In a Hotter World. The Effect of Temperature on Students' Performance*

Rosario Maria Ballatore[†] Alessandro Palma[‡] Daniela Vuri[§]

June 15, 2022

Abstract

We investigate the effect of higher temperatures on student performance in Italy. We align administrative data on mandatory students' tests with detailed weather information at municipality level. Our analysis considers standard test results and new emotional perceptions outcomes that allow to better understand how students respond to higher outdoor temperatures, in a setting characterized by low air conditioning penetration. Using test-to-test random exposure to temperature, our results show significant reduction in performance, with stronger effects for math and for younger students. We also found evidence of emotional disruption when temperatures the day of the test are high.

Keywords: student performance, cognitive ability, emotional stress, temperature, climate change

JEL: J21, J24, Q54, O15

---

*We are extremely thankful to Giulia Bovini, Luca Citino and Guido De Blasio for useful suggestions. We also thank the participants to the Bank of Italy webinars within the project "Gli effetti del cambiamento climatico sul sistema produttivo italiano". We are solely responsible for any and all errors. The views expressed in this paper are those of the authors and do not reflect those of the Bank of Italy.

[†]Bank of Italy, Regional Research Division, Largo Carlo Felice 13, Cagliari, Italy, e-mail: `rosariomaria.ballatore@bancaditalia.it`

[‡]Gran Sasso Science Institute (GSSI), Social Sciences Area, Via Francesco Crispi 7, 67100 L'Aquila, Italy and CEIS, Tor Vergata University, Via Columbia 2, Rome, Italy, e-mail: `alessandro.palma@gssi.it`

[§]Department of Economics and Finance, Tor Vergata University and CEIS Tor Vergata, Via Columbia 2, Rome, Italy, e-mail: `daniela.vuri@uniroma2.it`

1 Introduction

Education is the key ingredient of human capital development and a strong predictor of lifetime income (Card, 1999). A recent but influential body of research has shown that external factors, such as pollution externalities and hotter temperatures, negatively affect educational outcomes, especially in high-stake settings (Graff Zivin et al., 2018, among others). Despite the knowledge advances provided by these studies, we still know very little on the mechanisms affecting student performance and learning. This issue assumes even more relevance as scientists expect a warmer world and more extreme weather events in the near future.

Recent evidence from the Intergovernmental Panel on Climate Change (IPCC) warns that in a business-as-usual scenario we will reach the critical threshold of 1.5°C increase in the global temperature between 2030 and 2050.[1] While these figures may appear as small changes in the global climate, a large geographical and temporal heterogeneity exists in the distribution of these impacts as some regions and seasons are experiencing much higher warming increases than the global average. It is also well documented that those changes translate into a more extreme local weather. As recently documented by the latest report of the World Meteorological Organization (Pritchard et al., 2021), the frequency, duration and intensity of heat waves and other climate extremes have dramatically increased over the last 15 years, with this scenario becoming the "new norm". In 2020/2021, several exceptional heatwaves hit western North America: British Columbia in Canada registered a temperature peak of 49.6°C with associated nearly 600 deaths and in California the temperature reached the abnormal value of 54.4°C. In Europe, during the exceptional heat occurred in the second week of August, a weather station in Southern Italy reached 48.8°C, establishing a new European record. These phenomena are increasingly struggling policy makers to find an effective strategy to mitigate the externalities of a hotter world and to rapidly adapt to the "new abnormal" (Xu et al., 2018).

In this paper we study the effect of hotter temperatures on student performance by analyzing administrative data of low-stake mandatory tests in Italy from 2012 to

[1] According with the IPCC, "temperature extremes on land are projected to warm more than GMST (high confidence): extreme hot days in mid-latitudes warm by up to about 3°C at global warming of 1.5°C and about 4°C at 2°C, and extreme cold nights in high latitudes warm by up to about 4.5°C at 1.5°C and about 6°C at 2°C (high confidence). The number of hot days is projected to increase in most land regions, with highest increases in the tropics" (Allen et al., 2018).

2017, aligned with granular weather data available at municipality level. Studying the effect of temperature shocks on student performance is important for several reasons. Firstly, short run effects on performance could be a piece of evidence of a deterioration in student learning and skills, with potentially negative impacts in the long run on labor market outcomes and aggregate economic growth (Deschênes, 2014; Graff Zivin et al., 2018). Secondly, cognitive performance is crucial for many important steps in our life (such as public competitions, college admission, financial decisions) and evidence of its reduction at high temperatures could have potentially important implications for the optimal scheduling of cognitively demanding tasks (Graff Zivin and Neidell, 2013). Finally, the results of school assessments are often used to provide geographical comparisons among different areas within countries. We claim that a fare comparison should be conducted considering the heterogeneous exposure to temperatures stress.

Our empirical strategy is very simple. Since students cannot manipulate the date of the test[2], we exploit day-to-day variation in temperature across multiple tests of the same student that rule out possible bias due to sorting or other unobserved determinants of student performance, allowing to capture the causal effect of hotter temperatures.

We are not the first to study the relationship between temperature and school performance. Using multiple low-stake test observations within students, aligned with county-level weather conditions, Graff Zivin et al. (2018) estimate that math scores decline significantly above 26°C, with no effect for reading performance. Park (2020) estimates the effect of high temperatures on high-stake test scores linked with subsequent educational attainment (high school graduation and diploma status) in a large sample of US students. Using student fixed effects, the study finds that a one degree (F) increase in the temperature reduces performance by -0.009 s.d., with results in line with Graff Zivin et al. (2018). Moreover, the author did not find significant differences on mathematical versus verbal reasoning. However, about 70% of these schools benefited from air conditioning during the study period, which might have downward biased the results as the effect of temperature are largely attenuated in a colder controlled environment. Park et al. (2020) study how heat temperatures affects school learning using data from PSAT examinations in the United States. Their student fixed-effect estimates show that 1°F hotter school year reduces by 0.002 s.d. the test score, with

[2] All test dates are set several months in advance by Invalsi.

larger impacts for black and Hispanic students. On the contrary, in schools with air conditioning the effects seem much smaller. Focusing on learning, Cho (2017) considers the effect of summer heat on academic achievement using individual-level data on high-stake exam scores taken to college entrance in Korea during summer. School fixed-effect estimates unveil that an additional day with a maximum daily temperature above 34°C decreased the scores of math of 0.0042 s.d. and of English of 0.0064 s.d. compared with a day with temperature in between 28°C and 30°C.

Our contribution to the literature is twofold. First, since Invalsi data contain also information on students' perceptions while taking the test (such as anxiety), we consider emotional disruption on the complexity of the effect coming with high temperatures. Moreover, the universal coverage of the data (over 6 million students) enables us to run a rich heterogeneity analysis across grades to explore the effects at different student ages, without loosing statistical power.

Our estimates of the effects of temperature on school performance present also another advantage. The analyses is conducted in a context of low air conditioning penetration in school buildings. According to the survey on school infrastructures by the Ministry of Education, less than 2% of buildings in Italy are equipped with air conditioning. Our analysis therefore considers the effect of temperature stress net of a possible influence of a controlled environment.

Our main results point to a significant and negative effect of hotter temperature on test scores: a one degree (°C) increase in the maximum outdoor temperature decreases language test scores by 0.5 percentage points and the math score by 1. The differences between math and language scores and their magnitude are in line with previous research, which also highlights that temperature deferentially affects different parts of the brain that work on different subject areas (Graff Zivin et al., 2018, among others). We also estimate non-linear effects of temperatures using 3-degree temperature bins, finding that the effects are driven by extreme temperature values. All these findings are supported by extensive robustness checks that validate our main estimates. When exploring possible effects heterogeneity cross ages, we find significant effects only for younger students (grade 2^{nd} and 5^{th}) in both subjects and an absence of effects for older students (grade 8^{th} and 10^{th}). Finally, consistently with the effects across grades, we find sharp negative impacts on student emotional outcomes for temperature values above 34°C only for grade 5^{th}, while for grade 10^{th} students seem not to be significantly worried and affected by anxiety neither before or during the test.

