The Unequal Burden of Heat on Learning*

Gabriel Cruz
University of Maryland

Marcos Fabian
University of Maryland

June 19, 2026

Abstract

Heat reduces student learning through two channels, chronic exposure over the school year and acute exposure on exam day. Which channel dominates varies systematically with local climate and student vulnerability. We document this using a nationwide panel of sixty million test scores from eighteen million Mexican students linked to daily weather data. Chronic exposure drives losses in warm, humid, and disadvantaged settings; acute exposure dominates in temperate, dry, and advantaged ones. A stylized framework of learning accumulation and erosion rationalizes these patterns. A conditional cash transfer attenuates the chronic effect, pointing to social protection programs as potential climate adaptation tools.

*G. Cruz, University of Maryland, Economics Department: email gjcruz@umd.edu; M. Fabian, University of Maryland, School of Public Policy: email mfbnco@umd.edu. We would like to thank Daniel Agness, Judith Hellerstein, Susan Parker, Nolan Pope and Sergio Urzua for their useful comments and workshop participants at RIDGE (Environmental Economics, Montevideo) and CESifo (Junior Workshop on the Economics of Education, Munich).

1 Introduction

A growing body of research shows that heat affects students' performance on standardized exams (Venegas Marin et al., 2024). The literature has identified two distinct channels of exposure: chronic and acute. On the one hand, chronic exposure accumulates across the academic year and has been associated with mechanisms that include school attendance, less ability to concentrate during class, child labor, and vector-borne disease incidence (Garg et al., 2020; Arceo-Gomez and López-Feldman, 2024). On the other hand, acute exposure occurs on the date of the exam and can affect performance primarily through transitory cognitive and physiological impairment (Vu, 2022; Park, 2022; Zhang et al., 2024).

This literature often studies the two channels separately and treats their relative importance as homogeneous across student populations. Yet that importance is unlikely to be homogeneous, since individual and local characteristics, including historical climate, plausibly determine which channel dominates. For instance, students in historically warm regions face higher cumulative heat exposure across the school year, while students in temperate regions experience harmful heat only as occasional shocks. Differences in physiological adaptation, infrastructure investment, and household coping plausibly affect how each setting balances the two channels.[1] Similarly, a second source of expected heterogeneity is the differential vulnerability of student populations. Those with poorer school and home infrastructure, fewer compensatory household inputs, and lower physiological capacity to absorb heat may differ in their response to chronic and acute exposures.

Testing both exposure channels jointly across heterogeneous geography and degrees of vulnerability requires large-scale individual-level test-score data with enough variation in climate and student characteristics, which has rarely been available. We fill this gap with a nationwide administrative dataset from Mexico that follows more than eighteen million students in grades 3–9 from 2007 to 2013, with more than sixty million standardized test scores linked to daily municipal-level weather data. We estimate

[1] See, e.g., Wargocki et al. (2019); Yeganeh et al. (2018) on the temperature thresholds at which classroom cognition deteriorates; Park et al. (2020, 2021) on the role of adaptation in mitigating cumulative heat damage.

empirical models that exploit within-student variation in temperature exposure, controlling for municipality, year, grade effects, and state-specific time trends.

We document four main insights. First, we show that heat reduces test scores through both chronic and acute channels, and that which channel dominates depends at least on two dimensions: climate region and student vulnerability in the baseline. Second, we show that in warm and humid municipalities the cumulative academic-year channel drives most of the negative effect, with little role for exam-day temperature, whereas in temperate and dry municipalities exam-day temperature drives the effect with little role for cumulative exposure. Third, we document that the chronic effect is larger in highly-marginalized and rural municipalities and for elementary-school students, while the acute effect is larger in less vulnerable students. Finally, we exploit student-level data on Progresa receipt, a conditional cash transfer, and find that, consistent with the conceptual framework, beneficiaries show an attenuated cumulative-heat effect.

We rationalize this heterogeneity with a simple stylized framework that considers a student's exam score as a function of the learning accumulated over the school year. Chronic heat erodes that learning day by day, and a student's protective capacity (the household, community, and individual resources that protect against heat) offsets part of the loss. Acute heat on the exam day then reduces the score the student can produce from that accumulated learning. The relative importance of the two channels depends on how much learning a student brings to the exam. A student who arrives with little accumulated learning, whether because they live in a hot region or because they have weaker protective capacity, has little left for exam-day heat to take away, so most of the damage comes from the learning lost over the year.[2] A student who arrives with abundant learning has more to lose on the exam day itself, so exam-day heat matters more. The framework turns this logic into testable implications across climate regions and student vulnerability, which we take to the data.

This paper contributes to the literature in three ways. First, we identify the chronic

[2] This is consistent with diminishing returns to accumulated schooling, where an additional unit of instruction raises achievement by more where prior schooling is scarce (Agüero and Beleche, 2013; Kraft and Novicoff, 2024)

and acute channels jointly within a student panel that covers the full student population. Existing causal estimates typically address one channel at a time, on selected populations or single regions.[3] Identifying both channels is important because, on the one hand, estimating the chronic channel alone can overstate cumulative learning losses. Indeed, if the exam date is hot, we would attribute the exam-day stress to the more permanent learning losses. On the other hand, the distinction is policy relevant. Chronic damage calls for year-round protection, such as infrastructure, social protection, or attendance support, while acute damage points to the testing environment. Knowing which channel dominates and where suggests which policy response is relevant.

