ASIAN ECONOMIC JOURNAL
Journal of the East Asian Economic Association



# Effects of Heat on Mathematics Test Performance in Vietnam

Tien Manh Vu

Received 7 February 2021; Revised 31 July 2021; Accepted 19 December 2021

We examine the effects of average test-day temperature on the mathematics test performance of all Vietnamese students who took the national university and college entrance examinations in 2009. Using individual fixed effects, we find that an increase of 1°F results in an approximate 0.006 standard deviation loss of in standardized test scores by age and test problem. The negative effects are persistent regardless of whether students were from the hottest or coolest climate regions in the country. We also find that female students and students from rural areas and townships are most vulnerable to the effects of heat.

*Keywords:* heat, test performance, test score, Vietnam.

*JEL classification codes:* I29, J24, J16, O15.

*doi:* 10.1111/asej.12259

## I. Introduction

Climate and cognitive skills are important life factors that humans have not been able to alter immediately. Both are objects of great concern. Various studies regarding their interactions have examined how climate change, particularly heat waves, might influence cognitive skills and corresponding task performance (Garg *et al*., 2017; Graff Zivin *et al*., 2018; Park *et al*., 2018; Park, 2020; Park *et al*., 2020). Specifically, Horowitz (2009) suggested that a 1°C (1.8°F) increase in global temperature would reduce global income by approximately 3.8 percent. Examining approximately 10 million test-takers in the USA who repeated the PSAT, Park *et al*. (2020) showed that a 1°F (0.56°C) hotter school year leading up to the tests decreased learning outcomes by approximately 1 percent. Meanwhile, Park *et al*. (2020) demonstrated that hotter temperatures might have lowered the test performance of New York City high school students during 1999–2011 by up to 11 percent of the standard deviation (SD).

This line of research is crucial for developing countries. Poor countries are more likely to be in hot regions (Horowitz, 2009; Acemoglu and Dell, 2010). In addition, poor households are more likely to be involved in agriculture, to be

*Vu (corresponding author): Miyazaki International College, Miyazaki, Japan. Email: mvu@sky.miyazaki-mic.ac.jp.

exposed more to heat (Park *et al.*, 2018) and to suffer more from extreme weather because of residential 'sorting problems' (Graff Zivin and Neidell, 2013). Poor households have fewer means of prevention and protection (e.g. housing with cooling systems and favorable working conditions; i.e., with air-conditioning), fewer safety buffers (private savings) and fewer means of loss reduction (insurance) to combat the effects of climate change.

The link between cognitive skills and climate in developing countries is still being investigated. The main obstacles are inappropriate priorities and a lack of reliable data on cognitive skills that can be geocoded. Among the few studies on this topic, Park *et al.* (2018) suggested a sorting problem; that is, a negative relationship between hotter locations and lower household income within some developing countries with hot climates. Using repeated cross-sectional data from India for 2006–2014, Garg *et al.* (2017) showed a similar negative relationship between temperature in the calendar year prior to a standardized test and the test scores. Specifically, an additional 10 days with an average temperature above 29°C (85°F) relative to 15°C–17°C (59°F–63°F) decreased test scores by approximately 0.02–0.03 SD. However, without individual fixed effects, the estimations suffer from bias because past weather can be correlated with weather information used in the estimations. Horowitz (2009) showed a negative relationship between long-run average temperature and logarithm per capita GDP. In Horowitz's analyses, Vietnam was listed in the group with a relatively higher temperature but a relatively lower income. We chose Vietnam for our analyses for that reason and because approximately 21.26 out of Vietnam's 85.84 million people work in the agricultural sector (GSO, 2006).

We investigated the effects of test day average temperature on test-taking performance using the 2009 Vietnamese national university and college[1] test score data. The data contained 1 125 953 math test scores (779 123 test-takers). Using individual fixed effects, we examined 294 623 test-takers who took two or three separate mathematics examinations and created an unbalanced panel dataset (641 453 math test scores). We also investigated subsamples, including the hottest regions and coolest regions, heterogenous numbers of mathematics tests taken, gender, regional economic disparities and policy groups. We reported the results using standardized scores (by test problem and year of birth) but also checked the results against the zero-value-adjusted logarithm of the raw scores.

We found that a 1°F increase in the average temperature on the test day resulted in an approximately 0.006 SD decrease in the standardized test score. Specifically, the decrease was approximately 1.64 percent of the raw mathematics score of a typical student in the largest group of test-takers (born in 1991) if the temperature changed from 77.9°F (25.5°C)—the average temperature between 1950 and 2009—to 82.8°F (28.2°C). This temperature (82.8°F) was 75 percent of the highest recorded average daily temperature in Vietnam in the

[1] The exams were centrally controlled and had uniform test problems. Test-takers had up to three times to sit in different tests and needed to earn top scores each time in order to earn admittance.

same period. The results suggested that the Vietnamese economy may incur a similar loss in productivity due to global warming. We also found that female students and students from rural areas and townships were most vulnerable to these effects of heat. Our estimated results might serve as a reference point for a carbon tax policy to be implemented in Vietnam from 2022 as well as its ameliorating redistribution (especially for vulnerable groups). In addition, our results suggest an important implication for school infrastructure investment. The impact on testing performance due to heat can be offset with reasonable heat-preventing solutions based on cost–benefit analyses.