2 Data

The National Students' Assessment Survey (SNV) conduced by INVALSI is a national evaluation program designed to assess students' achievement in Italy at different points of their school career. Starting from the academic year 2009-10 the program took the form of a compulsory census and it is held on an annual basis. The assessment focuses on reading and mathematics competencies of students attending grades 2, 5, 8 and 10[3] by means of a standardized testing procedure. Students are asked to answer a series of questions of different difficulties aimed at testing different skills: reading comprehension, grammar and lexical competences for the reading test and problem solving and logical skills for mathematics. Students' scores provided with the SNV take the form of a standardized variable with mean 200 and a standard deviation of 40 within each grade/academic year and take into account the heterogeneity in the difficulty of the different items that make up the test through a statistical model called Rasch analysis. This procedure makes the resulting test scores comparable between grades and school years, letting us prefer this more accurate measure rather than the standard percentage of correct answers also provided with the survey. To better interpret our results we standardize these scores within grade and academic year with zero mean and a unitary standard deviation. Besides assessing competencies, Invalsi collects also a background questionnaire for grades 5 and 10 only (known as Student Questionnaire) containing information on students' perceptions while taking the tests, such as anxiety, feeling of performing badly or feeling fine during the assessment.

Each child is tested across multiple waves and it is possible to follow students over several school years. However, since the survey does not contain all grades, students are not observed in all the waves. To give an example: a student attending grade 5 in the academic year 2011-12 will be observed again in grade 8 in the year 2014-15 and in grade 10 in 2016-17. In the original sample of around 7,000,000 students 60 percent of them are observed at least twice, with the 10 percent being tested three times.

[3]Grades 2^{nd} and 5^{th} correspond to ISCED level 1 (primary schools), grade 8^{th} to ISCED level 2 (lower secondary), 10^{th} corresponds to ISCED level 3 (upper secondary school). Starting from the 2018-19 also grade 13 is tested.

The assessment is carried out every year in the first ten days of May for 2^{nd}, 5^{th} and 10^{th} grades and in the first half of June for grade 8.[4] Crucially for our analysis the days of assessment are the same for the whole national territory and cannot be manipulated by schools or regions. Furthermore the dates are set centrally at the beginning of each year, making impossible to predict climate conditions on the day of the test as the examination procedure is already scheduled. There is a difference between grades in the scheduling of reading and mathematics assessment: they take place within the same day for grades 8^{th} and 10^{th} and on two different days for gardes 2^{nd} and 5^{th} (primary schools).

We use information on geographic location of schools to match our data with climate conditions on the days of the assessment at the municipality level using information taken from Agri-4-Cast, which contains freely available daily data on minimum, maximum and average temperatures, as well as on rainfall and wind speed. In this paper, we focus on the maximum daily temperature rather than the its average as the tests take place in a time slot in which external temperatures are close to its maximum (around noon). Figure 1 shows the maximum temperature in the Italian municipalities on the days of the assessment in all the academic years and grades in our sample. The picture highlights a marked geographical heterogeneity, with the municipalities of Central and Southern Italy and those of the "Pianura Padana" valley more exposed to higher temperatures.

We limit our analysis to the six consecutive test waves from 2011-12 to 2016-17. Two reasons lie behind this restriction. Firstly, it is not possible to observe a unique student identifier before the academic year 2011-12, making it impossible to follow the student over time for those waves. Secondly, from the academic year 2017-18 the assessment procedure is computer-based and is carried out on multiple days, without any possibility to retrieve the exact day of the test.

Table 2 presents some descriptive statistics based on our final sample of more than 6,200,000 observations, 3,000,000 students, 31,000 schools in over 6,000 municipalities (on a total of 7,900 municipalities). The differences between language and math samples derives largely from the fact that the assessment procedure in the primary school is held in two different days. The average maximum temperature is about 23.8°C with peaks

[4]Differently from the other grades the testing procedure for grade 8 is part of the final examinations that take place at the end of the second cycle of education.

of 38°C in some locations, reflecting a large variation in temperature in the assessment dates.

3 Empirical strategy

To study the relationship between temperature and students' performance we estimate the following regression model:

$$y^{f}_{igsct} = \beta_0 + \beta_1 f(T_{ct}) + \beta_2 V_{ct} + \beta_3 Z_{gst} + \delta_i + \tau_t + \sigma_g + \theta_w + \pi_{rt} + \varepsilon_{igsct} \qquad (1)$$

The test score (y) in subject $f \in \{\text{language,math}\}$ of student i attending grade g in the academic year t is regressed on the maximum temperature experienced in the day of the test in municipality c. We control for weather conditions at municipal level (V_{ct}) such as total precipitation and wind-speed and for a vector of time-grade-school variables (Z_{gst}) including the share of immigrants, the share of female and the average class size at school level.

Our identification strategy is straightforward. We simply rely on the longitudinal structure of the data. This allows us to augment the specification with student fixed effects (δ_i), controlling for any time invariant characteristics of a child. The exploited variability comes from random variation in temperature exposure of students across different academic years and grades. Since the date of the test is centrally prearranged several months in advance without any possibility of manipulation for schools, it is reasonable to assume that the variability we use is as-good-as-random. As a consequence we are able to identify the causal effect of temperature on students' performance. Figure 2 shows the within student test-to-test variation in maximum temperature in our sample.

Our specification also includes academic year (τ_t), grade (σ_g) as well as day of the week (θ_w) fixed effects. We also control for a region specific non linear time trend, to take into account for time-varying factors common at regional level that may be correlated with temperature and may influence performance at the same time. The choice of region as the geographical unit depends upon the fact that Italian regions (in agreement with the Ministry of Education) have the possibility to legislate as far as education is concerned. Standard errors are clustered at municipality level to solve three potential issues: arbitrary spatial correlation across municipalities, autocorrelation in test scores over time and assignment of the same temperature to several children.

Temperature is included in our model using two distinct approaches: (1) a linear function of the maximum temperature registered in the day of the test, ranging from a minimum of 0.6 to a maximum of 38°C; (2) a full set of dummy variables in 3°C bins of maximum temperature from 14°C, with temperatures below that threshold taken as the reference category to explore the non-linearity between performance and heat exposure.

4 Results

Table 3 shows the results based on our baseline specification for the two outcomes of interest: language and mathematics. In columns (1) and (3) we simply include the linear effect of maximum temperature, while columns (2) and (4) include the non-linear effect by 3-degree bin of maximum temperature. All specifications include student, year, grade, day-of-the week and region×year fixed effects. The results in columns 1 and 3 indicate that hotter temperatures lead to a statistically significant decrease in performance. The estimate of -0.00477 in language and -0.0100 in math imply that an increase of 1°C lowers the language score by 0.5 percentage point and the math score by 1, therefore with an effect on math almost doubled. The discrepancy in impacts by subjects has been already highlighted by (Graff Zivin et al., 2018), who recognize that mathematical problem solving uses specific part of the brain not used by other subject areas, and temperature deferentially affects different part of the brain. Columns 2 and 4 present estimates for the two outcomes using the more flexible specification for temperature. The results show that child performance in language and math decreases monotonically in temperature, with a larger gradient for math. Figures 1 and 2 plot the corresponding estimates of Columns 2 and 4, showing clearly that the decline in performance is flatter for language than for math. If we look at the extreme temperatures, considering changes from values below 14°C to 35-39°C, the child's math score decreases by 0.202 of a standard deviation, while the language score by only 0.102 of a standard deviation. Even though the magnitude of these effects may appear small if compared to standard educational inputs, they in line with those estimated by Graff Zivin et al. (2018) and by Park (2020).

