ is approximately 10 ℃, each 1 ℃ increase in temperature decreases a score by 0.0031 (0.0310/10) log points (-0.0034 log points in column (1) of Table 3). We conduct a similar exercise in Figure 3, where the dependent variable is the dummy variable for admission to first-tier universities, and find similar results.

Studies show that AC can protect the human body from harms due to excess heat (Barreca et al., 2016) and it seems plausible that these protective effects might also extend to cognitive performance. Unfortunately, we do not have data on the availability of AC at test facilities. Moreover, AC use is supposed to be prohibited during the NCEE to ensure fairness across regions, some of which clearly do not have AC. Nonetheless, we explore the potential role of AC indirectly, by splitting our sample into urban districts and rural counties,[12] under the assumption that cities are more likely to have AC. Table A1 in the online appendix reports these results. The effects of temperature appear larger in urban districts than rural ones, although these differences are not significant at conventional levels. Whether the lack of difference suggests a limited protective role for air conditioning, the effectiveness of the policy ban on usage, or the noisiness of our AC measure remains an open question.

Since others have found that exposure to fine particulate matter less than 2.5 microns in diameter ($PM_{2.5}$) can also impair test performance (Ebenstein et al., 2016), one concern with our study is that our results may be confounded by air pollution levels in ways that are not fully captured by our controls for visibility. To examine this issue directly, we use data on the air pollution index (API) – a composite measure of pollution that ranks air quality based on its associated health risks (Ministry of Environmental Protection, 2006) – to examine the relationship between air quality and test performance.[13] The API is only available in major cities and thus our sample size for this analysis is greatly reduced. The estimates are reported in Table 5. Column (1) reports the baseline estimates from Table 3, column (1). Column (2) reports results from the same specification but only for the sample of cities covered by the API. In column (3)

[12] In China, districts (*qu*) and counties (*xian*) are in the same administrative level, but districts are typically located in urban cities.

[13] The data on $PM_{2.5}$ are only available since 2013.

we add controls for pollution as measured by the API. Though the sample size in columns (2) and (3) is less than half of column (1), the effect of temperature remains unchanged, which suggests that air pollution is not driving our temperature results.

While the results on API are reassuring, it remains possible that $PM_{2.5}$ could be confounding our results. To further probe this possibility, we utilize data from a more recent period when that data are available to examine the correlation between $PM_{2.5}$ and temperature. These results are reported in Tables A2 – A4 in the online appendix for the period 2013-2016. Regardless of functional form, the correlation coefficients are small, providing additional evidence that $PM_{2.5}$ is unlikely to explain the relationship between temperature and test performance in our setting.

4.2 Robustness Checks

Table 6 presents robustness checks for our main results. Column (1) is the baseline model. In column (2) we cluster the standard errors by prefecture (an administrative unit between province and county), to control for spatial and serial correlation within each prefecture. The effect of temperature is still statistically significant at the 1% significance level. In column (3), we calculate the average temperature between June 7th and 8th using daily maximum temperature, instead of daily mean temperature in the baseline model. The effect is robust, though the magnitude is slightly smaller, consistent with the observation that most of the testing period occurs before the hottest part of the day. In column (4) of panel A, we use level of score, instead of log of score in the baseline model, as the dependent variable. The point estimate is very close to the estimate when we use the log of score and evaluated at the mean level. In column (4) of panel B, we use the logit model and report the marginal effect evaluated at the mean level. The estimate is similar to the linear model. While the NCEE is held in most provinces on June 7th and 8th only, some provinces also have exams on June 9th. Therefore, in column (5), we calculate the average of temperature on June 7th-9th for provinces with a three-day exam. The results are robust. In the last column, instead of using individual-level score data in the baseline model, we average scores to county-year and then estimate the regression model to reflect the fact that the weather data are only at county-year level. Again our results remain robust.

5 Discussion and Conclusion

In this paper, we show that temperature plays an important role in high-stakes cognitive performance using data from the NCEE, the most important academic examination in China. In particular, a one-standard-deviation increase in temperature decreases total test scores by 9.62% of a standard deviation. This is approximately two times larger than the effects found by Park (2017) for similarly aged students in New York City and 1.5 times larger than that found by Graff Zivin et al. (2018) for younger children across the United States. The larger magnitude in our setting may be a reflection of the higher-stakes environment, the limited access to air conditioning, or fundamental differences in our study populations. It is also noteworthy that our estimates suggest that a one-standard-deviation change in temperature alters test performance by roughly the same magnitude as the impacts of the most successful interventions that directly target educational performance in developing countries J-PAL (2014).

Our results also imply that students in hotter regions may have disadvantages compared with their peers in cooler regions, highlighting potentially important concerns about equitable access to higher education within China under the NCEE system. We believe there exist at least three policy responses to remedy this inherent unfairness in the national testing and admissions system. First, the time of the NCEE might be shifted from June to cooler months, such as March, April, or May. In fact, the time of the NCEE was shifted once from July to June in 2003, to avoid the adverse effects of hot weather on students, but our results suggest that this shift was insufficient to fully address that concern. Second, AC could be installed and used in the exam rooms. Ironically, some regions prohibit the use of AC to enhance the fairness to regions where air conditioning is not available, which misses the important point that some regions are always hotter than others and that the use of AC may have leveled the playing field in the first place. Third, college admission authorities could adjust the test scores based on our estimates. For example, they may adjust upward (downward) test scores by 0.34% for counties with temperature above (below) provincial average every 1 ℃.