Our study contributes several important insights into the existing research. First, to our knowledge, this is one of only a few studies of developing countries. Our study confirms a negative effect of heat on cognitive performance based on national census test scores. Specifically, the study aligns with Graff Zivin *et al*. (2018), Park *et al*. (2020) and Park *et al*. (2018). However, it contributes new information because activated air-conditioning was likely nonexistent at test sites in Vietnam in 2009. Unlike Garg *et al*. (2017), we used the exact weather on-site on test days for each student who participated in up to three separate mathematics examinations. Second, our results validated the effects of heat among people living in a tropical climate. The results undermined the myth that people in hot climate regions are immune from heat effects. Third, the method and the specific settings of the examinations helped to improve the estimations. Our individual fixed effects helped to identify the effect of test day heat and to separate the effect of test-day heat from the effect of past weather and accumulated cognitive skills. In particular, examination settings with a short time interval between tests (fewer than 12 days) reduced the possibility of neglecting the effect of weather (heat) during the interval on learning outcomes. Previous studies using panel data with individual fixed effects often examined test scores with longer intervals between tests. Students retook the test subject, perhaps when they had already invested more time in their studies, aiming to get a higher score. The weather (heat) during the longer time interval might also have influenced the outcomes. Fourth, we dealt with the sorting problem (Graff Zivin and Neidell, 2013), which can create a spurious correlation between temperature and test scores, especially if individual fixed effects are not in place. For example, families with more wealth and educational attainment might flock to areas with a cooler climate and also invest more in their children's education. Therefore, their children are more likely to have better test scores. As a result, the correlation between test scores and the temperature inside the test venue (usually located in their city) is spurious. In addition, by using data for individuals above 17 years of age, our study expands the timeline research for a significant portion of literature, including Garg *et al*. (2017), Hoddinott and Kinsey (2001) and Thai and Falaris (2014), who used test scores and educational outcomes at primary and secondary schools.

The remainder of this paper is organized as follows. Section 2 summarizes the key settings of the national entrance examinations to universities and colleges in 2009, and Section 3 describes the data in detail. Section 4 follows with econometric methods and specifications. Section 5 presents the results, and Section 6 finishes with the discussion and conclusions.

## II. The Vietnamese National Entrance Examinations to Universities and Colleges in 2009

The Vietnamese National Entrance Examinations to Universities and Colleges in 2009 helped to construct comparable test scores and identify the effect of heat on test performance. The examinations were centrally organized by the Vietnamese Ministry of Education and Training (MOET) with uniform test problems. The MOET decided the test dates, major test sites and regulations involved in related issues (including the number of available placements per institute detailed at major/faculty level) and published all information in a guidebook for the examinations in February 2009.

Using the guidebook, students registered three desired institute names per application with MOET in descending order of preference. Test-takers could not change the names of the institutes once they had registered (MOET, 2008). In general, test-takers took the examinations in the location of the first institute in their three desired institutes. MOET termed each combination of three compulsory test subjects a "classification" and assigned an alphabet letter (or a code) to each classification.[2] There were 11 test classifications in total, with each classification having three test subjects. The six major classifications (A, B, D, M, T and V) included mathematics as a test subject. For example, A classification consisted of mathematics, physics and chemistry. B classification had only one test subject different from A classification (biology instead of physics). The three desired institutes must share the same test classification. From the guidebook, test-takers took the test classification corresponding with the chosen faculty/major of the three desired institutes and filled in the information in their test application.

Test-takers could send MOET up to two applications for universities and one for colleges, each with three predetermined test subjects because of the differences in examination timing. The examinations for entrance to university were on July 4 and 5 for A and V classifications and July 9 and 10 for B, D, M and T classifications. Meanwhile, the entrance examinations to colleges were on July 15 and 16 for all classifications. Test-takers could neither add new applications nor replace their existing applications after 17 April.

Both universities and colleges used a sum of three test subject scores per classification for student placement to the major (faculty). The universities and colleges might use several classifications for placements in each of their faculties (majors). However, a test score in one classification cannot be used for a

[2] The MOET abolished the centralized entrance examinations with the test classifications in 2014.

different classification. For example, students still had to take the two mathematics tests for A and B classifications separately, even if the institute used both A classification and B classification for placement. Based on the number of available slots declared in the guidebook of MOET (for each faculty/major per classification), each institute ranked the corresponding test scores from highest to lowest and cut the list where the number of available slots per classification for each faculty/major fit.

In general, the universities and colleges decided the test locations. They made use of their own or the facilities of the nearby schools and staff for testing and monitoring to reduce cost. In addition, MOET also gave test-takers the option to take the test at one of three designated test sites instead of traveling across regions (i.e., the northern, southern and central regions of Vietnam) to reach their desired universities.

At the test sites of respective universities, students were assigned to test rooms according to both targeted university (college) faculty and alphabetical order. However, MOET prohibited placing test-takers into test rooms according to residence origin (MOET, 2008). Test-takers in the same test room experienced the same ambient environment. MOET decided what test-takers could bring into the test room and limited options to pens, pencils, compasses, rulers and calculators (MOET, 2008). MOET required test monitors to randomly assign test seats prior to the arrival of test-takers at test sites and to reallocate seating between test subjects.

Grading was strictly anonymous and took place after the last test (July 16, 2009). Graders worked in positive environments (sometimes with air conditioners) to reduce the influence of summer heat. Therefore, the quality of grading is expected to be homogenous and independent from the surrounding environment, including the weather at the grading site.

Because test-takers did not know the cut-off point for placements, they had to put their best efforts into solving the test problems on the day of the test. Moreover, the competition was fierce because even the “easiest” institutions selected only one out of every two test-takers. If test-takers failed, it will be costly to wait for another year to retake the exam again. Therefore, the test score distribution of the entrance examinations did not have a huddle at the passing score. Other experiments to measure cognitive skills were not always able to provide enough incentives for individuals to exert their best efforts to solve the test problems, especially for several lengthy tests taken at different times.

## III. Data

We used test score data from the 2009 Vietnamese national college and university entrance examinations and linked the data with the Daily (Weather) Summaries. The test score data are a census of all math test-takers who took two entrance exams: (a) to university from 4–5 July and 9–10 July 2009 (984, 477 individual test scores) and (b) to college (277 988 individual scores) on 15–16 July 2009.

The raw data were available for analysis from the website of the Higher Education Department, Ministry of Education and Training of Vietnam in 2009.