One concern in interpreting our results could be related to the fact that temperatures may affect cognitive performance only when they exceed the values students are used to experience. For example students from southern Italy may be more accustomed to study when it is hot, so that temperatures in their usual range are ineffective for them but could be detrimental for the performance of others. This motivates a robustness exercise where we augment our specification with the interaction between the maximum temperature the day of the test and a dummy indicating whether the temperature exceed its three years average the same months of the assessment.[5] Since the tests are taken in a hot period of the year, with relatively high temperatures, only 20 percent of the students in our sample experienced an excess of heat the day of the test, that ranges from 0.003 to 11.8°C. Table 4 shows the results of this exercise using both the linear maximum temperature interacted with the excess of heat dummy (columns 1 and 3) and the linear temperature interacted with multiple dummies signaling 3 $C°$ bins interval of excess of heat (columns 2 and 4). Evidence displayed in the table says that one degree °C increase over the range of temperature students used to experience has the same effect of the same temperature variation below this threshold. This is consistent with the fact that human body does not quickly get used to temperatures, as far as brain activity is concerned.

A debated issue when estimating the effect of temperature using test-to-test variation among different academic years relates to the possibility that students or schools learn from past tests exposure to warm temperatures and engage in compensatory behaviors in subsequent assessments. This is what the literature refers as *avoidance behavior*. In our specific framework it could be that students put more effort in studying for the test when they assume the day of assessment is going to be hot. Another possibility is that teachers act to compensate for the disruption of performance when they know, from their past experience, that extremely high temperatures affect students' performance. To check for this we exploit the fact that for grades 2 and 5 the assessment of language and math take place in two different but close days. Therefore, we run a regression where we control for student×grade×academic year fixed effect as in Park (2020) lever-

[5] May for grades 2, 5 and 10. June for grade 8.

aging exogenous variation in temperatures observed in two close days between subjects to identify the effect of interest. As the time span between the two tests is very short (approximately two days), it is very unlikely that avoidance behaviors take place. Figure 4 displays the non linear estimates using this identification strategy for grades 2 and 5. Even though the results are imprecise for higher temperatures (given the low variability in temperature between close days) we observe a clear declining negative effect of temperature on students test scores and we take this as an evidence that avoidance behavior does not represent an issue in our framework.

Additional robustness checks are presented in Table 5 where we show that our results are qualitatively the same irrespective of the fixed effects structure of the specification. In Table 6 we present other robustness analyses. First, running a model that adds grade×academic year fixed effects to control for grade-specific time trends (column 2), does not alter the results. We also run our regression dropping students who change municipality between tests (*movers*). This exercise is motivated by the fact that inclusion of movers could bias our results if decision of moving is correlated both with climate conditions and with test scores (Graff Zivin et al., 2018). Results presented in column (3) suggest that migration is not an issue in our framework. Moreover, in column (4) we control for maximum temperature the day before the test, with a point estimate that is slightly reduced in language but remains unaffected in math. In Figure 5 we plot non linear estimates of the effect of temperature on test scores in the same fashion of Figure 3 using maximum temperatures 20 days before and after the test as an additional falsification test. We do not use days very close to the assessment as temperatures are highly serially correlated. Finally, Figure 5 shows that there is no evidence of the effect of temperature when we consider days different form the one of the test. Particularly the results obtained using lead values of maximum temperature reassure us that our findings are not driven by unobserved confounding.

5 Evidence of the effects of temperature by age

In this section we offer a new set of results by breaking down our analysis by grade. These estimates allow to disentangle the effect of thermal stress across different students' ages. For this purpose we employ school fixed effects specification instead of using student fixed effects because we lose the longitudinal dimension of the data when we run regressions by grade. Thus the variability we exploit is the test-to-test variation in temperatures within school between adjacent academic years. Although this specification slightly differs from the one used in our baseline estimates presented in Table 3, evidence provided in Table 5 show that results are fully consistent across models. Although very little is known yet on how the thermoregulatory mechanism of the human body changes with age, in particular for the brain, the medical literature provides some evidence that growth and maturation are accompanied with physiological changes in the thermoregulation system, suggesting that it improves with age and declines in the elderly (Van Someren, 2007).[6] Considering that our data allow to cover four grades, which correspond to a age range between 7 and 15 years old, our estimates allow to explore the effects of temperature in the most critical window of growth of the grade schooling phase.

Panel A and B of Figure 6 show these estimates for 2^{th}, 5^{th}, 8^{th} and 10^{th}, respectively for language and math skills. The graphs report estimates across different temperature bins as in column (2) and (4) of Table 3. Both for math and language, we observe a significant negative effect of temperature on test score for higher temperature bins only for grades 2^{th} and 5^{th}, corresponding respectively to 7 and 10 year-old students. In particular, for language the effects become significant for temperature bins exceeding 34°C, while for math the effects manifest when temperature exceeds 31°C. Conversely, the effects almost disappear for grades 8^{th} and 10^{th} in both cognitive domains, with weakly significant and negative impacts for grade 8^{th} at the highest temperature bin (35-39°C), especially for math. Overall this set of results shows that students in middle adolescence (age 13-15) are much less sensitive to thermal stress than students pre-teen age (age 7-10). These results are in line with previous medical literature, even though they should be interpreted with cautious in absence of a well identified mechanism.

[6]Previous research has highlighted some specific mechanisms. The different morphology of the human body between children and adults make children's sweating capacity lower, particularly at extreme temperatures (Falk and Dotan, 2008). Specifically, the reduced sweating capacity in children has been identified in a lower sweating rate per gland and not to a lower number of sweat glands, i.e. a higher density of heat-activated sweat glands but a smaller size of sweat glands causing a lower sensitivity of the sweating mechanism and metabolic capacity (Székely and Garai, 2018).

6 Temperature and emotional disruption

An additional aspect to be considered when interpreting our findings relates to the effect of temperature on emotional disruption before and during the test. In order to address this issue we take advantage of an additional questionnaire administered by Invalsi exclusively to 5^{th} and 10^{th} graders which collects additional information on student feelings and motivation during the tests. In particular we focus on four questions that are collected from 2011/12 to 2016/17 regarding the student's perceived stress before and during the test: i) being worried before the test; ii) feeling the test was not going well; iii) feeling anxiety during the test; iv) feeling fine during the test. Questions 2 and 4 mirror each other and can be considered a double check on the accuracy of the students' answers. Table 2 displays summary statistics for these variables, transformed into dummy indicators (e.g. anxiety is equal to 1 when students is anxious while taking the test). Since we have only grades 5 and 10 and the questions are asked for a limited number of years, our identification strategy relies on controlling for the same set of variables V_{ct} and Z_{gst} used in equation 1 plus school, year, weekday and region×year fixed effects, without student fixed effects.

Panel A and B of Figure 7 display estimates for each of the four emotional outcome variables (transformed into dummy variables) using the more flexible specification for temperature. These figures convey a different message: the relationship between temperature and emotional disruption is rather flat and not significantly different from zero for 5^{th} grade students, except at extreme temperatures, i.e. above 35°C. Our emotional outcomes are instead insensitive to temperature for 10^{th} grade students.

7 Conclusions

In this paper, we merge administrative data on students with meteorological data to provide evidence of the relationship between temperature and students' performance. We find that changes in temperature lead to significant decreases in cognitive performance, with effects larger for math and for younger students. We also find evidence of emotional disruption at extreme climate conditions the day of the test.