Though our empirical setting is China, our results have important implications for other developing countries that utilize standardized testing to gate access to institutions of higher learning or access to particular professions. Whether these results generalize to a developed country setting, where AC is more prevalent, remains an open question. Nonetheless, the significant effect of temperature on cognitive performance suggests a potential channel through which future climate change may affect economic well-being.

References

Bai, Chong-en, Wei Chi, and Xiaoye Qian (2014). "Do college entrance examination scores predict undergraduate GPAs? A tale of two universities." *China Economic Review* 30: 632-647.

Barreca, Alan, Karen Clay, Olivier Deschênes, Michael Greenstone, and Joseph S. Shapiro (2016). "Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the twentieth century." *Journal of Political Economy* 124, no. 1: 105-159.

Bowler, K., and R. Tirri (1974). "The temperature characteristics of synaptic membrane ATPases from immature and adult rat brain." *Journal of Neurochemistry* 23, no. 3: 611-613.

Cai, Xiqian, Yi Lu, Jessica Pan, and Songfa Zhong (forthcoming). "Gender gap under pressure: Evidence from China's National College Entrance Examination." *Review of Economics and Statistics*.

Chen, Yan, and Onur Kesten (2017). "Chinese college admissions and school choice reforms: A theoretical analysis." *Journal of Political Economy* 125, no. 1: 99-139.

China Education Online (2015). http://gaokao.eol.cn/zhiyuan/zhinan/201512/t20151225_1351335.shtml.

China Education Online (2016). http://www.eol.cn/html/g/report/2016/index.shtml.

Deboer, Tom (1998). "Brain temperature dependent changes in the electroencephalogram power spectrum of humans and animals." *Journal of sleep research* 7, no. 4: 254-262.

Deschênes, Olivier, and Michael Greenstone (2007). "The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather." *The American Economic Review* 97, no. 1: 354-385.

Deschênes, 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, no. 4: 152-185.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth (2016). "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution." *American Economic Journal: Applied Economics* 8, no. 4: 36-65.

Fisher, Anthony C., W. Michael Hanemann, Michael J. Roberts, and Wolfram Schlenker (2012). "The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather: Comment." *The American Economic Review* 102, no. 7: 3749-3760.

Garg, Teevrat, Maulik Jagnani, and Vis Taraz (2016). “Human capital costs of climate change: evidence from test scores in India”.

Goodman, Joshua, Michael Hurwitz, Jisung Park, and Jonathan Smith (2018). “Heat and learning.” *NBER Working Paper W24639*.

Graff Zivin, Joshua, and Matthew Neidell (2014). “Temperature and the allocation of time: Implications for climate change.” *Journal of Labor Economics* 32, no. 1: 1-26.

Graff Zivin, Joshua, Solomon M. Hsiang, and Matthew Neidell (2018). “Temperature and human capital in the short and long run.” *Journal of the Association of Environmental and Resource Economists* 5, no. 1: 77-105.

Guven, Cahit (2012). “Reversing the question: Does happiness affect consumption and savings behavior?” *Journal of Economic Psychology* 33, no. 4: 701-717.

Hocking, Chris, Richard B. Silberstein, Wai Man Lau, Con Stough, and Warren Roberts (2001). “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: 719-734.

Hyde, Dale, John R. Thomas, John Schrot, and W. F. Taylor (1997). *Quantification of special operations mission-related performance: Operational evaluation of physical measures*. No. NMRI-97-01. Naval Medical Research Institute, Bethesda MD.

J-PAL (2014). https://www.povertyactionlab.org/policy-lessons/education/ increasing-test-score-performance.

Jia, Ruixue, and Hongbin Li (2017). “Access to elite education, wage premium, and social mobility: The truth and illusion of China’s College Entrance Exam.”

Koretz, Daniel, and Edward Deibert (1995). “Setting standards and interpreting achievement: A cautionary tale from the National Assessment of Educational Progress.” *Educational Assessment* 3, no. 1: 53-81.

Ministry of Environmental Protection of China (2006). http://jcs.mep.gov.cn/hjzl/200604/ t20060428_76218.shtml.

Oswald, Andrew J., Eugenio Proto, and Daniel Sgroi (2015). “Happiness and productivity.” *Journal of Labor Economics* 33, no. 4: 789-822.

Park, Jisung (2017). “Hot temperature, high stakes exams, and avoidance behavior: Evidence from New York City public schools”.

Robelen EW (2002). “An ESEA primer”. *Educ Week* Feburary:21.

Schiff, Steven J., and George G. Somjen (1985). "The effects of temperature on synaptic transmission in hippocampal tissue slices." *Brain Research* 345, no. 2: 279-284.

Schwarz, Norbert, and Gerald Clore (1983). "Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states". *Journal of Personality and Social Psychology* 45: 513–23.

Sievertsen, Hans Henrik, Francesca Gino, and Marco Piovesan (2016). "Cognitive fatigue influences students' performance on standardized tests." *Proceedings of the National Academy of Sciences* 113, no. 10: 2621-2624.

US Legal (2014). "Standardized test [education] law and legal definition", https://definitions.uslegal.com/s/standardized-test-education.

Vasmatzidis, Ioannis, Robert E. Schlegel, and Peter A. Hancock (2002). "An investigation of heat stress effects on time-sharing performance." *Ergonomics* 45, no. 3: 218-239.

Yablonskiy, Dmitriy A., Joseph JH Ackerman, and Marcus E. Raichle (2000). "Coupling between changes in human brain temperature and oxidative metabolism during prolonged visual stimulation." *Proceedings of the National Academy of Sciences* 97, no. 13: 7603-7608.

Zhu, Hong Zhen, and Shiyan Lou (2011). *Development and reform of higher education in China*. Elsevier, 2011.