We examined mathematics test scores for several important reasons. First, mathematics test scores are the best measurement of human capital investment in Vietnam. Mathematics is compulsory for all grades (from 1 to 12) in Vietnamese general education. In contrast, the other test subjects were introduced in schools starting in grade 7 at the earliest. Second, mathematics test scores enabled us to implement individual fixed effects in the panel data with the largest number of observations and the largest variations in test sites and times. Mathematics test scores accounted for nearly one-third of all test subjects' scores in the NEEU data. Literature subject is also compulsory in Vietnamese general education but does not appear in as many test classifications and examination days as mathematics does. Third, the use of mathematics test scores presented the fewest issues, including sample selection, compared with all other test subjects used in the NEEU. The mathematics test format was uniform across test times. Mathematics tests uniformly lasted 180 min, taking place from 7:15 to 10:15 am on July 4, 2009 and from 2:15 to 5:15 pm on July 9 and 15, 2009. A selection issue might occur when students choose to take two or three examinations, including mathematics as a test subject. Specifically, 484 500, 242 416 and 52 207 students took one, two and three different mathematics tests, respectively. To examine whether this selection might alter our results, we performed heterogeneity checks among the subsamples based on the number of mathematics tests taken). In contrast, although other subjects received as much research attention as did mathematics, they presented various issues, such as more severe sample selection issues. The majority of the other test subjects took place once and at fewer test locations in Vietnam. Some were presented in different test formats, including multiple-choice questions or tests with less time allotted (e.g. only 90 min). Therefore, the other test subjects were prone to produce comparable outcomes (e.g. due to different durations of heat exposure).

After omitting individuals with missing and unidentified information necessary for analysis, we obtained 1 125 952 scores from 779 123 students. The data consist of raw test scores (maximum point = 10),[3] year of high school graduation and relative resident location (province, district and area types as defined by the provincial departments of education, namely urban city, township, rural and "economic difficulty area"). We were able to identify test-takers who took one, two and three mathematics tests to university and college (hereafter, university).

Another dataset—the Daily (Weather) Summaries (available from the US National Oceanic and Atmospheric Administration)—contains information on the daily weather in Vietnam since 1950. The data include weather information measured at specific weather stations. These data specify weather indicators, including daily temperature and daily precipitation for 353 257 days from

[3] In the data, we rescale to 1000 as the maximum point of the raw score.

January 1, 1950, to December 31, 2009, from every meteorological station in service (44 existing stations during the years 1950–2009) in Vietnam.[4]

We used ground-level weather station data, which we preferred to gridded weather data for reasons that Auffhammer *et al.* (2013) indicated. First, gridded weather data produce similar average temperatures in each cell. Second, gridded data often have significant spatial correlation, depending on the data inference method. Third, gridded data are often based on weather station data, and missing information in the weather station data leads to artificial variation in the gridded data. Furthermore, the highest resolution for gridded data is approximately 0.25° × 0.25° (approximately 27–28 km$^2$ block); however, to the best of our knowledge, such high resolution is not readily available in Vietnam. Meanwhile, we found that the average distance between test sites and weather stations was 20–24 km (12.4–14.9 miles) in our data. Later, we also checked the robustness of the results by further limiting the data to a distance equal to or less than 8 km (approximately 5 miles) between weather station and test site.

Using the test score data and the MOET exams guidebook, we geocoded the corresponding latitude and longitude of school locations from their full addresses. We merged the corresponding information from the Daily (Weather) Summaries with the test score data by choosing the nearest weather station to the testing site.[5] We used the exact weather on each test day.

We standardized the raw math score to form comparable outcomes ($Z-score_{it}$). The *Z*-score is based on the mean ($Mean_{a,p}$) and SD ($SD_{a,p}$) from the origin census data for each combination of year of birth ($a$) and test problem ($p$); as such, $Z-score_{it} = \left(raw\,score_{it} - Mean_{a,p}\right)/SD_{a,p}$. $Mean_{a,p}$ and $SD_{a,p}$ were at the country level but were specified for a given year of birth and test problem.[6] We reported estimations using $Z-scores$ as our main results. In addition, we used logarithm-adjusted raw scores as we did for the robustness checks ($ln(score) = ln(raw\,score + 1)$).

We used test-day average temperature as the main independent variable (*Avg. Temp*). Ideally, the temperature should be measured at the test site when students are taking the test. Unfortunately, the only test-day temperature data available were the maximum, minimum and average temperatures. Accordingly, the

4 During the history of development, new weather stations were built and some were no longer used in Vietnam. Therefore, the average number of weather stations increased from 5 stations in 1950 to 30 stations in 2009.

5 Some weather stations were located on islands and did not correspond to any test or school locations.

6 We also tested at country level if the mean of temperature among weather stations varied significantly among the 3 days (July 3, July 8 and July 14, 2009) having the mathematics tests. The *p*-values of the tests were 0.785 (between July 3 and July 8, 2009), 0.357 (between July 8 and July 14, 2009) and 0.334 (between July 3 and July 14, 2009). The test results suggested that at country level, the mean of raw test scores at country level ($Mean_{a,p}$) by year of birth and test problem were less likely affected by neither the country average temperature of the test days nor the temperature at the specific test sites.

average temperature turned out to be the best choice among the available data. We acknowledge this as a data limitation. Table 1 shows the detailed summary statistics. Further distribution of test-takers into weather stations is presented in Appendix B.

We acknowledged the drawbacks of being unable to account for a humidity indicator. Higher humidity levels can magnify the effects of heat on test performance. Unfortunately, such information was not available in the Daily Weather Summaries for Vietnam. Therefore, we had to make a strong assumption that humidity levels were homogenous across test sites during July 2009.

## IV. Econometric Methods and Specifications

We applied individual fixed effects ($\mu_i$) to identify the effect of heat, proxied by test day average temperature ($Avg.Temp_{it}$), to the standardized test score ($Z-score_{it}$). We estimated the following equation:

$$Z-score_{it} = \rho_1 Avg.Temp_{it} + \rho_{2t}\mathrm{C}_{it} + \rho_3\mu_i + \mathrm{t} + \epsilon_{it}. \tag{1}$$

The time and date of the exams were used as a set of time dummies, $t$. $\mathrm{C}_{it}$ is a vector of time-variant control variables, including test classification, university-classification rank, university dummy and distance between the high school and the test site. University-classification rank is a set of categorical variables for each interval between 0, 1, 5, 10, 25, 50, 75, 90, 95, 99 and 100 percent in the ranking (university) list by test classification. For each university, we calculated the corresponding university-classification mean raw score to determine its order in the rank list. Each university dummy also captured major differences in infrastructure (e.g. buildings with certain cooling conditions) per university-classification rank. However, university dummy and university classification do not overlap because a university could use several classifications for the placement.