Our analysis of the effects of hotter temperatures on student performance have important direct policy implications. First, our findings help policy makers design effective strategies to circumvent the side effects of extreme heat to make school assessments more even, mitigating the impacts of external factors that differentially affect individuals who live in different places or who take the tests in the most exposed periods. Moreover, our analysis stimulates the debate about the quality standard of school facilities as school buildings in many advanced economies are seldom equipped with air conditioning. In the case of Italy, official data collected by the Ministry of Education and Research (MIUR) show that less than 2% of school buildings benefited from air conditioning in the academic year 2020-2021.[7] Moreover, a report of the Ministry of Health highlights that there is no systematic and updated legislation aimed at regulating the hygienic and functional requirements of school environments in relation to ventilation standards and air quality thresholds.[8]. These figures, together with the negative effects presented in our study, claim a systematic policy intervention that would allow a significant improvement of human capital investments and a more even chance of success to each individual.

[7]See `https://dati.istruzione.it/opendata/opendata/catalogo/elements1/leaf/?area=Edilizia%20Scolastica&datasetId=DS0176EDITIPORISCSTA2021`

[8]Source: `https://www.salute.gov.it/imgs/C_17_pubblicazioni_1892_allegato.pdf`

References

Allen, M. R., Babiker, M., Chen, Y., de Coninck, H., Connors, S., van Diemen, R., Dube, O. P., Ebi, K. L., Engelbrecht, F., Ferrat, M., et al. (2018). Summary for policymakers. In *Global Warming of 1.5: An IPCC Special Report on the impacts of global warming of 1.5\C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* IPCC.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3:1801–1863.

Cho, H. (2017). The effects of summer heat on academic achievement: a cohort analysis. *Journal of Environmental Economics and Management*, 83:185–196.

Deschênes, O. (2014). Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics*, 46:606–619.

Falk, B. and Dotan, R. (2008). Childrenâs thermoregulation during exercise in the heatâa revisit. *Applied Physiology, Nutrition, and Metabolism*, 33(2):420–427.

Graff Zivin, J., Hsiang, S. M., and Neidell, M. (2018). Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists*, 5(1):77–105.

Graff Zivin, J. and Neidell, M. (2013). Environment, health, and human capital. *Journal of Economic Literature*, 51(3):689–730.

Park, R. J. (2020). Hot temperature and high stakes performance. *Journal of Human Resources.*

Park, R. J., Goodman, J., Hurwitz, M., and Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy*, 12(2):306–39.

Pritchard, H. D., Turner, J., et al. (2021). State of the global climate in 2020.

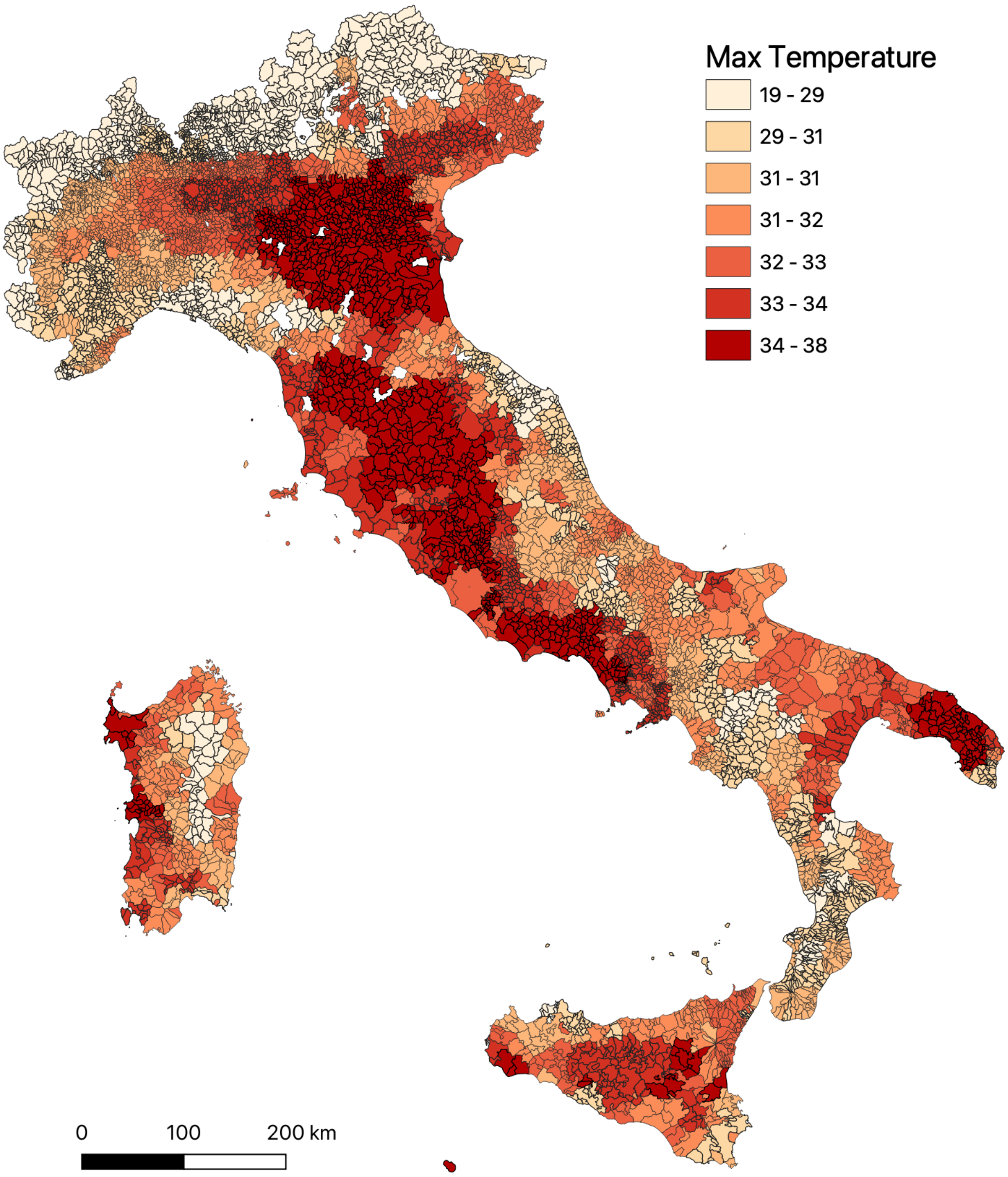
Székely, M. and Garai, J. (2018). Thermoregulation and age. *Handbook of clinical neurology*, 156:377–395.

Van Someren, E. J. (2007). Thermoregulation and aging. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 292(1):R99–R102.

Xu, Y., Ramanathan, V., and Victor, D. G. (2018). Global warming will happen faster than we think.