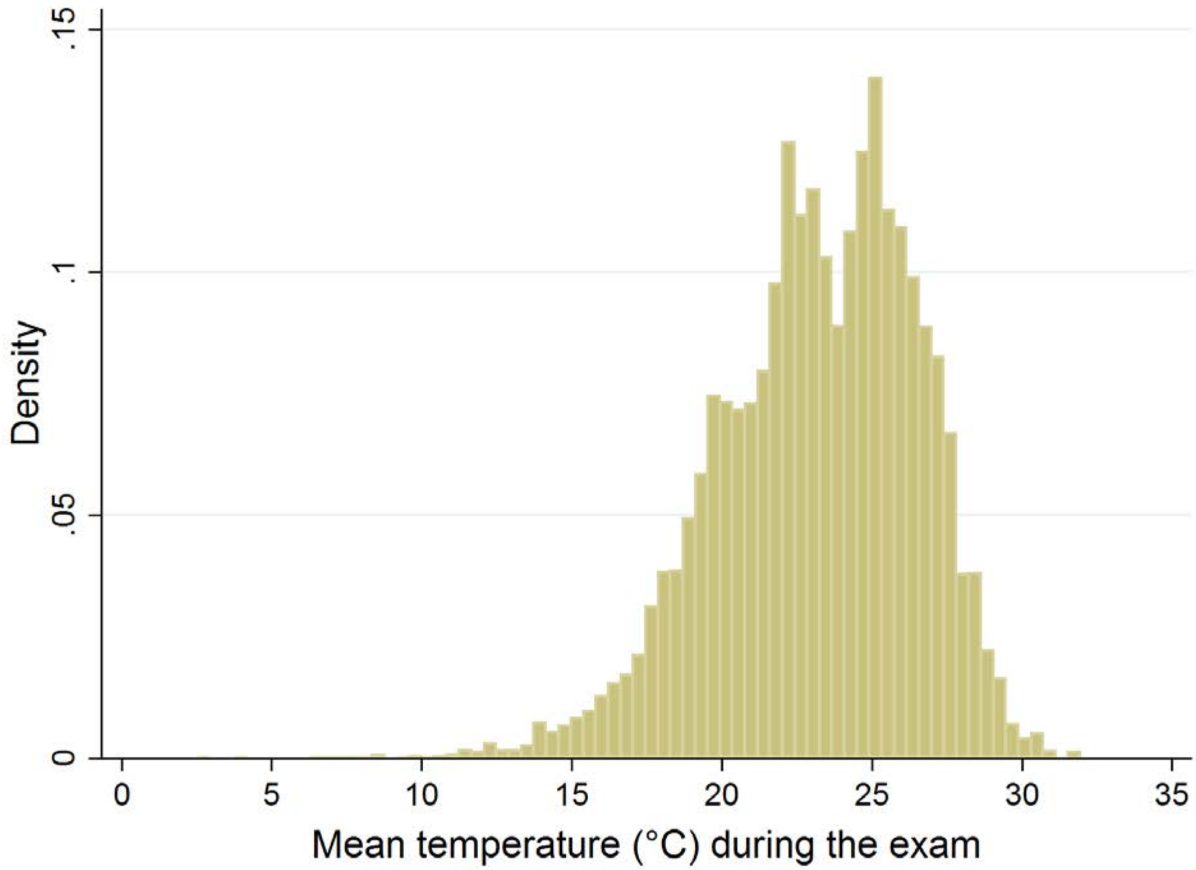

Figure 1: Histogram of mean temperature (°C) during the exam

Notes: Mean temperature over this two-day period is defined as the average of the daily average temperature on June 7th and 8th over 2005-2011. As is standard practice, the daily average temperature is the average of the daily maximum and minimum temperatures.