Second, we show that the relative importance of the two channels varies systematically with baseline climate and student vulnerability. That heat effects vary across these dimensions is documented in the climate-economy literature (Burke et al., 2015; Dell et al., 2014; Venegas Marin et al., 2024); we show that a unified framework of learning accumulation and erosion accounts for the pattern across both channels. Third, we provide evidence that social protection can attenuate the chronic effect of heat. Linking our test-score database with an individual-level registry of Mexico's nationwide conditional cash transfer Progresa, we show that beneficiaries experience a weaker chronic effect. By design, the program supplements poor households' income and conditions transfers on school attendance, two features that plausibly raise protective capacity through compensatory household investments and reduced heat-induced absenteeism. This finding moves the adaptation literature beyond physical infrastructure such as air conditioning and identifies social protection as an instrument of climate adaptation that middle-income countries already run at scale.

The remainder of the paper is organized as follows. Section 2 describes the data sources and institutional background. Section 3 presents the conceptual framework and derives the testable implications. Section 4 describes the empirical approach. Section 5 presents the results. Section 6 concludes.

[3] For instance, Park et al. (2020) estimate the cumulative effect of school-year heat on learning among exam retakers, netting out exam-day temperature as a confound, while Park (2022) estimates the effect of exam-day heat on performance in a single school district. Neither estimates the two channels jointly.

2 Data and Background

2.1 Student Achievement Data

Our learning outcome data come from the ENLACE[4] test, a census-based standardized test administered annually by the SEP[5] between 2006 and 2013 to assess student achievement at the end of each academic year. It evaluated all students in grades 3–6 (primary) and 7–9 (lower secondary) in mathematics, Spanish, and a rotating third subject.[6] The test was low-stakes for students: scores had no bearing on grades, promotion, or school placement (de Hoyos et al., 2021; Behrman et al., 2025).

The test was administered nationally once per academic year, over three consecutive days for each level — primary schools first, followed by secondary schools one week later.[7]

We construct a student-level panel covering the six academic years 2007–08 through 2012–13 by linking individual test records across years using anonymized student identifiers derived from civil registry records, following de Hoyos et al. (2021). Table C.1 documents six cohorts that can be tracked longitudinally as they progress from primary through lower secondary school, with follow-up ranging from four to six consecutive years depending on the cohort's entry grade. Our sample comprises approximately 64 million student-year observations from over 18 million unique students across roughly 2,300 municipalities.

Because test forms and scaling procedures differ across grades and years, raw scores are not comparable across these dimensions. We standardize each subject's score within grade–year cells to have mean zero and standard deviation one, and average the two standardized scores to construct our primary outcome variable.[8] All results are therefore expressed in student-level standard deviation units. We leverage this large student-level panel by linking it with high-resolution daily weather data at the municipality

[4]Evaluación Nacional del Logro Académico en Centros Escolares

[5]Secretaría de Educación Pública

[6]We use only mathematics and Spanish, the two subjects tested in every year. A third subject rotated annually (e.g., natural sciences, history, civic education).

[7]See Appendix 1.2 for the test calendar of each year.

[8]Raw scores are reported on a 200–800 IRT scale with approximately 50 items per subject.

level, which we describe next.

2.2 Weather Data

Our weather data come from two gridded products developed by the Climate Hazards Center at UC Santa Barbara. For temperature, we use CHIRTS-ERA5 v3, which provides daily maximum and minimum temperature at 0.05° resolution (∼5.5 km) covering our period of interest. CHIRTS-ERA5 combines monthly temperature rasters from CHIRTS, which combines satellite and station data to perform well even where weather stations are sparse, with daily variation from the ERA5 reanalysis.[9] For precipitation, we use CHIRPS v3.0, which provides daily rainfall at the same spatial resolution by blending satellite imagery with in-situ station data. Both products share the same grid geometry, allowing us to construct a unified daily climate panel at the municipality level.

We aggregate daily grid-cell values to the municipality level using area-weighted averages, following the standard approach in this literature (Barreca et al., 2016). Each raster cell is assigned to the municipalities it intersects, with weights proportional to the share of the cell's area falling within each municipal boundary. The weight matrix is computed once from municipal boundaries and applied uniformly across all days and years.[10]

We construct two measures of heat exposure corresponding to the two channels identified in the literature. Our first measure captures chronic exposure: the average daily maximum temperature experienced by students over the academic year (Park et al., 2020), from August through the day the standardized nationwide test was administered each year, excluding winter and spring breaks. We use daily maximum temperature on the assumption that schooling takes place during daytime hours, when temperatures peak. Our second measure captures acute exposure: the average daily maximum temperature during the days of ENLACE administration. Exam-day temperatures are substantially higher than academic-year averages because the test is administered in late spring (April–June), when temperatures peak in much of Mexico.