Additionally, the university-classification rank helps to reduce the possibility of a test site sorting problem (Graff Zivin and Neidell, 2013). For example, higher-ranked universities might be in plains areas, whereas lower-ranked universities might be in the highlands. During summertime, good universities would have higher test site temperatures (see left graph of Figure 1), whereas the others might enjoy lower temperatures. Meanwhile, good students generally chose good universities to take the test at. As a result, the correlation between test site temperature and test score will become positive, as shown in the linear trend line in the left graph of Figure 1. The correlation is spurious. Given that several universities for different university-classification ranks could be located within scope of a weather station (having the same average temperature at the test sites), the combination of university-classification rank and university

**Table 1 Summary statistics of selected variables**

| *Variable* | *All* | *Gender* | | *Take* | | |
|---|---|---|---|---|---|---|
| | | *Male* | *Female* | *One* | *Two* | *Three* |
| | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* |
| *Z*-score | 0.099 | 0.134 | 0.071 | –0.025 | 0.171 | 0.261 |
| Gender | 0.551 | 1.000 | 1.000 | 0.525 | 0.555 | 0.620 |
| High school grad. Year | 2008.735 | 2008.65 | 2008.805 | 2008.633 | 2008.801 | 2008.847 |
| *Exams choice* | | | | | | |
| Distance to test site (km) | 118.991 | 121.334 | 117.081 | 122.154 | 117.735 | 113.097 |
| July 9, 2009 (P.M) | 0.369 | 0.317 | 0.412 | 0.393 | 0.357 | 0.333 |
| July 15, 2009 (P.M) | 0.215 | 0.204 | 0.224 | 0.173 | 0.219 | 0.333 |
| Take 2 | 0.431 | 0.427 | 0.434 | 0 | 1 | 0 |
| Take 3 | 0.139 | 0.118 | 0.156 | 0 | 0 | 1 |
| *Test day weather* | | | | | | |
| Avg. Temp (°F) | 83.449 | 83.289 | 83.579 | 83.555 | 83.459 | 83.091 |
| *N (scores)* | 1 125 953 | 505 676 | 620 277 | 484 500 | 484 832 | 156 621 |
| ID (students) | 779 123 | 358 000 | 421 123 | 484 500 | 242 416 | 52 207 |
| | Regional disparity* | | | | Policy group | |
| | Most difficulty | Rural | Townships | Urban city | Ethnic minority | State honored |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| *Z*-score | –0.124 | 0.126 | 0.264 | 0.385 | –0.311 | 0.147 |
| Gender | 0.547 | 0.549 | 0.567 | 0.545 | 0.554 | 0.524 |
| High school grad. Year | 2008.712 | 2008.72 | 2008.756 | 2008.821 | 2008.665 | 2008.554 |
| *Exams choice* | | | | | | |
| Distance to test site (km) | 139.386 | 138.090 | 87.409 | 44.991 | 143.591 | 149.199 |
| July 9, 2009 (P.M) | 0.390 | 0.347 | 0.374 | 0.380 | 0.444 | 0.368 |
| July 15, 2009 (P.M) | 0.207 | 0.233 | 0.198 | 0.204 | 0.198 | 0.199 |
| Take 2 | 0.408 | 0.442 | 0.440 | 0.442 | 0.324 | 0.415 |

(*Continues*)

**Table 1 (continued)**

| *Variable* | *All* | *Gender* | | *Take* | | |
|---|---|---|---|---|---|---|
| | | *Male* | *Female* | *One* | *Two* | *Three* |
| | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* |
| Take 3 | 0.121 | 0.147 | 0.133 | 0.173 | 0.040 | 0.116 |
| *Test day weather* | | | | | | |
| Avg. Temp (°F) | 83.117 | 83.875 | 82.886 | 83.672 | 83.920 | 83.449 |
| *N (scores)* | 363 890 | 434 146 | 201 627 | 126 290 | 48 686 | 43 915 |
| ID (students) | 260 146 | 295 676 | 139 456 | 83 845 | 39 541 | 31 509 |

*Note:* * Each province's department of education self-defined each commune into the four categories in each year in accordance with national regulations and approval from MOET. Gender = 1 if female, 0 if otherwise.

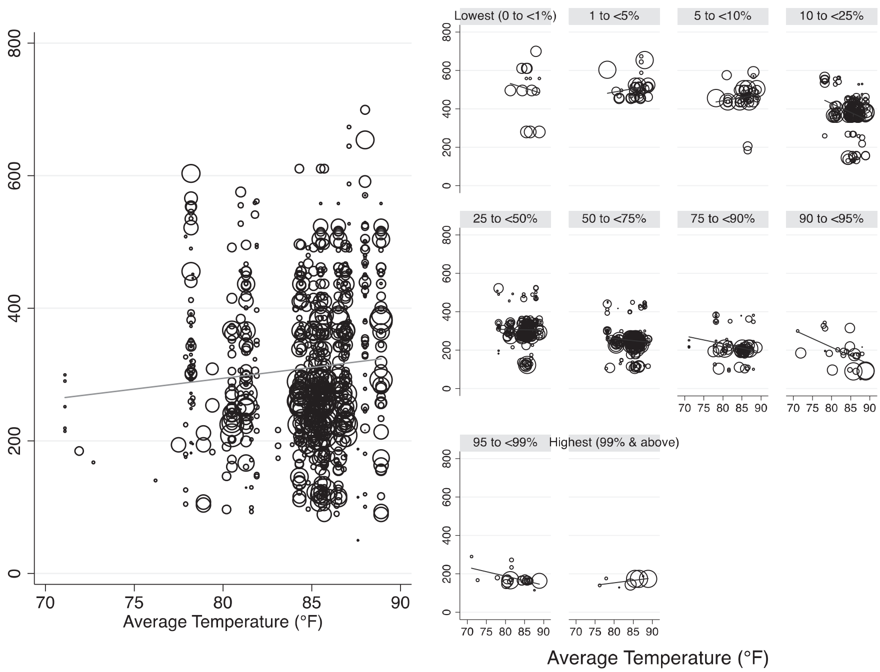
dummies would dissolve the spurious correlation (see the right graph of Figure 1 for the possibility).