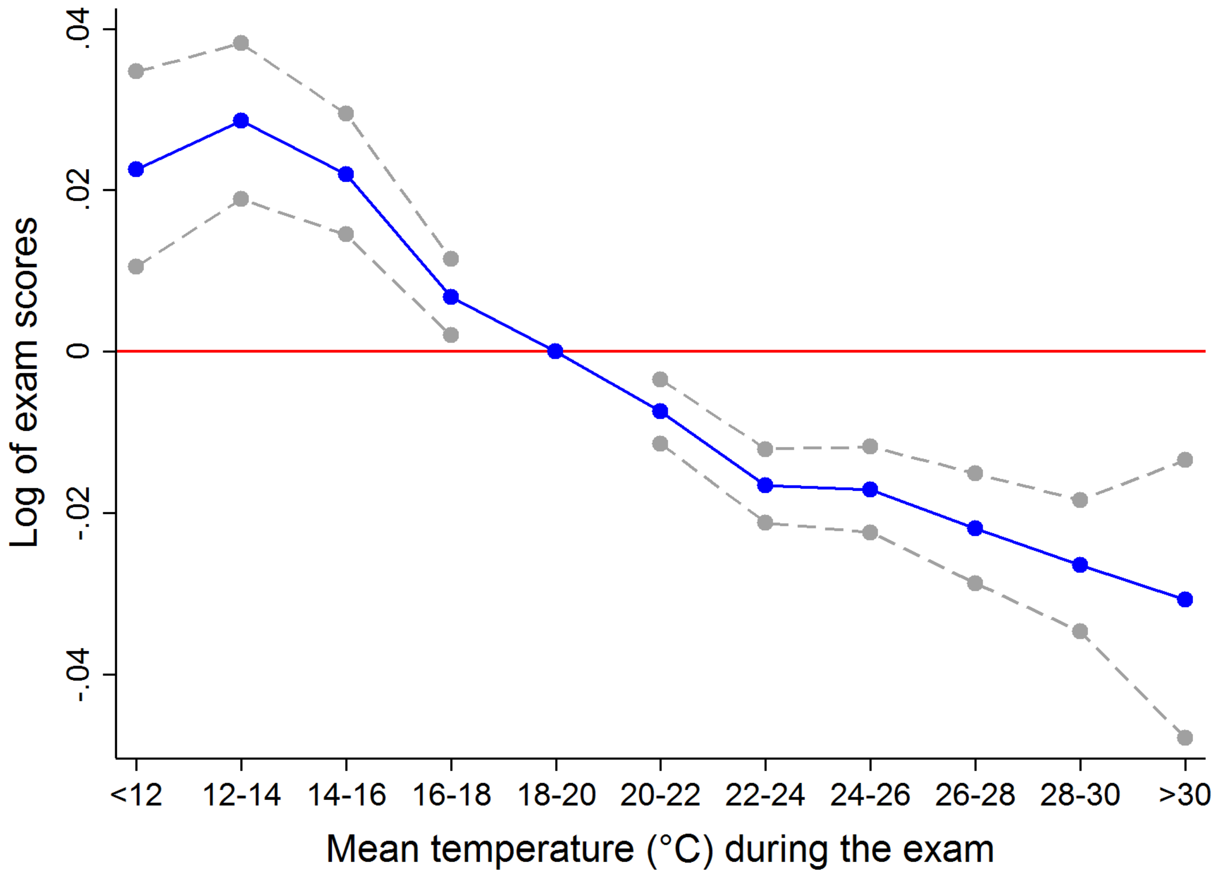

Figure 2: Relationship between temperature and log of exam scores

Notes: Each point estimate represents the effect of replacing a day with temperature in the 18-20 °C interval (reference group) with a day with temperature in the corresponding interval. Control variables include: precipitation, relative humidity, wind speed, sunshine duration, pressure, visibility, county fixed effects, and year fixed effects. Whiskers denote the 95% confidence interval, after adjusting for spatial and serial correlation within each county.

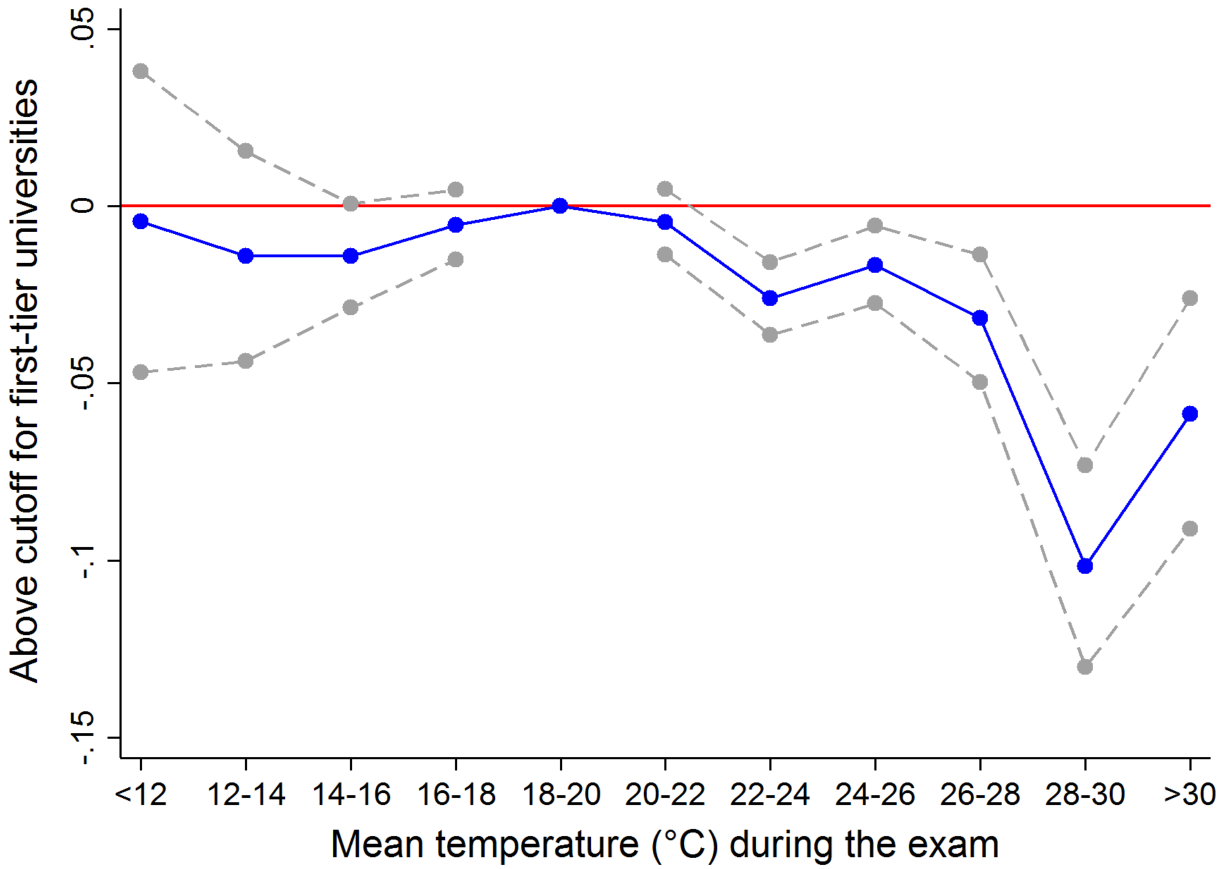

Figure 3: Relationship between temperature and probability of getting into first-tier universities

Notes: Each point estimate represents the effect of replacing a day with temperature in the 18-20 °C interval (reference group) with a day with temperature in the corresponding interval. Control variables include: precipitation, relative humidity, wind speed, sunshine duration, pressure, visibility, county fixed effects, and year fixed effects. Whiskers denote the 95% confidence interval, after adjusting for spatial and serial correlation within each county.