[9] Details at `https://chc.ucsb.edu/data`.

[10] Details appear in Appendix 1.1

Exploiting Mexico's geographic diversity, we classify municipalities as warm or temperate and as dry or humid using INEGI's Köppen–García climatological normals (see Figure D.1).[11] Roughly 40 percent of dry municipalities are temperate and 60 percent are warm, so the two classifications do not coincide and temperature and moisture capture distinct sources of climatic variation. Panel A of Table C.2 reports mean values that illustrate the geographic variation underlying our analysis. Warm municipalities average 29.2°C and experience 24.8 school days above 34°C per year, compared to 25.2°C and 1.6 such days in temperate areas.[12]

Figure D.2 illustrates the geographic and temporal variation in heat exposure by mapping the number of school days above 28°C for four academic years spanning the sample period. Finally, we control for precipitation throughout our main specifications, measured as mean daily rainfall (mm/day) over the same academic-year window. Precipitation data are constructed from the same grid geometry and aggregation procedure as temperature for consistency across weather variables.

2.3 Conditional Cash Transfer: Progresa Program

Progresa was Mexico's flagship conditional cash transfer program, providing income support to poor households conditional on children's regular school attendance. Launched in the late 1990s, the program had reached national scale well before our study period (Parker and Todd, 2017). Importantly, family enrollment did not occur through voluntary application. Instead, eligibility was determined through a combination of locality-level poverty maps and detailed household socioeconomic censuses. Potentially eligible households were identified administratively and contacted directly by program personnel. Families who accepted the program's rules were incorporated into the beneficiary registry and subject to periodic re-certification of eligibility, but there was no open sign-up or competitive selection process among the poor. Although the program operated nationwide, beneficiary status varied within schools, where more than half of the schools in our estimation sample contain both recipients and non-recipients.

[11] Available at `https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463769361`. The classification procedure is described in Appendix 1.1.

[12] In Supplemental Appendix B we analyze nonlinear effects, accounting for school days per municipality-year in six temperature bins (<18, 18–22, 22–26, 26–30, 30–34, ≥34°C).

Our individual-level panel identifies students in grades 6 through 9 as Progresa beneficiaries through administrative records, allowing us to distinguish beneficiaries from non-beneficiaries (De Hoyos et al., 2018). We restrict the conditional cash transfer analysis to these grades, where beneficiary status is directly observed. In summary, our dataset combines the individual-level panel structure needed to control for unobserved student heterogeneity with the geographic and socioeconomic variation needed to test how chronic and acute heat exposures interact with climate region and student vulnerability.

3 Conceptual Framework

Our stylized framework is motivated by the climate-economy and human-capital literatures. From the climate-economy literature, Burke et al. (2015) show that the marginal effect of temperature on aggregate economic output depends on local historical climate conditions, while Dell et al. (2014) argue that a bundle of household, school, and community resources mediates the effect of heat. From the human-capital literature, we draw on the idea that human capital is built from learning that accumulates over time (Ben-Porath, 1967; Todd and Wolpin, 2003). We combine these insights by modeling a student's exam score as a function of accumulated learning over the academic year. However, while heat can erode the learning, the resources protecting the student can offset part of the loss. The framework separates two channels, *chronic* heat accumulated across the school year and *acute* heat on the exam day, and characterizes how their relative importance varies across climate regions and student vulnerability.

Setup. Consider a student i who attends school in municipality m during academic year y. The school year comprises D instructional days ($d = 1, \ldots, D$), at the end of which the student takes a standardized exam. Let T_{md} denote the maximum temperature on school day d, and let T_{my}^{exam} denote the maximum temperature on the day of the exam. We summarize chronic exposure by the academic-year mean temperature $T_{my} \equiv \frac{1}{D}\sum_{d=1}^{D} T_{md}$ and acute exposure by T_{my}^{exam}.

Learning during the school year. Each instructional day delivers up to one unit of learning. Heat erodes that day's learning in proportion to the day's temperature, so the learning lost on school day d is $r\,T_{md}$, with $r > 0$.[13]

Protective capacity. Students differ in their capacity to offset heat-induced learning losses using a bundle of household, school, and community resources. We summarize this with a *protective capacity*[14] $\kappa_i \in [0, 1]$. The student offsets a fraction κ_i of each day's heat-induced learning loss, so realized learning on day d is $1 - (1 - \kappa_i)\,r\,T_{md}$. A student with $\kappa_i = 1$ fully offsets the day's heat loss; a student with $\kappa_i = 0$ absorbs it in full.

Learning accumulated over the year. The student reaches the exam with the learning accumulated over the year,[15]

$$L_i = \sum_{d=1}^{D} \Big[1 - (1 - \kappa_i)\,r\,T_{md} \Big] = D - (1 - \kappa_i)\,r\,D\,T_{my} \qquad (1)$$
$$= D\Big[1 - (1 - \kappa_i)\,r\,T_{my}\Big]$$

where D is the scale of learning and $1 - (1 - \kappa_i)\,r\,T_{my}$ is the effective fraction of learning preserved against the year's heat. Because the daily erosion is linear in temperature, the cumulative loss depends on the year's temperatures only through their mean. The academic-year average T_{my} is therefore the relevant measure of chronic exposure. Learning accumulated by the exam is lower in hotter municipalities (higher T_{my}) and for students with weaker protective capacity (smaller κ_i).