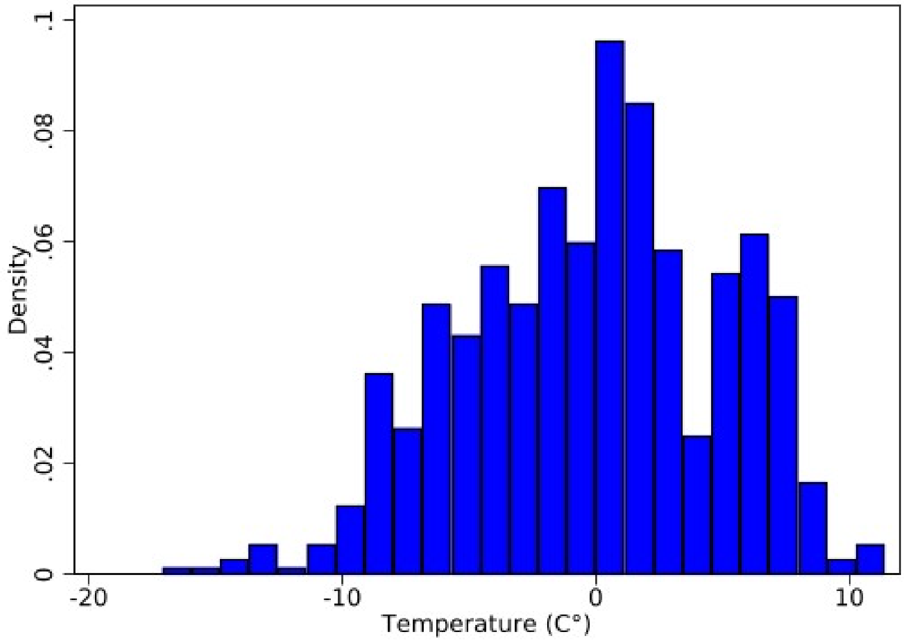
Figures

Figure 1: Maximum Temperature During The Tests



Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. The figure displays the average maximum temperature registered in each municipality during the days of the test. Temperature is measured in°C.

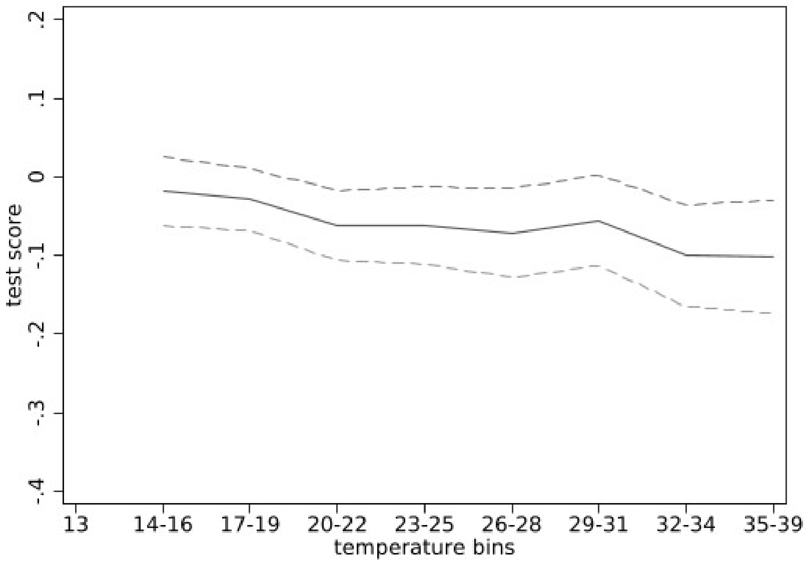
Figure 2: Test-To-Test Temperature Variation



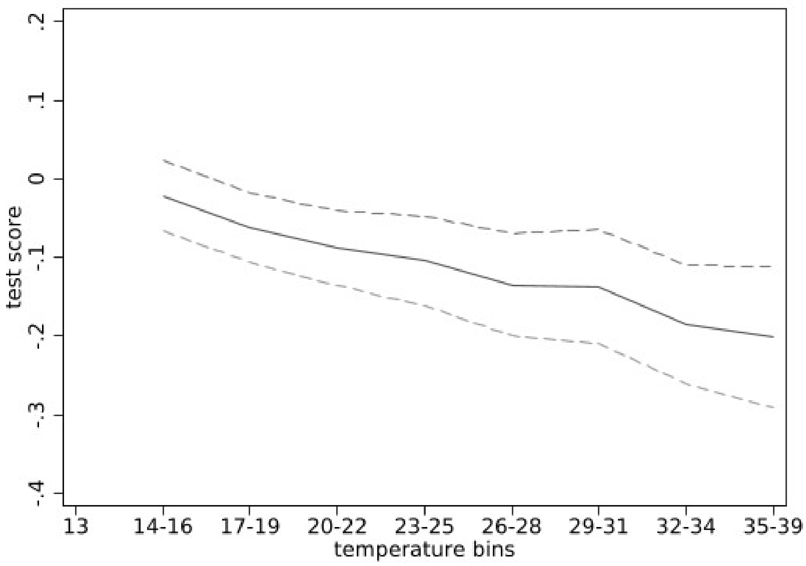
Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. The figure displays the absolute test-to-test variation in maximum temperature registered the day of the assessment experienced by students. Temperature is measured in°C.