Table 1: Summary statistics

Variable	Mean	SD	Min	Max
Panel A: Score (0-750)				
Full sample	518.96	60.40	60.00	750.00
Art track	512.66	57.24	74.00	749.17
Science track	521.20	62.56	60.00	750.00
Panel B: Probability of above cutoff for first-tier universities				
Full sample	0.29	0.45	0.00	1.00
Art track	0.20	0.40	0.00	1.00
Science track	0.32	0.47	0.00	1.00
Panel C: Weather				
Temperature (°C)	23.21	3.29	2.55	31.96
DD>=20 (degree days)	3.59	2.64	0.00	12.00
DD<20 (degree days)	0.38	1.09	0.00	17.45
Precipitation (cm)	0.54	1.01	0.00	15.42
Relative humidity (%)	69.20	15.34	13.56	99.74
Wind speed (m/s)	2.30	0.86	0.26	16.22
Sunshine duration (hour)	5.77	3.73	0.00	14.17
Pressure (hpa)	965.33	53.67	581.45	1014.39
Visibility (km)	13.32	5.99	0.27	29.76

Notes: The NCEE data covers all students enrolled into college during 2005-2011. The observations for the full sample: 14,042,417. The observations for the art track: 3,699,915. The observations for the science track: 8,972,856. The sum of observations between the art track and the science track is not equal to the observations of the all track due to the existence of a small number of specialized tracks. The score scale is 0-750 for most provinces. We normalize the score scale to 750 for provinces that are not using the same scale. All students need to take three compulsory subjects: Chinese, mathematics, and a foreign language (typically English). The students in the art track need to take one combined subject comprising politics, history, and geography, and the students in the science track need to take one combined subject comprising physics, chemistry, and biology. The data on the cutoff of the first-tier universities are only available for the art and science tracks. The Tier 1 cut-off score is the minimum qualifying score for students to apply to Tier 1 universities. It is determined by the Provincial Admission Offices based on each year's admission quota and the distribution of student scores within the province and track. It varies annually across provinces and subject tracks. The weather variables are averaged using daily values on June 7th and 8th, when the NCEE is held.

Table 2: Statistical test on the difference of weather variables between provinces using independent exams and national exams

	Provinces using own exams (1)	Provinces using national exams (2)	Unconditional difference (3)	Conditional difference (4)
Temperature	23.7723	21.8715	1.9008***	-0.0002
	(2.8438)	(3.8719)	(0.0574)	(0.0293)
Precipitation	0.6865	0.3951	0.2914***	0.0001
	(1.1850)	(0.7776)	(0.0167)	(0.0138)
Humidity	73.9069	62.3013	11.6056***	-0.0011
	(12.9731)	(17.4421)	(0.2598)	(0.1505)
Wind	2.2474	2.2465	0.0009	-0.0001
	(0.9213)	(0.8385)	(0.0148)	(0.0087)
Sunshine	4.8838	6.6744	-1.7906***	-0.0003
	(3.6954)	(3.6573)	(0.0618)	(0.0423)
Pressure	980.7174	921.5846	59.1328***	0.0002
	(34.6706)	(72.4946)	(0.9668)	(0.0571)
Visibility	11.6256	17.9341	-6.3085	-0.0001
	(4.9484)	(6.5800)	(0.0984)	(0.0448)
Observations	6,811	7,367	---	---

Notes: Column (1) reports the average of county-year weather variables for provinces that use their own exams. Column (2) reports the average of county-year weather variables for provinces that use national exams. Column (3) reports the unconditional difference between columns (1) and (2). Column (4) reports the difference between columns (1) and (2) conditional on county fixed effects and year fixed effects. Standard errors are clustered at the county level and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Effect of temperature on log of exam score

	Dependent variable: Log of exam scores					
	All track		Art track		Science track	
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	-0.0034***	---	-0.0036***	---	-0.0018***	---
	(0.0004)	---	(0.0004)	---	(0.0004)	---
DD>=20	---	-0.0029***	---	-0.0031***	---	-0.0012***
	---	(0.0004)	---	(0.0004)	---	(0.0004)
DD<20	---	0.0061***	---	0.0061***	---	0.0046***
	---	(0.0006)	---	(0.0005)	---	(0.0007)
Precipitation	-0.0008	-0.0004	0.0002	0.0007	-0.0004	0.0002
	(0.0005)	(0.0005)	(0.0006)	(0.0006)	(0.0006)	(0.0007)
Humidity	0.0000	-0.0000	-0.0002**	-0.0003***	0.0003***	0.0002**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Wind	0.0039***	0.0036***	0.0008	0.0005	0.0024***	0.0020***
	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0007)	(0.0007)
Sunshine	0.0025***	0.0025***	0.0018***	0.0018***	0.0023***	0.0023***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Pressure	-0.0000	0.0000	0.0004***	0.0004***	-0.0006***	-0.0006***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Visibility	0.0001	0.0000	0.0004*	0.0004*	-0.0002	-0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	14,042,417	14,042,417	3,699,915	3,699,915	8,972,856	8,972,856
R-squared	0.2697	0.2699	0.4035	0.4037	0.2738	0.2740