[13]High temperatures reduce the learning drawn from a school day through mechanisms documented in the literature, including impaired concentration, reduced attendance, child labor, and vector-borne disease (Cho, 2017; Garg et al., 2020; Park et al., 2020).

[14]Dell et al. (2014) refer to these characteristics as "adaptive capacity"; we use "protective capacity" because we find the term less misleading for our framework. In practice it includes household resources (e.g. compensatory investments), school resources (infrastructure, e.g. AC), community conditions (healthcare; diseases environment), and the student's own physiological tolerance to heat.

[15]This cumulative formulation follows the human-capital and education-production-function tradition (Ben-Porath, 1967; Todd and Wolpin, 2003).

Human capital and the test score. An increasing and concave function f turns accumulated learning into human capital, reflecting diminishing returns of learning or schooling to human capital.[16]

$$s_i = f(L_i)\left(1 - a\,T_{my}^{\text{exam}}\right), \qquad f' > 0, \quad f'' < 0, \tag{2}$$

Heat on the exam day can destroy a share of the score by a fraction $a\,T_{my}^{\text{exam}} \in [0,1)$, with $a > 0$. The multiplicative form says exam-day heat can only draw on the score the student's human capital produces.[17]

The chronic and acute channels. Taking partial derivatives of equation 2 with respect to T_{my} and T_{my}^{exam}, the chronic effect is given by:

$$|\beta_C| = \frac{\partial s_i}{\partial T_{my}} = -\,f'(L_i)\,(1-\kappa_i)\,rD\left(1 - a\,T_{my}^{\text{exam}}\right), \tag{3}$$

and the acute effect is

$$|\beta_A| = \frac{\partial s_i}{\partial T_{my}^{\text{exam}}} = -\,a\,f(L_i). \tag{4}$$

The chronic effect scales with the marginal return $f'(L_i)$, which concavity makes larger when accumulated learning is scarce. The acute effect scales with the level $f(L_i)$, larger when accumulated learning is abundant. In simpler words, a transitory exam-day shock therefore costs more to students with more to lose when the exam date arrives, while academic-year heat costs more to the most exposed and vulnerable students during the year.

[16] As is common in the human-capital and education literature (Ben-Porath, 1967; Kraft and Novicoff, 2024). In the Mexican-ENLACE context, Agüero and Beleche (2013) find test scores to be concave in the days of schooling completed before the exam, which directly reinforces this assumption.

[17] Following the insights in Graff Zivin et al. (2018), exam-day heat acts through a transitory physiological/cognitive impairment that affects access to knowledge already acquired. Students have no less human capital on hot days, but they struggle to access that knowledge, which is reflected in the score.

3.1 Testable Implications

Implication 1. *Chronic exposure effects dominates in warm regions, acute in temperate regions.*[18] Compare a warm municipality with a cool one, where their average academic year temperatures are $T_{\text{warm}} > T_{\text{cool}}$. Fixed everything else equal, a warm municipality erodes more learning during the academic year, so $L_{\text{warm}} < L_{\text{cool}}$. Therefore, by concavity, students in warm municipalities face the larger chronic effect, since they are more exposed throughout the year, but the smaller acute effect, since they reach the exam with less learning for exam-day heat to take away; students in temperate areas face the reverse.[19]

Implication 2. *Chronic exposure effects dominate for vulnerable students, acute for advantaged students.* Compare a vulnerable student (low κ) with an advantaged one (high κ) in the same municipality. Holding everything else equal, the vulnerable student offsets less of the year's heat, so they absorb a larger share of each day's erosion (higher $1 - \kappa$) and reach the exam with less learning ($L_{\text{vuln}} < L_{\text{adv}}$). Their chronic effect is larger directly, through the larger uncompensated erosion, and is reinforced by the higher marginal return $f'(L)$ that comes with less accumulated learning. Their acute effect is smaller, since arriving with less learning leaves exam-day heat less to take away. Advantaged students face the reverse.

Implication 3 (raising protective capacity). An intervention that raises a student's protective capacity κ_i raises L_i and lowers the uncompensated erosion $(1 - \kappa_i)$. Both attenuate the chronic effect $|\beta_C|$.

[18] The same comparison applies to the dry/humid dimension since effective heat depends on both temperature and humidity. Humid air retains heat and cools more slowly, so at a given temperature humid municipalities experience higher effective heat over the school year (Agarwal et al., 2025).