Figure 3: Non-Linear Effect Of Temperatures On Students Test Score

A. Language

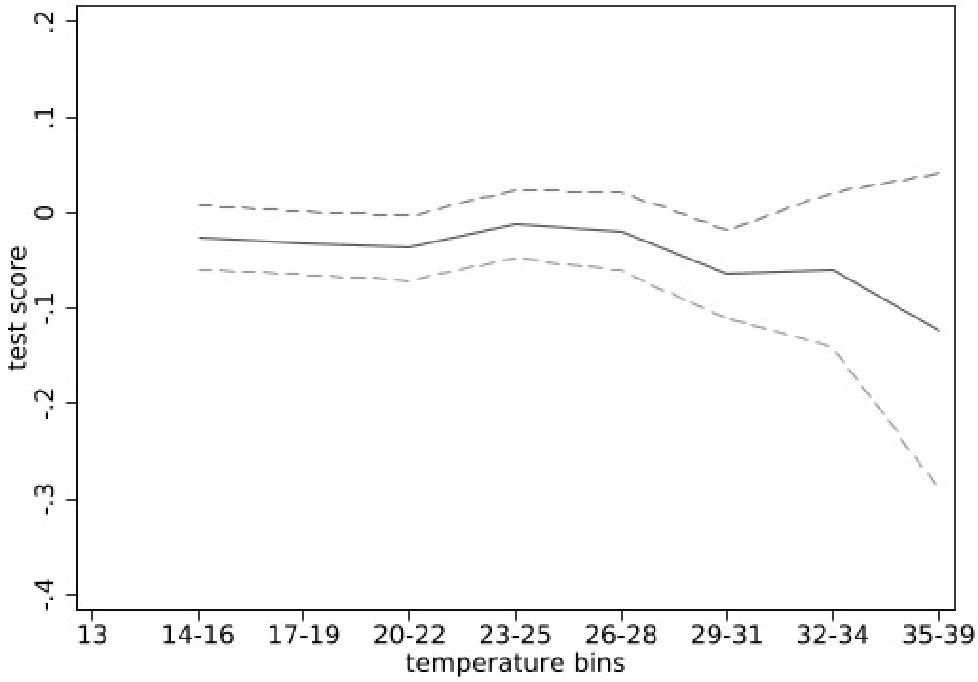


B. Math



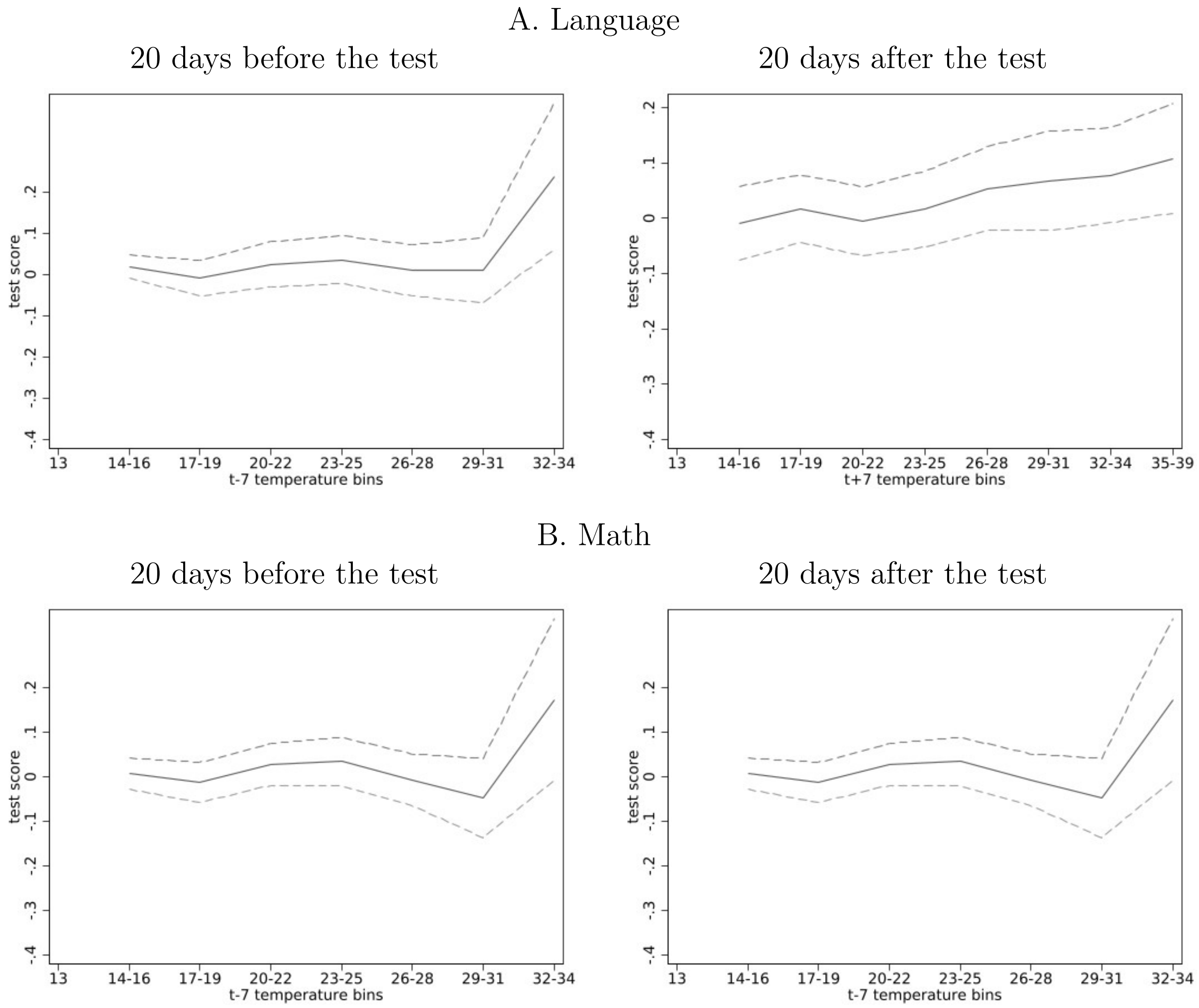
Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. The figure displays non linear estimates of the effect of temperatures the day of the test on students test score. Bins are grouped by 3°C maximum temperature intervals. Temperatures less than 14°C are the benchmark category. The dependent variable is the standardized test score with zero mean and unitary standard deviation. Estimates include controls for rainfall and wind-speed the day of the test at municipal level, share of female students, share of immigrants, share of early enrolled, share of retained, average class size at school level, along with student, academic year, grade, day-of-week and region×academic year fixed effects. Standard errors are clustered on municipalities.

Figure 4: Robustness Check: Between Subjects Non-Linear Effect Of Temperatures On Students Test Score



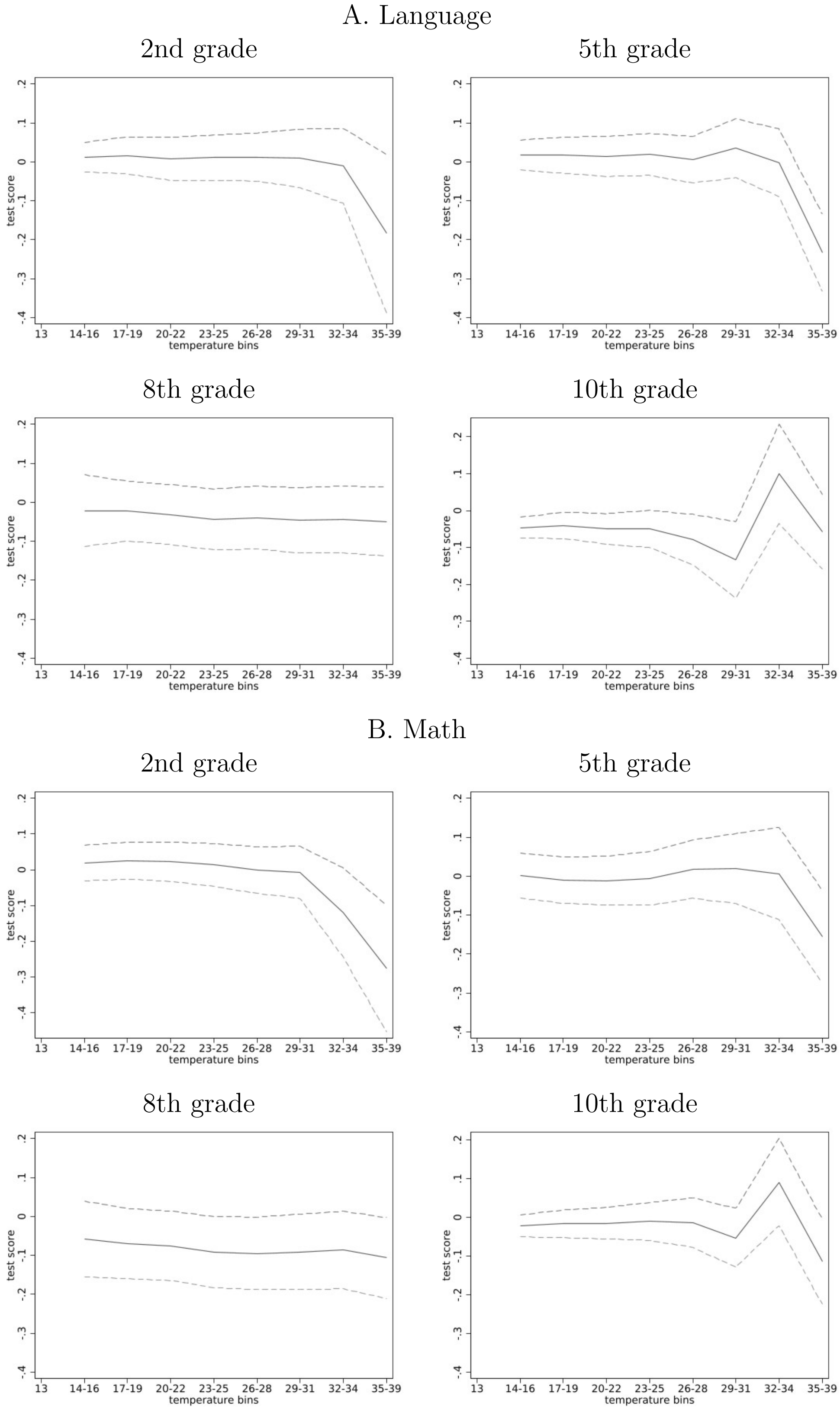
Notes: Pooled sample of grades 2 and 5 in academic years from 2011-12 to 2016-17.The figure displays non linear estimates of the effect of temperatures the day of the test on students test score. Bins are grouped by 3°C maximum temperature intervals.Temperature less than 14°C are the benchmark category. The dependent variable is the standardized test score with zero mean and unitary standard deviation. Estimates include controls for rainfall and wind-speed the day of the test at municipal level, along with student×grade×academic year, subject, grade, day-of-week and region×academic year fixed effects. Standard errors are clustered on municipalities.