Notes: The dependent variable is the log of the exam score. All students need to take three compulsory subjects: Chinese, mathematics, and foreign language (typically English). Students in the art track need to take one combined subject comprising politics, history, and geography, and students in the science track need to take one combined subject comprising physics, chemistry, and biology. The observations for all tracks does not equal the sum of observations from the art and science tracks, due to the existence of a small number of specialized tracks. Regression models also include county fixed effects and year fixed effects. Degree days (DD) ≥20 (<20) is the number of degrees above (below) 20 °C. Standard errors are clustered at the county level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Effect of temperature on the probability of above cutoff for first-tier universities

	Dependent variable: Above cutoff for first-tier universities					
	All track		Art track		Science track	
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	-0.0060***	---	-0.0083***	---	-0.0052***	---
	(0.0012)	---	(0.0015)	---	(0.0011)	---
DD>=20	---	-0.0075***	---	-0.0095***	---	-0.0068***
	---	(0.0013)	---	(0.0016)	---	(0.0012)
DD<20	---	-0.0021	---	0.0019	---	-0.0036**
	---	(0.0014)	---	(0.0015)	---	(0.0015)
Precipitation	-0.0185***	-0.0200***	-0.0220***	-0.0232***	-0.0159***	-0.0175***
	(0.0018)	(0.0019)	(0.0020)	(0.0022)	(0.0017)	(0.0019)
Humidity	-0.0010***	-0.0008***	-0.0017***	-0.0015***	-0.0007***	-0.0005**
	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0002)	(0.0002)
Wind	-0.0010	0.0001	0.0023	0.0031*	-0.0023	-0.0011
	(0.0016)	(0.0016)	(0.0019)	(0.0019)	(0.0016)	(0.0016)
Sunshine	0.0002	0.0002	0.0025***	0.0025***	-0.0005	-0.0005
	(0.0006)	(0.0006)	(0.0007)	(0.0007)	(0.0006)	(0.0006)
Pressure	-0.0017***	-0.0019***	-0.0022***	-0.0024***	-0.0013***	-0.0016***
	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Visibility	-0.0033***	-0.0032***	-0.0056***	-0.0055***	-0.0021***	-0.0021***
	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0007)	(0.0007)
Observations	12,672,771	12,672,771	3,699,915	3,699,915	8,972,856	8,972,856
R-squared	0.0550	0.0552	0.0666	0.0668	0.0568	0.0570

Notes: The dependent variable is a dummy variable, which equals to one if a student's score is above or equal to the cutoff of the first-tier universities and zero otherwise. All students need to take three compulsory subjects: Chinese, mathematics, and foreign language (typically English). Students in the art track need to take one combined subject comprising politics, history, and geography, and students in the science track need to take one combined subject comprising physics, chemistry, and biology. The data on the cutoff of the first-tier universities are only available for the art and science tracks. Regression models also include county fixed effects and year fixed effects. Degree days (DD) ≥20 (<20) is the number of degrees above (below) 20 °C. Standard errors are clustered at the county level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effect of weather and air pollution on log of exam scores

	Dependent variable: Log of exam scores		
	(1)	(2)	(3)
Temperature	-0.0034***	-0.0031**	-0.0032**
	(0.0004)	(0.0015)	(0.0016)
Precipitation	-0.0008	-0.0016	-0.0015
	(0.0005)	(0.0018)	(0.0018)
Humidity	0.0000	0.0000	0.0000
	(0.0001)	(0.0004)	(0.0004)
Wind	0.0039***	0.0034**	0.0035*
	(0.0007)	(0.0018)	(0.0018)
Sunshine	0.0025***	0.0024***	0.0025***
	(0.0003)	(0.0008)	(0.0009)
Pressure	-0.0000	0.0003***	0.0003***
	(0.0001)	(0.0001)	(0.0001)
Visibility	0.0001	-0.0003	-0.0002
	(0.0002)	(0.0005)	(0.0005)
API	---	---	0.0000
	---	---	(0.0001)
Observations	14,042,417	6,321,398	6,321,398

Notes: The dependent variable is the log of the exam score. All weather and air pollution variables are calculated using the average between June 7th and 8th. Column (1) reports the baseline estimates from Table 1, column (1). Column (2) reports results from the same specification but only for the sample of cities covered by the air pollution index (API). In column (3) we add controls for pollution as measured by the API. The regression models also include county fixed effects and year fixed effects. Standard errors are clustered at the county level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Robustness checks

Panel A	Dependent variable: Log of exam scores					
	Baseline	Clustering prefecture	Max Temp	Level of score	Three days	County-year
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	-0.0034***	-0.0034***	-0.0026***	-1.7453***	-0.0022***	-0.0026***
	(0.0004)	(0.0009)	(0.0003)	(0.2062)	(0.0004)	(0.0003)
Observations	14,042,417	14,042,417	14,042,417	14,042,417	14,042,417	14,042,417
Panel B	Dependent variable: Above cutoff for first-tier universities					
	Baseline	Clustering prefecture	Max Temp	Logit	Three days	County-year
Temperature	-0.0060***	-0.0060**	-0.0044***	-0.0289***	-0.0029***	-0.0078***
	(0.0012)	(0.0026)	(0.0008)	(0.0058)	(0.0010)	(0.0009)
Observations	12,672,771	12,672,771	12,672,771	12,672,771	12,672,771	14,177

In panel A, the dependent variable is the log of exam score except for column (4), where the dependent variable is the level of score. In panel B, the dependent variable is a dummy variable which equals to one if the student's score is above or equal to the cutoff of the first-tier universities and zero otherwise. Column (1) is the baseline model. In column (2), we cluster standard errors by prefecture, to control for serial and spatial correlation within prefecture. Noted that prefecture is an administrative unit between province and county. In column (3), we use average of daily maximum temperature on June 7th and 8th. In column (4) of panel A, we use the level of score as the dependent variable. In column (4) of panel B, we use the logit model and reported the marginal effects evaluated at the mean level. In column (5), we include temperature on June 9th for provinces with exams held on June 7th-9th. In column (6), we collapse observations by county-year, and estimate the model using count-year observations. Standard errors are clustered at the county level except in column (2) and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