[19] This implication holds when $f'(L_{\text{warm}})/f'(L_{\text{cool}}) > \big(1 - a\,T^{\text{exam}}_{\text{cool}}\big)/\big(1 - a\,T^{\text{exam}}_{\text{warm}}\big)$. That is, when the gain in marginal return from a more depleted stock outweighs the larger exam-day score loss in warm regions. Two mild conditions are enough. First, f is sufficiently concave, supported by evidence that the marginal return to instructional time falls toward zero where schooling is already high (Kraft and Novicoff, 2024), and by Mexican ENLACE estimates where the gain from an extra school day falls to zero as days accumulate (Agüero and Beleche, 2013). Second, exam-day heat does not destroy a large share of the score, plausible because a single day's heat is unlikely to generate extreme losses, as our estimates also suggest. Read in reverse, the chronic effect being larger in warm regions, as we find, is consistent with both conditions.

4 Empirical Approach

To test the framework's implications, we estimate the chronic and acute channels jointly in a student fixed-effects regression. The repeated testing of students across years allows us to control for unobserved time-invariant individual characteristics. Identification comes from year-to-year variation in temperature within municipalities, net of state-specific trends, which provides quasi-random changes in exposure across grade levels, geography, and time. Specifically, we estimate:

$$test_{ismgy} = \beta_C T_{my} + \beta_A T_{my}^{\text{exam}} + \delta P_{my} + \eta_i + \rho_m + \gamma_y + \phi_g + \tau_{sy} + \epsilon_{ismgy} \tag{5}$$

where $test_{ismgy}$ denotes the standardized test score of student i in school s, municipality m, grade g, and academic year y.[20] T_{my} is the mean daily maximum temperature in municipality m during academic year y, our measure of chronic exposure, and T_{my}^{exam} is the maximum temperature on the day of the exam, our measure of acute exposure. P_{my} is the corresponding mean daily precipitation. The term η_i captures student fixed effects, ρ_m municipality fixed effects, γ_y academic-year fixed effects, ϕ_g grade fixed effects, and τ_{sy} state-specific linear time trends. Standard errors are clustered at the municipality level. The coefficients β_C and β_A are the empirical counterparts of the chronic and acute effects in the conceptual framework (equations 3 and 4).

With student fixed effects, identification relies on comparisons of the same student across years as heat exposure varies, conditional on municipality, year, and grade effects. The state-specific trends absorb differential secular changes in test performance across Mexican states, ensuring that the identifying variation comes from deviations in local temperature from the trajectory predicted by the student's municipality and state trend. Under this specification, β_C and β_A each measure the change in a student's test score (in standard-deviation units) associated with a 1°C increase in academic-year and exam-day maximum temperature, respectively.[21]

We also estimate a binned specification that replaces the continuous temperature

[20] In the main specification we use the average standardized score between Spanish and Mathematics as the dependent variable.

[21] A 1°C increase is equivalent to a 1.8°F increase.

measure with counts of days in temperature ranges, reported in Appendix B. However, we prefer the linear specification throughout, since our interest is the relative importance of the chronic and acute channels, which a single coefficient per channel makes directly comparable. A binned specification fits the cumulative academic-year channel but not the acute one, which falls on a single exam day. We still account for the convexity of temperature effects by estimating the model separately by climate region, which lets the temperature slope differ across regions.

5 Effects of Heat on Learning

5.1 Heat and Test Scores across Climate Zones

We begin by estimating the chronic channel in isolation. Panel A of Table 1 regresses test scores on academic-year heat alone, without controlling for exam-day temperature. In the full sample, a 1°C increase in the average daily maximum temperature during the school year is associated with a 0.019 standard deviation reduction in test scores. Columns (2)–(5) estimate this across four climate groups. The results vary across climate zones: it is largest in warm municipalities (-0.023) and humid municipalities (-0.021), smaller in temperate municipalities (-0.012), and smallest in dry municipalities (-0.009).

Panel B decomposes the temperature effect into chronic academic-year exposure and acute exam-day exposure. In the full sample, the chronic coefficient is -0.013 and the acute coefficient is -0.004, both statistically significant, with chronic exposure accounting for roughly three-quarters of the total effect. This decomposition shows that the two channels are empirically separable, each potentially operating through a different channel: cumulative academic-year heat erodes accumulated human capital over the school year, while exam-day heat causes a transitory cognitive and physiological impairment that reduces the score the student can produce from that stock.

The relative importance of chronic and acute temperature exposure varies across climate types. In warm and humid municipalities, the chronic channel dominates, while the acute effect is small and statistically insignificant. This pattern is consistent with

Table 1: Effects of Temperature on Test Scores by Climate Type

	(1) All Sample	(2) Warm	(3) Temperate	(4) Humid	(5) Dry
Panel A: Avg Max Temp (°C)					
Full Academic Year	-0.0189***	-0.0234***	-0.0121**	-0.0210***	-0.0096**
	(0.0038)	(0.0051)	(0.0049)	(0.0054)	(0.0041)
Panel B: Exam Day Temp (°C)					
Full Academic Year	-0.0133***	-0.0230***	-0.0020	-0.0197***	-0.0031
	(0.0044)	(0.0062)	(0.0046)	(0.0074)	(0.0044)
Exam day	-0.0036***	-0.0003	-0.0059***	-0.0008	-0.0060***
	(0.0013)	(0.0020)	(0.0017)	(0.0023)	(0.0016)
Observations	64,005,332	33,576,955	30,251,695	41,240,101	22,513,653
Mean Temp	26.9	29.1	24.4	27.3	26.2
SD	0.34	0.37	0.28	0.30	0.37