Figure 5: Robustness check: Non-Linear Effect Of Temperatures On Students Test Score



Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. The figure displays non linear estimates of the effect of temperatures 20 days before and after the test on students test score. Bins are grouped by 3°C maximum temperature intervals.Temperatures less than 14°C are the benchmark category. The dependent variable is the standardized test score with zero mean and unitary standard deviation. Estimates include controls for rainfall and wind-speed 20 days before or after the test at municipal level, share of female students, share of immigrants, share of early enrolled, share of retained, average class size at school level, along with student, academic year, grade, day-of-week and region×academic year fixed effects. Standard errors are clustered on municipalities.

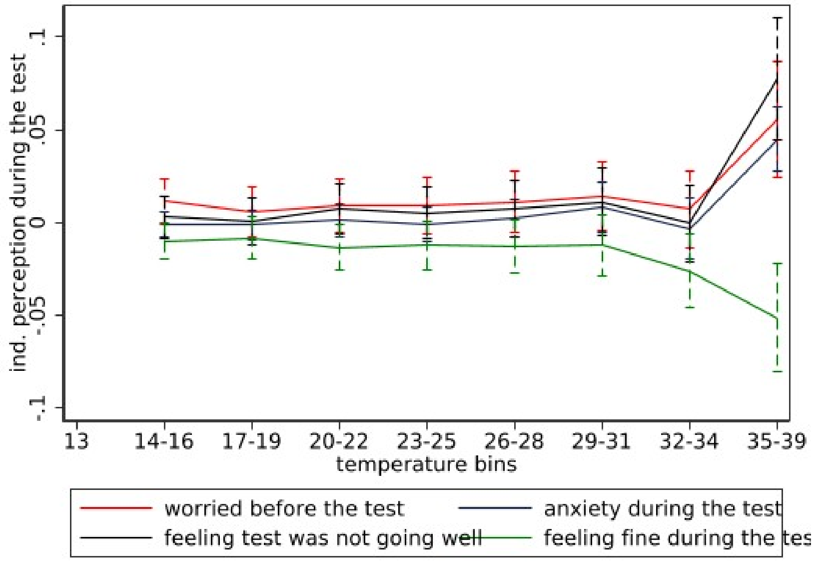
Figure 6: Non-Linear Effect Of Temperatures On Students Test Score Across Grades



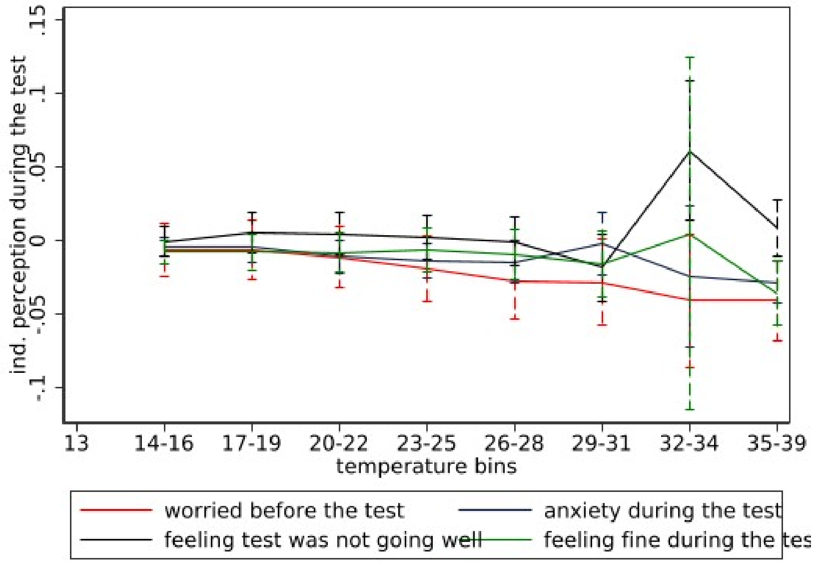
Notes: Academic years from 2011-12 to 2016-17. Non linear estimates of the effect of temperatures the day of the test on students test score (standardized). Bins are grouped by 3°C maximum temperature intervals. Temperatures less than 14°C are the benchmark category. Estimates include controls for rainfall and wind-speed the day of the test at municipal level, dummy variables for father or mother having a tertiary education degree, dummy for non employed status of parents, dummy for foreign citizenship, dummy for being female, dummy for being early enrolled or a retained student and the escs indicator of socioeconomic background. At school level we control for share of female students, share of immigrants, share of early enrolled, share of retained, average class size. We add school, academic year, day-of-week and region×academic year fixed effects. Standard errors are clustered on municipalities.

Figure 7: Non-Linear Effect Of Temperatures On Emotional Outcomes

A. 5^{th} grade



B. 10^{th} grade



Notes: Pooled sample academic years 2011-12 to 2016-17 for grade 5 and from 2011-12 to 2014-15 for grade 10. The figure displays non linear estimates of the effect of temperatures the day of the test on students emotional outcomes. Bins are grouped by 3°C maximum temperature intervals. Temperatures less than 14°C are the benchmark category. The dependent variables are dummies for each emotional perceptions retrieved from students' questionnaire. Estimates include controls for rainfall and wind-speed the day of the test at municipal level, dummy variables for father or mother having a tertiary education degree, dummy for non employed status of parents, dummy for foreign citizenship, dummy for being female, dummy for being early enrolled or a retained student and the escs indicator of socioeconomic background. At school level we control for share of female students, share of immigrants, share of early enrolled, share of retained, average class size. We further include school, academic year, day-of-week and region×academic year fixed effects. Standard errors are clustered on municipalities.

Tables

Table 1: Summary Statistics

	mean	S.D	min.	max.
	Panel A: Language sample			
Test scores:				
Correct answers (%)	68.32	17.72	0	100
Standardized scores	0.103	0.96	-6.51	4.69
Climate conditions:				
Max. temperature (C°)	23.76	4.27	0.6	38
Rainfall (mm/day)	0.98	3.24	0	103
Wind-speed (m/s)	2.50	1.13	0.1	13.1
Obs.		6,201,067		
# of students		3,099,802		
# of schools		31,015		
# of municipalities		6,724		
	Panel B: Math sample			
Test scores:				
Correct answers (%)	59.23	20.12	0	100
Standardized scores	0.08	0.98	-6.05	4.98
Climate conditions:				
Max. temperature (C°)	24.16	4.42	2	38
Rainfall (mm/day)	1.44	4.47	0	103
Wind-speed (m/s)	2.46	1.11	13.1	
Obs.		6,212,649		
# of students		3,017,323		
# of schools		31,045		
# of municipalities		6,725		

Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years 2011-12 to 2016-17. The scores are standardized with 0 mean and unitary standard deviation within grade and academic year.

Table 2: Summary Statistics - Emotional Perceptions During The Test

	mean	S.D	min.	max.
	Panel A: 5th grade			
1(Worried before the test)	0.455	0.498	0	1
1(Anxiety during the test)	0.135	0.341	0	1
1(Feeling test was not going well)	0.366	0.481	0	1
1(Feeling fine during the test)	0.456	0.498	0	1
Obs.	2,580,445			
	Panel B: 10th grade			
1(Worried before the test)	0.259	0.438	0	1
1(Anxiety during the test)	0.101	0.302	0	1
1(Feeling test was not going well)	0.251	0.434	0	1
1(Feeling fine during the test)	0.355	0.478	1	1
Obs.	1,266,951			

Notes: Pooled sample academic years 2011-12 to 2016-17 for grade 5 and 2011-12 to 2014-15 for grade 10. The emotional perceptions are dummy indicators retrieved from self reported answers on a student questionnaire. These variables are available for grades 5 and 10 only.