The distances to the test venues and university dummies did not show perfect multicollinearity when we used individual fixed effects. The institutions of the entrance examinations in 2009 allowed some students to take the test in their region. The total number of regions is three. The students did not have to travel across regions to the registered university.

Although the individual could take a mathematics test for one classification at one $t$, the classification variable still varied within individual. A test-taker could take different classifications for different $t$. For example, the test-taker could take A classification (mathematics, physics and chemistry) on July 4 and 5 and B classification (mathematics, biology and chemistry) on July 9 and 10.

Our panel data are unbalanced. Students took mathematics tests on two or three different days (July 4, 9 and 15). We named the number of tests taken by a student "take." However, the "take" was predetermined. The test-takers had to decide the test classification in their applications several months before the tests took place. They could withdraw from taking the registered examinations but could not send new applications to MOET.

**FIGURE 1 Distribution of student raw score, university mean mathematics raw score by university rank and average test day temperature. A circle radius represents the number of test-takers to a university. The center point of the circle indicates the mean mathematics raw score of each university classification**



We acknowledged that the number of "takes" contained a selection problem and had several corresponding arguments. First, students had to decide on test classification in advance, as early as 10th grade, for test preparation and then made their final decisions on the test applications more than 2 months prior to the examinations. Each "take" was associated with only a targeted university that used three predetermined test subjects for the given test classification. The individual fixed effects might eliminate the endogeneity. However, we also conducted robustness checks using different sample selections by number of "takes." Second, although the choices of which university to enter were endogenous, the temperature on the test sites were exogenous to students. Students were unlikely to choose a particular university because of an expectation for a favorable temperature on the test day.

We did not include the test-takers who took only a mathematics test (take = 1) in the estimation of Equation (1). The excluded number was 484 500. However, we used a high school fixed effect instead (for take = 1). We also applied dummies for districts wherein the test-takers registered using the family registration. We acknowledge that the estimation using the high school fixed effect cannot be as accurate as Equation (1) for controlling time-invariant factors related to the individuals. However, we used its results to roughly compare among subsamples by number of mathematics tests taken.

Finally, we checked for the robustness of the results against measurement errors and heterogeneity. Following Park *et al.* (2018), we limited our estimations to a sample in which the distance between the nearest weather station and test site was equal to or less than 8 km (approximately 5 miles).[7] Similarly, we analyzed the data using students from the 25 percent hottest regions and 25 percent coolest regions. We also examined heterogeneity by regional economic disparity, policy group, number of takes and gender by re-estimating equations (1) for each sample restriction.

## V. Results

We found that the average temperature at the test site on the test day had a significant impact on the *Z*-scores. An increase of 1°F reduced the *Z*-score by approximately 0.006 SD, as shown in Panel A of column 4 in Table 2, which is our main specification. For example, a sudden change from 77.9°F (25.5°C)—the average temperature between 1950 and 2009—to 82.8°F (28.2°C), 75 percent of the highest recorded average daily temperature in the same period, would cause a reduction of 0.03 SD in the *Z*-score. For students who were born in 1991 and took classification A (see Appendix C), the effect would be equivalent to a 1.64 percent loss ($0.006 \times 4.9 \times 165 \div 296$) in the raw score

[7] In the non-restricted sample, the average distance was approximately 22–23 km, as in Appendix A.

**Table 2 Effects of test day average temperature on test *Z*-score**

| | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* |
|---|---|---|---|---|---|---|
| *A: All regions* | | | | | | |
| Avg. Temp. (°F) | −0.0014*** | −0.0022*** | −0.0041*** | −0.0060*** | −0.0087*** | −0.0049*** |
| | (0.0004) | (0.0003) | (0.0004) | (0.0008) | (0.0005) | (0.0010) |
| $R^2$ | 0.141 | 0.174 | 0.193 | 0.238 | 0.149 | 0.251 |
| *N* (scores) | 641 453 | 641 453 | 641 453 | 641 453 | 357 580 | 357 580 |
| ID (students) | 294 623 | 294 623 | 294 623 | 294 623 | 160 384 | 160 384 |
| *B: 25% hottest regions* | | | | | | |
| Avg. Temp. (°F) | −0.0421*** | −0.0413*** | −0.0280*** | −0.0106*** | −0.0127** | −0.0107 |
| | (0.0022) | (0.0022) | (0.0023) | (0.0028) | (0.0056) | (0.0069) |
| $R^2$ | 0.183 | 0.220 | 0.243 | 0.277 | 0.152 | 0.250 |
| N (scores) | 66 635 | 66 635 | 66 635 | 66 635 | 33 242 | 33 242 |
| ID (students) | 30 923 | 30 923 | 30 923 | 30 923 | 14 733 | 14 733 |
| *C: 25% coolest regions* | | | | | | |
| Avg. Temp. (°F) | 0.0109*** | 0.0081*** | −0.0005 | −0.0154*** | −0.0007 | −0.0059** |
| | (0.0008) | (0.0008) | (0.0008) | (0.0020) | (0.0009) | (0.0026) |
| $R^2$ | 0.126 | 0.162 | 0.185 | 0.230 | 0.155 | 0.270 |
| N (scores) | 177 592 | 177 592 | 177 592 | 177 592 | 59 543 | 59 543 |
| ID (students) | 81 922 | 81 922 | 81 922 | 81 922 | 26 951 | 26 951 |
| *D: All regions and ln(score) as outcome* | | | | | | |
| Avg. Temp. (°F) | −0.0040*** | −0.0037*** | −0.0034*** | −0.0038*** | −0.0059*** | −0.0042*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0006) | (0.0004) | (0.0007) |
| $R^2$ | 0.501 | 0.511 | 0.518 | 0.540 | 0.527 | 0.570 |
| N (scores) | 641 453 | 641 453 | 641 453 | 641 453 | 357 580 | 357 580 |
| ID (students) | 294 623 | 294 623 | 294 623 | 294 623 | 160 384 | 160 384 |
| *Dummies* | | | | | | |
| Time and date of exams | Yes | Yes | Yes | Yes | Yes | Yes |
| Test classification | No | Yes | Yes | Yes | No | Yes |