Notes: The dependent variable is the average standardized score between Mathematics and Spanish. Panel A reports estimates from a restricted version of equation (5) that includes only the average daily maximum temperature during the full academic year, without controlling for exam-day temperature. Panel B adds the maximum temperature on the exam day as a separate regressor, corresponding to the full equation (5). Columns 2–5 estimate the model separately by climate type: warm vs. temperate and humid vs. dry. All specifications include student, municipality, year, and grade fixed effects, state-specific linear time trends, and control for mean daily precipitation. Standard errors clustered at the municipality level in parentheses. Mean and SD report the mean and standard deviation of the temperature variable after absorbing all fixed effects.

Implication 1 of the conceptual framework: sustained heat throughout the school year erodes accumulated human capital, so students arrive at the exam with a reduced stock, implying a larger chronic marginal effect and a smaller acute one. In temperate and dry municipalities, the pattern reverses.

The bottom rows of Table 1 report the mean and residual standard deviation of the temperature variable after absorbing all fixed effects. Mean academic-year temperature ranges from 24.4°C in temperate municipalities to 29.1°C in warm ones, but the identifying variation is similar across climate zones—roughly 0.3°C—indicating that the differences in estimated effects are not driven by differences in the amount of temperature variation available for identification.

To further show how baseline heat relates to the two channels, we rank municipalities into more granular tiers by their historical heat exposure and re-estimate the chronic/acute decomposition within each tier. As expected, the chronic effect becomes larger and the acute effect smaller as historical exposure rises (Appendix Figure D.3).[22]

The Supplemental Appendix reports several additional checks. Table C.4 presents the chronic and acute estimates across specifications with incrementally richer fixed effects. The chronic coefficient is stable once municipality fixed effects are included, and the acute coefficient attenuates once state-by-year fixed effects enter but remains statistically significant. Adding air pollution and humidity as controls leaves these estimates unchanged. Replacing temperature with the Heat Index, which accounts for temperature and humidity jointly, yields qualitatively similar results, though larger in magnitude (Table C.3). Finally, future and past academic-year temperatures do not predict current test scores, which rules out confounding local trends (Table C.5).

5.2 Heterogeneous Effects across Vulnerability Dimensions

The conceptual framework links both the chronic and acute channels to two sources of heterogeneity that compound. Within a climate region, the chronic effect is larger and the acute effect smaller for more vulnerable students, with the reverse for less vulnerable ones. The two dimensions reinforce each other, so vulnerable students in warm municipalities face the largest chronic effect and advantaged students in temperate municipalities the largest acute effect. We test these patterns along three dimensions: municipality-level socioeconomic marginalization, rural-urban classification, and grade cycle (primary vs. lower secondary).

5.2.1 Municipality characteristics

We use two municipality-level proxies for protective capacity: socioeconomic marginalization and rural-urban status. Table 2 reports estimates by marginalization (Panel A) and rural-urban status (Panel B). Two patterns emerge, both consistent with the conceptual framework. First, the chronic effect is concentrated in more disadvantaged municipalities. In high-marginalization municipalities, the academic-year temperature

[22]The tiers are based on the Heat Index, which combines temperature and relative humidity into a measure of perceived heat stress (Steadman, 1979; Rothfusz and Headquarters, 1990). We present the main results by four climate regions, for simplicity of exposition and for statistical power. Broader groups retain more observations for the finer heterogeneity analysis in the next section.

coefficient is -0.045 and statistically significant, whereas in low-marginalization municipalities it is not statistically different from zero. The magnitude is even larger in warm and humid climates, where high-marginalization municipalities experience test-score declines exceeding -0.06 standard deviations, consistent with Implications 1 and 2.

Second, rural municipalities are considerably more affected than urban ones.[23] The academic-year coefficient is -0.031 standard deviations in rural areas, compared with no significant effect in urban areas. In urban municipalities, the effect operates primarily through exam-day temperature, consistent with Implication 2: urban students, with higher protective capacity from better infrastructure, are partially shielded from chronic exposure but retain a higher potential score that acute heat can destroy.

5.2.2 Grade cycle

A third dimension of protective capacity is the student's age and school cycle. Younger children plausibly have lower individual protective capacity—less physiological resilience to heat and greater dependence on school-based learning—placing them in the vulnerable group in the framework. Our data allow us to split the sample into elementary (grades 3–6) and middle school (grades 7–9).

Table 3 shows that the chronic effect is concentrated among elementary school students, where a 1°C increase reduces scores by 0.012 standard deviations in the full sample, with larger effects in warm and humid municipalities, consistent with Implication 2. For middle school students, the chronic coefficient is small and not statistically significant in most specifications.[24] The acute effect follows the opposite pattern: middle school students exhibit a large and significant negative response (-0.010 in the full sample), roughly three times the magnitude found for elementary students. This con-

[23] Rural communities are not limited to a single climate zone: the share of rural population ranges from 18.2% in dry municipalities to 27.8% in warm ones (Table C.2, Panel B).