Table 3: Effect of Hotter Temperatures on Student Performance

	Language		Mathematics	
	(1)	(2)	(3)	(4)
Max. temperature (°C)	-0.00477***		-0.0100***	
	(0.00150)		(0.00192)	
14-16°		-0.0185		-0.0214
		(0.0222)		(0.0228)
17-19°		-0.0288		-0.0621***
		(0.0203)		(0.0224)
20-22°		-0.0620***		-0.0880***
		(0.0227)		(0.0242)
23-25°		-0.0616**		-0.105***
		(0.0251)		(0.0292)
26-28°		-0.0713**		-0.135***
		(0.0289)		(0.0331)
29-31°		-0.0552*		-0.137***
		(0.0293)		(0.0369)
32-34°		-0.100***		-0.185***
		(0.0329)		(0.0386)
35-39°		-0.102***		-0.202***
		(0.0370)		(0.0458)
Constant	0.211***	0.155***	0.482***	0.344***
	(0.0371)	(0.0257)	(0.0485)	(0.0311)
Obs.	6,201,067	6,201,067	6,212,649	6,212,649
R-squared	0.763	0.763	0.760	0.760

Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. Regression of standardized test scores on maximum temperature observed the day of the test at municipal level. Estimates include controls for rainfall and wind-speed the day of the test at municipal level, share of female students, share of immigrants, share of early enrolled, share of retained, average class size at school level, along with student, academic year, grade, day-of-week and region×academic year fixed effects. Max. temperature is in Celsius degree (°C). Column (1) and (3) include the linear effect of maximum temperature the day of the test at municipal level. Columns (2) and (4) include the its non-linear effect by 3-degree bins of temperature. Standard errors, in parentheses, are clustered on municipalities. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Robustness: Excess of Heat

	Language		Mathematics	
	(1)	(2)	(3)	(4)
Max. temperature (°C)	-0.00652***	-0.00534**	-0.0107***	-0.0107***
	(0.00155)	(0.00225)	(0.00204)	(0.00225)
Max. temperature×1(excess>0)	0.000975		0.000445	
	(0.000613)		(0.000561)	
Max. temperature×1(0≥excess<3)		0.000960		0.000428
		(0.000596)		(0.000550)
Max. temperature×1(3≥excess<6)		0.000649		0.000490
		(0.000987)		(0.000879)
Max. temperature×1(6≥excess<9)		-0.000359		0.000230
		(0.000998)		(0.000998)
Max. temperature×1(9≥excess<12)		-0.000436		0.00148
		(0.00279)		(0.00314)
Constant	0.246***	0.219***	0.496***	0.496***
	(0.0438)	(0.0516)	(0.0494)	(0.0531)
Obs.	6,201,067	6,201,067	6,212,649	6,212,649
R-squared	0.763	0.763	0.760	0.760

Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. In columns (1) and (3) regressions of standardized test scores on maximum temperature observed the day of the test at municipal level and maximum temperature interacted with a dummy indicating that temperature the day of the assessment exceed its three years average in the month of the test. In column (2) and (4) regressions of standardized test scores on maximum temperature observed the day of the test at municipal level and maximum temperature interacted with dummies indicating intervals of excess of heat the day of the assessment compared to the three years average temperature in the month of the test. All estimates include controls for rainfall and wind-speed the day of the test at municipal level, share of female students, share of immigrants, share of early enrolled, share of retained, average class size at school level, along with student, academic year, grade, day-of-week and region×academic year fixed effects. Max. temperature and bins of excess of heat are in Celsius degree (°C). Standard errors, in parentheses, are clustered on municipalities. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Robustness: Comparison Across Different Model Specifications

	Language				Mathematics			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Max. temperature	-0.00337***	-0.00950***	-0.00352***	-0.00477***	-0.00478***	-0.0181***	-0.0124***	-0.0100***
	(0.000778)	(0.000917)	(0.00123)	(0.00150)	(0.000742)	(0.00213)	(0.00180)	(0.00192)
Constant	0.234***	0.331***	0.181***	0.211***	0.364***	0.705***	0.539***	0.482***
	(0.0240)	(0.0253)	(0.0316)	(0.0371)	(0.0283)	(0.0508)	(0.0451)	(0.0485)
Obs.	6,200,847	6,201,067	6,201,067	6,201,067	6,212,433	6,212,649	6,212,649	6,212,649
R-squared	0.197	0.761	0.763	0.763	0.224	0.754	0.760	0.760
School×grade FE	x				x			
Student FE		x	x	x		x	x	x
Academic year FE	x	x	x	x	x	x	x	x
Grade FE		x	x	x		x	x	x
Weekday of the test FE	x		x	x	x		x	x
Region×academic year FE	x			x	x			x

Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. Regression of standardized test scores on maximum temperature observed the day of the test at municipal level. Estimates in Columns (1) and (5) we use a specification with school, academic year,day-of-week and region×academic year fixed effects and are based on the same sample of student fixed-effect estimates of Table 3. We further control for rainfall and wind-speed the day of the test at municipal level, dummy variables for father or mother having a tertiary education degree, dummy for non employed status of parents, dummy for foreign citizenship, dummy for being female, dummy for being early enrolled or a retained student and the escs indicator of socioeconomic background. At school level we control for share of female students, share of immigrants, share of early enrolled, share of retained, average class size. Columns (2) and (6) include student, year and grade fixed effects. Columns (3) and (7) include student, year, grade and regionÃyear fixed effects. Columns (4) and (8) include model specification as in Table 3. Dependent variable is in Celsius degree (°C). Standard errors, in parentheses, are clustered on municipalities. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Robustness: Other Specifications

	Panel A. Language			
	(1)	(2)	(3)	(4)
Max. temperature (°C)	-0.00477***	-0.00458***	-0.00337***	-0.00291
	(0.00150)	(0.00167)	(0.00128)	(0.00192)
Constant	0.211***	0.206***	0.173***	0.224***
	(0.0371)	(0.0407)	(0.0320)	(0.0384)
Obs.	6,201,067	6,201,067	4,725,589	6,201,067
R-squared	0.763	0.763	0.764	0.763
	Panel B. Mathematics			
	(1)	(2)	(3)	(4)
Max. temperature (°C)	-0.0100***	-0.00725***	-0.00633***	-0.0105***
	(0.00192)	(0.00225)	(0.00160)	(0.00221)
Constant	0.482***	0.412***	0.362***	0.538***
	(0.0485)	(0.0574)	(0.0399)	(0.0518)
Observations	6,212,649	6,212,649	4,420,859	6,212,649
R-squared	0.760	0.760	0.761	0.760
Baseline	x			
Baseline + grade×academic year		x		
Non-movers only			x	
Control for max. T previous day				x

Notes: Pooled sample of grades 2, 5, 8 and 10 in academic years from 2011-12 to 2016-17. Regression of standardized test scores on maximum temperature observed the day of the test at municipal level. Estimates include controls for rainfall and wind-speed the day of the test at municipal level, share of female students, share of immigrants, share of early enrolled, share of retained, average class size at school level, along with student, academic year, grade, day-of-week and region×academic year fixed effects. Max. temperature is in Celsius degree (°C). Column (1) is the baseline specification. In column (2) the specification is augmented with the Grade×academic year fixed effect. Column (3) is the baseline specification in a sample of students that do not change municipality between tests (non movers). In column (4) we control for previous day maximum temperature. Standard errors, in parentheses, are clustered on municipalities. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.