(*Continues*)

**Table 2 (continued)**

| | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* |
|---|---|---|---|---|---|---|
| University-classification rank | No | No | Yes | Yes | No | Yes |
| University | No | No | No | Yes | No | Yes |
| Distance to test site | No | No | No | Yes | No | Yes |
| *Individual fixed effect* | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes:* In panel D, dependent variable is ln(score) = ln(raw score + 1) instead. Column (5) and (6) limit to where distance between weather station and test site is ≤8 km. Student clustered robust standard errors are in parentheses. ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

(performance). Although the effect seems to be economically negligible at the individual level, the effect will be economically significant at the national level, especially for countries with a population as large as Vietnam's (ranked 15th in the world with approximately 97 million people in 2019).[8] This may, indeed, be the case if we consider a negative impact of heat on problem solving in the workplaces of all workers in a given nation and link the effect to the productivity of the entire nation.

The mechanism determining individual performance relates to heat tolerance. Mathematical problem solving depends on brain function in an area of the prefrontal cortex and neural circuits that is sensitive to heat (Hocking *et al.*, 2001; Graff Zivin *et al.*, 2018). In addition, high temperature would decrease cerebral blood flow (Cho, 2017) and, thus, reduce the efficiency of the test performance.

As seen in Table 2, our results are robust across specifications. The adverse effect of temperature maintains its statistical significance even if we use the outcomes as zero-adjusted logarithms of the raw scores (see Panel D in Table 2). Moreover, when the distance between the weather station and the test site is ≤8 km (5 miles), we still find a negative impact of higher average temperature, as in Panel A of columns 5 and 6 in Table 2. One may consider that students from hotter regions might have higher heat tolerance despite the fact that the majority of time-invariant factors were captured in the student fixed effect. However, as shown in Panel B of column 4 in Table 2, our results indicated that adverse effects of higher temperature would exist even among students from the hottest areas in a tropical country.

Our findings validated the results for an Asian country with a hot climate located close to the equator and expanded the inference power of Graff Zivin *et al.* (2018),[9] which was based on the US data. Our estimations are similar to Park's result (Park, 2020), which was −0.007 SD of the *Z*-score in the specification with student fixed effects. However, our results were based on national census data, whereas Park *et al.* (2020) relied on the data of New York City high school students. We are less likely to be concerned with failing to account for weather impacts on test preparation that occurred during test intervals because test intervals were only 5–11 days long. Our results concur with the findings of Cho (2017) regarding the adverse effect of heat (above 93.2°F or 34°C, compared with temperatures between 82.4 and 86°F, or 28 to 30°C). More importantly, our results were estimated from continuous average temperature and completely controlled for all time-invariant factors.

Our results are robust across the number of mathematics tests taken, as in columns 2 and 3 of Table 3, wherein student fixed effects were used. The

[8] The World Bank Data (link: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=VN&most_recent_value_desc=true).

[9] Graff Zivin *et al.* (2018) find that higher temperatures always decrease math test scores of US children when beyond 78.8°F (26°C).

**Table 3 Effects of average test day temperature by number of mathematics tests taken**

| | *Number of mathematics tests taken* | | |
|---|---|---|---|
| | *One* | *Two* | *Three* |
| *Variables* | *(1)* | *(2)* | *(3)* |
| Avg. Temp (°F) | –0.0027 | –0.0057*** | –0.0075*** |
| | (0.0017) | (0.0010) | (0.0015) |
| *Dummies* | | | |
| Time and date of exams | Yes | Yes | Yes |
| Test classification | Yes | Yes | Yes |
| University-classification rank | Yes | Yes | Yes |
| University | Yes | Yes | Yes |
| Distance to test site | Yes | Yes | Yes |
| High school | Yes | - | - |
| District (family reg.) | Yes | - | - |
| *Individual fixed effect* | No | Yes | Yes |
| $R^2$ | 0.355 | 0.230 | 0.273 |
| N (scores) | 484 500 | 484 832 | 156 621 |
| ID (students) | 484 500 | 242 416 | 52 207 |

*Notes*: High school clustered robust standard errors are in parentheses in (1), while individual clustered robust standard errors are for the rest columns. ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

corresponding coefficients of the average temperature were negative and statistically significant in both specifications. For students taking just a test (column 1 of Table 3), the coefficient is negative and statistically insignificant, perhaps because the individual fixed effects were not in place.

When dividing the data for heterogeneity analysis, we found that average test-day temperature still has a negative effect on $Z$-scores in most estimations, as seen in Table 4. Rural students showed the highest impact of heat per degree Fahrenheit, as in column 2 of Table 4. Furthermore, women were more vulnerable to heat at test sites (column 8 of Table 4). Compared with men, the impact on women is statistically significant and larger. Our results show that some groups are statistically resilient to the effect of daily temperature, including men and people living in economically depressed areas, as in column 1 of Table 4.

Unfortunately, we did not have sufficient data to explain the specific mechanisms underlying the differences in results among these groups. We acknowledge that there might be more than one mechanism to explain these differences. One possible explanation may be differences in nutrition intakes and heat compensation. For example, nutrition intakes and heat prevention measures may be lowest in students living in rural areas. A future study when more detailed information is available might provide better explanations for the observed differences.