[24] In temperate and humid municipalities, the academic-year coefficient for middle school is positive and statistically significant, a pattern that could reflect selective attrition if weaker students are more likely to drop out in hotter years. This result is also consistent with Arceo-Gomez and López-Feldman (2024), who find that heat can have a positive effect on academic performance in colder areas in Mexico, and with the fact that temperate municipalities in our sample do not experience significant extreme heat events (Panel A of Table C.2).

Table 2: Heterogeneous Effects of Temperature on Test Scores by Municipality Characteristics

	(1) All Sample	(2) Warm	(3) Temperate	(4) Humid	(5) Dry
Panel A: Area-level Socioeconomic Marginalization Index (higher = more disadvantaged)					
Low					
Avg Max Temp (°C)	-0.0055	-0.0157**	0.0062	-0.0077	-0.0039
	(0.0046)	(0.0071)	(0.0070)	(0.0098)	(0.0045)
Exam Day Temp (°C)	-0.0062***	-0.0013	-0.0080***	-0.0052*	-0.0066***
	(0.0013)	(0.0024)	(0.0023)	(0.0027)	(0.0016)
Observations	45,525,277	21,770,295	23,653,340	26,516,723	18,820,018
Medium					
Avg Max Temp (°C)	-0.0164**	-0.0270***	0.0049	-0.0143	-0.0046
	(0.0065)	(0.0088)	(0.0092)	(0.0087)	(0.0090)
Exam Day Temp (°C)	0.0004	0.0025	-0.0074**	-0.0003	0.0011
	(0.0023)	(0.0029)	(0.0034)	(0.0028)	(0.0034)
Observations	11,851,010	6,994,395	4,850,711	8,487,487	3,357,903
High					
Avg Max Temp (°C)	-0.0448***	-0.0608***	-0.0122	-0.0459***	0.0327
	(0.0131)	(0.0166)	(0.0182)	(0.0137)	(0.0210)
Exam Day Temp (°C)	0.0054	0.0086	0.0006	0.0054	-0.0033
	(0.0045)	(0.0062)	(0.0043)	(0.0046)	(0.0072)
Observations	6,331,580	4,657,653	1,669,360	6,030,622	300,589
Panel B: Rural and Urban					
Rural					
Avg Max Temp (°C)	-0.0314***	-0.0385***	-0.0056	-0.0284***	-0.0031
	(0.0079)	(0.0108)	(0.0104)	(0.0101)	(0.0088)
Exam Day Temp (°C)	0.0036	0.0045	-0.0018	0.0026	0.0012
	(0.0025)	(0.0033)	(0.0030)	(0.0031)	(0.0030)
Observations	11,950,125	7,579,866	4,362,054	9,310,672	2,635,264
Urban					
Avg Max Temp (°C)	-0.0067	-0.0162***	0.0015	-0.0084	-0.0041
	(0.0042)	(0.0062)	(0.0053)	(0.0081)	(0.0044)
Exam Day Temp (°C)	-0.0060***	-0.0027	-0.0066***	-0.0055**	-0.0062***
	(0.0013)	(0.0021)	(0.0018)	(0.0024)	(0.0016)
Observations	51,836,051	25,878,490	25,831,162	31,783,441	19,844,986

Notes: The dependent variable is the average standardized score between Mathematics and Spanish. Panel A reports estimates by area-level socioeconomic marginalization (CONAPO index): Low (very low marginalization), Medium (low and medium marginalization), and High (high and very high marginalization). Panel B reports estimates separately for rural and urban municipalities. Within each panel, the first row reports the coefficient on the average daily maximum temperature during the full academic year and the second row reports the coefficient on the maximum temperature on the exam day. Columns 2–5 estimate the model separately by climate type: warm vs. temperate and humid vs. dry. All specifications follow equation (5) and include student, municipality, year, and grade fixed effects, state-specific linear time trends, and control for mean daily precipitation. Standard errors clustered at the municipality level in parentheses.

trast aligns with Implication 2: elementary students, with lower κ, absorb more of the year's heat erosion and are more affected by the chronic channel, while middle school students retain more human capital at the exam that acute heat can destroy.

Table 3: Heterogeneous Effects by School Level

	(1) All Sample	(2) Warm	(3) Temperate	(4) Humid	(5) Dry
Panel A: By School Level					
Elementary School					
Avg Max Temp (°C)	-0.0116**	-0.0214***	0.0046	-0.0241***	0.0005
	(0.0046)	(0.0063)	(0.0046)	(0.0056)	(0.0047)
Exam Day Temp (°C)	-0.0037**	-0.0005	-0.0067***	0.0003	-0.0049***
	(0.0015)	(0.0027)	(0.0010)	(0.0018)	(0.0017)
Observations	38,459,645	19,967,850	18,373,500	24,619,755	13,671,830
Middle School					
Avg Max Temp (°C)	0.0146	-0.0112	0.0251**	0.0265**	0.0078
	(0.0075)	(0.0095)	(0.0119)	(0.0118)	(0.0079)
Exam Day Temp (°C)	-0.0103***	-0.0020	-0.0130***	-0.0125***	-0.0125***
	(0.0026)	(0.0036)	(0.0045)	(0.0034)	(0.0037)
Observations	20,762,465	10,992,158	9,711,599	13,468,010	7,212,321

Notes: Each column reports estimates from the same specification as Table 2. The sample is split by school level: elementary school (grades 3–6) and middle school (grades 7–9). Standard errors clustered at the municipality level in parentheses.