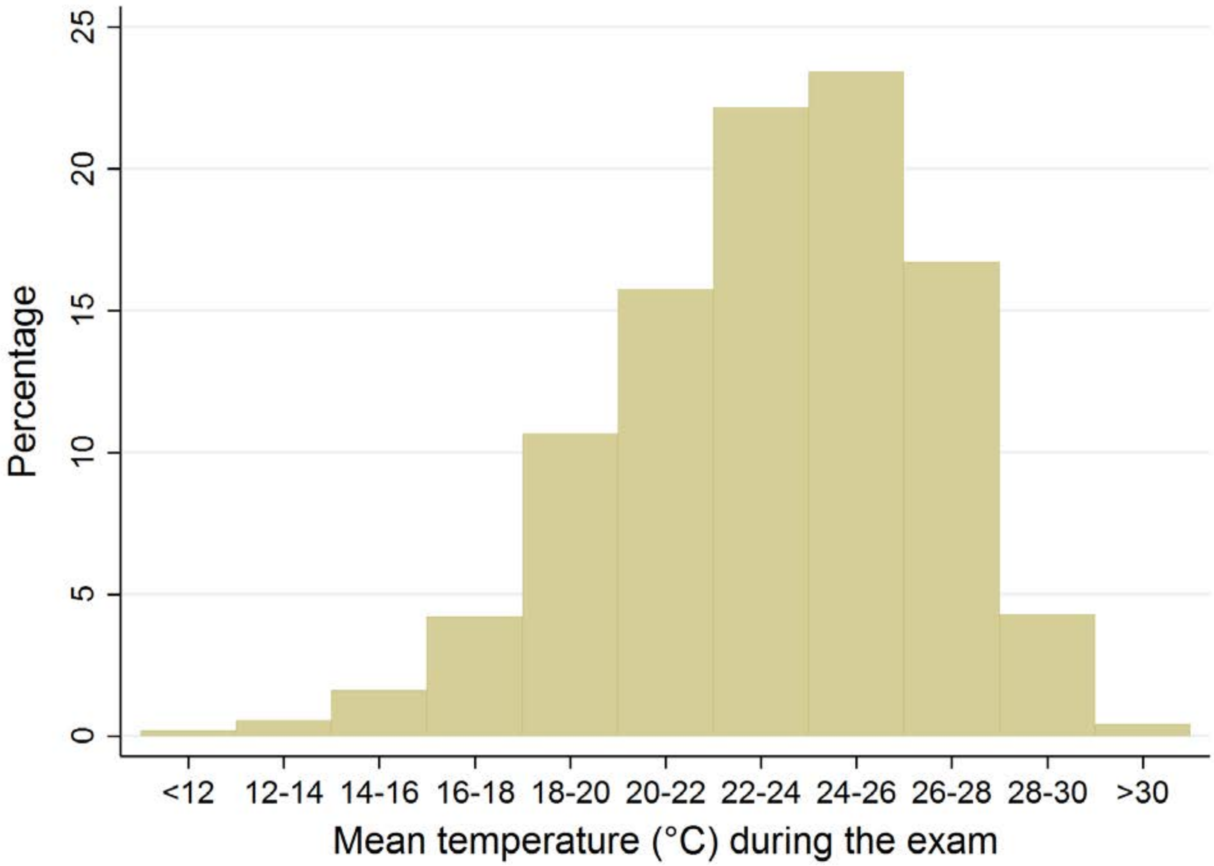

Figure A1. Distribution of 2 ℃ indicators.

Notes: The height of each bar denotes the percentage of each 2 ℃ indicator.

Table A1: Effect of weather on log of exam scores across urban and rural areas

	Dependent variable: Log of exam scores					
	Full		Urban		Rural	
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	-0.0034***	---	-0.0039***	---	-0.0026***	---
	(0.0004)	---	(0.0005)	---	(0.0005)	---
DD>=20	---	-0.0029***	---	-0.0031***	---	-0.0026***
	---	(0.0004)	---	(0.0006)	---	(0.0005)
DD<20	---	0.0061***	---	0.0079***	---	0.0026***
	---	(0.0006)	---	(0.0008)	---	(0.0008)
Precipitation	-0.0008	-0.0004	-0.0018***	-0.0013*	0.0015**	0.0015*
	(0.0005)	(0.0005)	(0.0006)	(0.0007)	(0.0008)	(0.0008)
Humidity	0.0000	-0.0000	-0.0002	-0.0002*	0.0003***	0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Wind	0.0039***	0.0036***	0.0048***	0.0043***	0.0030***	0.0030***
	(0.0007)	(0.0007)	(0.0010)	(0.0010)	(0.0009)	(0.0009)
Sunshine	0.0025***	0.0025***	0.0026***	0.0026***	0.0021***	0.0021***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Pressure	-0.0000	0.0000	-0.0002	-0.0001	0.0003*	0.0003*
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Visibility	0.0001	0.0000	-0.0003	-0.0003	0.0010***	0.0010***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	14,042,417	14,042,417	9,418,385	9,418,385	4,624,032	4,624,032