**Table 4 Effects of average test day temperature by regional disparity, policy group, and gender**

| | *Regional disparity* | | | | *Policy group* | | *Gender* | |
|---|---|---|---|---|---|---|---|---|
| | *Most difficulty* | *Rural* | *Townships* | *Urban city* | *Ethnic minority* | *State honored* | *Male* | *Female* |
| *Variable* | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* | *(7)* | *(8)* |
| Avg. Temp (°F) | 0.0007 | –0.0108*** | –0.0082*** | –0.0023 | –0.0003 | –0.0020 | –0.0018 | –0.0081*** |
| | (0.0014) | (0.0013) | (0.0020) | (0.0035) | (0.0050) | (0.0044) | (0.0013) | (0.0010) |
| *Dummies* | | | | | | | | |
| Time and exams date | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Test classification | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| University-classification rank | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| University | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to test site | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| *Individual fixed effect* | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.282 | 0.237 | 0.222 | 0.241 | 0.400 | 0.249 | 0.227 | 0.252 |
| N (scores) | 192 762 | 255 658 | 115 415 | 77 618 | 17 719 | 23 312 | 275 496 | 365 957 |
| ID (students) | 89 018 | 117 188 | 53 244 | 35 173 | 8574 | 10 906 | 127 820 | 166 803 |

*Notes*: Individual clustered robust standard errors are in parentheses. ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

## VI. Discussion and Conclusions

We examined the immediate effects of heat on the mathematics test performance (*Z*-scores) of test-takers in 2009 Vietnamese national university entrance examinations. We found that a 1°F increase in the average temperature on the test day was associated with a decreased *Z*-score equal to 0.006 SD.

Our results have several policy implications. Although our results are based on data from the NEEU in 2009, the impact of heat, potentially due to global warming and climate change, might persist for the medium and long run of the economic development path. Vietnam passed the *Laws on Environment Protection* in November 2020 (effective from January 1, 2022), with the article 139 as a foundation for carbon tax policy. Our results might serve as a reference index for this future carbon tax policy of Vietnam as well as its ameliorating redistribution. In addition, our results might also be used as a reference for school administrators when considering investments in school infrastructure. A reasonable heat-diffusion solution determined through cost–benefit analysis might offset the harms economically.

In addition to previously stated limitations, we acknowledge some other drawbacks of our study that should be considered in future work. First, there are selection issues due to some students not taking the entrance exams. The test-taking rate was only approximately 45 percent on average (Vu and Yamada, 2020) among students completing high school in 2009. Some of the students did not have plans to continue in higher education. Second, we acknowledge the measurement errors and biasness it may cause due to the lack of exact temperature by hour and location. Third, we do not have an abundance of variations in temperature across seasons, especially at lower temperatures. Therefore, we were unable to test whether the impact of heat is nonlinear (Dell *et al.*, 2014). Additionally, all measured temperatures occurred in July; therefore, the findings might apply only during the hotter portions of the year.

## ACKNOWLEDGMENTS

I acknowledge insightful comments and suggestions from Shinsuke Uchida, the journal Editors, and participants from the 2019 Spring Meetings of the Japanese Economic Association, the 14th Applied Econometrics Conference in the Osaka University, the 95th Annual Virtual Conference of the Western Economic Association International and the 2020 (Virtual) World Congress of the European Association of Labour Economists, the Society of Labor Economists and Asian and Australasian Society of Labour Economics. I thank the Higher Education Department, the Ministry of Education and Training of Vietnam for publishing the data of test scores—2009 National Entrance Exams to University and College for analysis and all related information in the department's website in 2009. I also thank the US National Centers for

Environment Information and the National Oceanic and Atmospheric Administration for allowing me to use the Daily (Weather) Summaries. The work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, grant number 18K12784, 19H00619, 20H01506 and 21K01455 and research grants from the Asian Growth Research Institute and Kyushu University.

## Conflict of Interest

No conflict of interest

## REFERENCES

Acemoglu., D. and M. Dell, 2010, Productivity differences between and within countries. *Am Econ J: Macroecon*, **2**, pp. 169–88. https://doi.org/10.1257/mac.2.1.169.

Auffhammer, M., S. M. Hsiang, W. Schlenker and A. Sobelz, 2013, Using weather data and climate model output in economic analyses of climate change. *Rev Environ Econ Policy*, **7**, pp. 181–98. https://doi.org/10.1093/reep/ret016.

Cho, H., 2017, The effects of summer heat on academic achievement: A cohort analysis. *J Environ Econ Manag*, **83**, pp. 185–96. https://doi.org/10.1016/j.jeem.2017.03.005.

Dell, M., B. Jones and B. Olken, 2014, What do we learn from the weather? The new climate-economy literature. *J Econ Lit*, **52**(3), pp. 740–98. https://doi.org/10.1257/jel.52.3.740.

Garg, T., Jagnani, M. and V. Taraz, 2017, Temperature and human capital in India. https://doi.org/10.2139/ssrn.2941049

General Office of Statistics (GSO), 2006, Result Summaries: 2016 Census on Rural, Agriculture and Aquaculture [cited August 26, 2019]. Available from URL: https://www.gso.gov.vn/ItemPreview.aspx?ItemID=8057

Graff Zivin, J., S. M. Hsiang and M. Neidell, 2018, Temperature and human capital in the short and long run. *J Assoc Environ Res Econ*, **5**, pp. 77–105. https://doi.org/10.1086/694177.

Graff Zivin, J. and M. Neidell, 2013, Environment, health and human capital. *J Econ Lit*, **51**, pp. 689–730.

Hocking, C., R. B. Silberstein, W. M. Lau, C. Stough and W. Roberts, 2001, Evaluation of cognitive performance in the heat by functional brain imaging and psychometric testing. *Comp Biochem Physiol Pt A: Mol Integ Physiol*, **128**, pp. 719–34. https://doi.org/10.1016/S1095-6433(01)00278-1.

Hoddinott, J. and B. Kinsey, 2001, Child growth in the time of drought. *Oxf Bull Econ Stat*, **63**(4), pp. 409–36. https://doi.org/10.1111/1468-0084.t01-1-00227.

Horowitz, J. K., 2009, The income–temperature relationship in a cross-section of countries and its implications for predicting the effects of global warming. *Environ Res Econ*, **44**, pp. 475–93. https://doi.org/10.1007/s10640-009-9296-2.

Ministry of Education and Training of Vietnam. (MOET), 2008, *Decision of the Minister of MOET on regulations on the national entrance exams to full-time regular university and college programs (In Vietnamese: Quyết định của Bộ trưởng Bộ Giáo dục và Đào tạo về việc ban hành "Quy chế tuyển sinh đại học, cao đẳng hệ chính quy")*. [cited December 18, 2018]. https://thuvienphapluat.vn/van-ban/Giao-duc/Quyet-dinh-05-2008-QD-BGDDT-Quy-che-tuyen-sinh-dai-hoc-cao-dang-he-chinh-quy-62356.aspx.

Park, J. R., J. Goodman, M. Hurwitz and J. Smith, 2020, Heat and Learning. *Am Econ J: Econ Policy*, **12**, pp. 306–39. https://doi.org/10.1257/pol.20180612.

Park, J. R., 2020, Hot temperature and high stakes cognitive performance. *J Hum Res*. https://doi.org/10.3368/jhr.57.2.0618-9535R3.

Park, J. R., M. Bangalore, S. Hallegatte and E. Sandhoefner, 2018, Households and heat stress: estimating the distributional consequences of climate change. *Environ Develop Econ*, pp. 1–20. https://doi.org/10.1017/S1355770X1800013X.

Thai, T. Q. and E. M. Falaris, 2014, Child schooling, child health, and rainfall shocks: Evidence from rural Vietnam. *J Develop Stud*, **50**(7), pp. 1025–37. https://doi.org/10.1080/00220388.2014.903247.

Vu, T. M. and H. Yamada, 2020, Persistent legacy of the 1075–1919 Vietnamese imperial examinations in contemporary quantity and quality of education. AGI Discussion Paper Series, WP2020-12, 2020. [Cited July 22, 2020]. http://id.nii.ac.jp/1270/00000176/.

## Appendix A

*Distance between nearest weather station and test sites*

| *Variable* | *Mean* | *Standard deviation* |
|---|---|---|
| *Test date: July 4, 2009* ($N = 468\ 002$) | | |
| Weather station and test sites | 23.884 | 31.459 |
| *Test date: July 9, 2009* ($N = 415\ 835$) | | |
| Weather station and test sites | 21.487 | 21.784 |
| *Test date: July 15, 2009* ($N = 242\ 116$) | | |
| Weather station and test sites | 20.625 | 21.608 |

*Note:* Distance unit is in kilometer.

## Appendix B

*Students in test sites correspondent with weather stations*

| *Station name/Location* | *July 4, 2009* | *July 9, 2009* | *July 15, 2009* |
|---|---|---|---|
| | Frequency | Frequency | Frequency |
| Lao Cai | 0 | 0 | 153 |
| Son La | 2600 | 2302 | 1699 |
| Noi Bai Intl (Hanoi) | 168 889 | 128 384 | 67 174 |
| Nam Dinh | 1060 | 12 588 | 10 555 |
| Phu Lien (Hai Phong) | 17 691 | 16 718 | 15 877 |
| Lang Son | 0 | 0 | 915 |
| Thanh Hoa | 2431 | 4445 | 0 |
| Vinh | 19 760 | 16 098 | 6256 |
| Dong Hoi (Quang Binh) | 846 | 902 | 0 |
| Phu Bai (Hue) | 15 006 | 27 254 | 4498 |
| Da Nang Intl (Da Nang) | 23 247 | 12 418 | 16 028 |
| Quang Ngai | 3990 | 956 | 3335 |
| Pleiku | 459 | 23 | 678 |
| Quy Nhon | 32 686 | 28 054 | 4074 |
| Ban Me Thuot | 5964 | 8024 | 0 |
| Nha Trang | 7896 | 9192 | 4310 |
| Tan Son Nhat Intl (Ho Chi Minh city) | 128 748 | 106 256 | 102 689 |
| Rach Gia (Phan Thiet) | 260 | 42 221 | 2231 |
| Ca Mau | 36 469 | 0 | 1644 |
| *N* (test scores) | 468 002 | 415 835 | 242 116 |

## Appendix C

*Raw mathematics test scores by year of birth and test problem (Main groups only)*

| | | University entrance examinations | | | College entrance examinations | | |
|---|---|---|---|---|---|---|---|
| *Classification* | *YOB* | N | Mean | Standard deviation | N | Mean | Standard deviation |
| A | | | | | | | |
| | 1991 | 343 609 | 296.29 | 165.62 | 138 523 | 530.44 | 206.52 |
| | 1990 | 84 412 | 259.60 | 160.66 | 33 190 | 478.51 | 203.59 |
| | 1989 | 18 784 | 237.05 | 154.90 | 6896 | 430.85 | 197.34 |
| | 1988 | 6309 | 248.53 | 164.53 | 2207 | 419.74 | 193.38 |
| B | | | | | | | |
| | 1991 | 180 104 | 333.01 | 212.39 | 12 065 | 371.61 | 217.69 |
| | 1990 | 48 723 | 254.46 | 179.29 | 5503 | 328.56 | 208.12 |
| | 1989 | 10 510 | 214.93 | 160.70 | 1301 | 261.16 | 184.84 |
| | 1988 | 3306 | 218.90 | 166.42 | 396 | 250.63 | 180.91 |
| D | | | | | | | |
| | 1991 | 112 025 | 314.62 | 213.84 | 28 466 | 387.79 | 208.73 |
| | 1990 | 22 658 | 253.49 | 178.85 | 6558 | 343.77 | 200.00 |
| | 1989 | 4360 | 221.55 | 161.79 | 1150 | 296.96 | 182.94 |
| | 1988 | 1299 | 226.71 | 170.56 | 285 | 285.71 | 186.60 |
| M | | | | | | | |
| | 1991 | 7593 | 143.49 | 114.95 | 4707 | 239.45 | 169.00 |
| | 1990 | 2551 | 119.56 | 96.70 | 1969 | 194.31 | 160.32 |
| | 1989 | 752 | 104.02 | 84.97 | 670 | 169.59 | 158.50 |
| | 1988 | 220 | 103.41 | 92.47 | 219 | 168.84 | 159.82 |
| T | | | | | | | |
| | 1991 | 10 229 | 128.66 | 82.79 | 453 | 212.58 | 145.81 |
| | 1990 | 4764 | 110.83 | 66.69 | 408 | 156.80 | 136.42 |
| | 1989 | 1844 | 99.62 | 58.89 | 221 | 117.18 | 96.41 |
| | 1988 | 782 | 99.13 | 62.77 | 97 | 109.54 | 89.15 |

*Note:* The table consists of 1 110 118 test scores. The total number of mathematics test scores available for analysis is 1 125 952.