Notes: The dependent variable is the log of the exam score. All weather variables are calculated using the average between June 7th and 8th. Columns (1) - (2) report the estimates for the full sample. Columns (3) - (4) report the estimates for urban districts and columns (5) - (6) report the estimates for rural counties. Degree days (DD) ≥20 (<20) is the number of degrees above (below) 20 °C. Regression models also include county fixed effects and year fixed effects. Standard errors are clustered at the county level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Correlation coefficients between $PM_{2.5}$ and temperature

	$PM_{2.5}$					
	Raw Correlation			Correlation conditional on fixed effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	0.1589	---	---	0.1817	---	---
DD>=20	---	0.1217	---	---	0.1989	---
DD<20	---	-0.1855	---	---	-0.0190	---
<12 ℃	---	---	-0.0197	---	---	-0.0015
12-14 ℃	---	---	-0.0404	---	---	0.0032
14-16 ℃	---	---	-0.0859	---	---	0.0070
16-18 ℃	---	---	-0.1110	---	---	-0.0115
18-20 ℃	---	---	-0.1478	---	---	-0.0312
20-22 ℃	---	---	-0.0559	---	---	-0.0957
22-24 ℃	---	---	0.1158	---	---	-0.0143
24-26 ℃	---	---	0.1206	---	---	-0.0111
26-28 ℃	---	---	0.1039	---	---	0.1585
28-30 ℃	---	---	-0.1409	---	---	-0.0098
>30 ℃	---	---	-0.0539	---	---	-0.0010

Notes: This table reports the correlation coefficients between $PM_{2.5}$ and temperature. All variables are calculated using the average between June 7th and 8th during the period 2013-2016. Columns (1)-(3) report the raw correlation coefficients and columns (4)-(6) report the correlation coefficients conditional on county fixed effects and year fixed effects. The variable "Temperature" is the average temperature between June 7th and 8th during the period 2013-2016. The variable "Degree days (DD) ≥20 (<20)" is the number of degrees above (below) 20 ℃ constructed using the variable "Temperature". The non-parametric approach that ranges from <12 ℃ to >30 °C, with 2 ℃ bins in between is constructed using the variable "Temperature".

Table A3: Raw correlation coefficients between $PM_{2.5}$ and temperature indicators

	$PM_{2.5}$ ($\mu g/m^3$)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<20	20-40	40-60	60-80	80-100	100-120	120-140	>140
<12 °C	0.0063	0.0229	-0.0110	-0.0133	-0.0090	-0.0060	-0.0037	-0.0058
12-14 °C	0.0318	0.0233	-0.0167	-0.0243	-0.0165	-0.0110	-0.0068	-0.0107
14-16 °C	0.1230	0.0040	-0.0476	-0.0409	-0.0326	-0.0146	-0.0135	-0.0138
16-18 °C	0.1060	0.0354	-0.0427	-0.0509	-0.0528	-0.0173	-0.0147	-0.0296
18-20 °C	0.1242	0.0848	-0.0744	-0.0921	-0.0695	-0.0242	0.0146	-0.0406
20-22 °C	-0.0381	0.1337	-0.0498	-0.0475	-0.0341	-0.0526	0.0120	0.0081
22-24 °C	-0.1626	-0.0173	0.0615	0.0633	0.0979	0.0114	-0.0218	-0.0077
24-26 °C	-0.1408	-0.0892	0.1460	0.0899	-0.0254	0.0443	-0.0171	0.0240
26-28 °C	-0.0093	-0.0684	-0.0115	0.0222	0.0793	0.0163	0.0472	0.0532
28-30 °C	0.2677	-0.0676	-0.0924	-0.0303	-0.0497	0.0230	-0.0241	-0.0377
>30 °C	0.1119	-0.0369	-0.0273	-0.0177	-0.0120	-0.0080	-0.0050	-0.0078

Notes: This table reports the raw correlation coefficients between $PM_{2.5}$ and temperature indicators. All variables are calculated using the average between June 7th and 8th during the period 2013-2016.

Table A4: Correlation coefficients between $PM_{2.5}$ and temperature indicators conditional on county fixed effects and year fixed effects

	$PM_{2.5}$ (μg/m^3)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<20	20-40	40-60	60-80	80-100	100-120	120-140	>140
<12 °C	0.0015	0.0041	-0.0091	0.0004	0.0061	0.001	0.0001	-0.0013
12-14 °C	-0.0075	0.0071	-0.0037	-0.0004	0.0069	-0.006	-0.0057	0.0044
14-16 °C	0.0294	-0.0058	-0.0285	-0.0007	0.0134	-0.0018	0.0063	0.0153
16-18 °C	0.0635	-0.0211	-0.0361	0.0167	-0.0052	0.004	0.001	0.0069
18-20 °C	0.0579	0.0195	-0.0473	-0.0011	-0.0439	0.0051	0.0358	0.002
20-22 °C	0.0072	0.109	-0.0136	-0.0654	-0.0602	-0.0612	-0.0115	0.0033
22-24 °C	-0.0562	0.0279	0.025	-0.0309	0.0637	0.0231	-0.0579	-0.0631
24-26 °C	-0.0093	-0.0723	0.0698	0.1029	-0.0662	-0.0066	-0.0496	-0.0194
26-28 °C	-0.0533	-0.0466	-0.0015	-0.0309	0.0767	0.0137	0.1096	0.0899
28-30 °C	0.0546	-0.0308	-0.0428	0.0273	0.0147	0.0427	-0.0085	-0.0158
>30 °C	0.0431	-0.0407	0.0021	0.0017	0.0131	0.0034	0.0007	-0.0006

Notes: This table reports the correlation coefficients between $PM_{2.5}$ and temperature indicators conditional on county fixed effects and year fixed effects. All variables are calculated using the average between June 7th and 8th during the period 2013-2016.