In summary, the heterogeneity results across climate zones, municipality characteristics, and school/grade level align with the conceptual framework: the chronic channel is strongest where protective capacity is low, while the acute channel is strongest where protective capacity is stronger. Implication 3 suggests that interventions that raise protective capacity should attenuate the chronic channel. We test this directly using individual-level data on conditional cash transfer receipt.

5.3 Can Social Protection Attenuate the Chronic Effect?

Implication 3 of the conceptual framework implies that raising a student's protective capacity κ should attenuate the chronic effect of heat, since higher κ means less uncompensated erosion over the school year and, by the concavity of f, a smaller marginal cost of each additional degree of heat. We examine this using individual-level data on receipt of the Progresa conditional cash transfer, which provides income to poor households conditional on children's school attendance. Progresa can attenuate the chronic effect by expanding household resources and reducing heat-induced absenteeism (Cho, 2017; Garg et al., 2020), both of which slow the erosion of learning over the school year.

We therefore interact Progresa only with academic-year temperature, since it offsets chronic erosion but cannot protect against the transitory cognitive and physiological impairment that exam-day heat causes.

We estimate equation (6) by interacting the temperature measure with a student-level indicator for Progresa receipt:

$$
\begin{aligned}
test_{ismgy} = {} & \beta_C \, T_{my} + \beta_A \, T_{my}^{\text{exam}} + \theta \, CCT_{iy} + \lambda \, CCT_{iy} \times T_{my} \\
& + \delta \, P_{my} + \eta_i + \rho_m + \gamma_y + \phi_g + \tau_{sy} + \epsilon_{ismgy}
\end{aligned} \tag{6}
$$

where CCT_{iy} indicates whether student i received the Progresa transfer in academic year y, and all other variables and fixed effects are defined as in equation (5). The coefficient β_C captures the effect of academic-year temperature for non-beneficiaries, and λ captures the differential effect for Progresa recipients—the object of interest. Because only 2.7% of students lose the benefit during the sample period, the CCT main effect θ is largely absorbed by the student fixed effect and is not independently informative. The low attrition rate also implies that the composition of the beneficiary group remains stable over time, mitigating concerns about differential selection out of treatment. The sum $\beta_C + \lambda$ gives the total temperature effect for beneficiaries. We restrict the sample to grades 6–9, where Progresa receipt is directly observed.

Table 4 reports the results. The interaction between Progresa receipt and academic-year temperature is positive and statistically significant across all specifications, indicating that the CCT attenuates the chronic effect of heat, consistent with the corollary. In the full sample, a 1°C increase in academic-year temperature reduces scores by 0.012 standard deviation for non-beneficiaries, but the interaction term (0.018) nearly fully offsets this effect: the net marginal effect for beneficiaries is not statistically different from zero.

Students from households poor enough to qualify for Progresa (who would otherwise be expected to be more vulnerable given their low κ) are effectively shielded from the chronic channel. This pattern holds across climate zones: in warm, dry, and humid municipalities, the total effect for beneficiaries is not significantly different from zero, while in temperate municipalities the marginal effect is positive and significant.[25]

Table 4: Conditional Cash Transfers and the Temperature–Learning Relationship

	(1) All Sample	(2) Warm	(3) Temperate	(4) Humid	(5) Dry
Avg Max Temp (°C)	-0.0122**	-0.0292***	-0.0037	-0.0113	-0.0089*
	(0.0052)	(0.0077)	(0.0064)	(0.0099)	(0.0049)
Progresa × Avg Max Temp	0.0184***	0.0323***	0.0246***	0.0198***	0.0131***
	(0.0019)	(0.0039)	(0.0035)	(0.0020)	(0.0032)
Exam Day Temp (°C)	-0.0058***	-0.0005	-0.0075**	-0.0068**	-0.0067***
	(0.0017)	(0.0023)	(0.0033)	(0.0030)	(0.0024)
Beneficiaries Effect	0.0062	0.0031	0.0209***	0.0085	0.0042
	(0.0052)	(0.0073)	(0.0066)	(0.0096)	(0.0059)
Observations	31,382,549	16,643,069	14,625,454	20,313,995	10,883,865

Notes: The dependent variable is the average standardized score between Mathematics and Spanish. All specifications follow equation (6). Avg Max Temp is the average daily maximum temperature during the academic year. Progresa × Avg Max Temp is the interaction between the student-level CCT receipt indicator and academic-year temperature. Exam Day Temp is the maximum temperature on the test day. Marginal effect: beneficiaries reports the sum $\hat{\beta}_C + \hat{\lambda}$, i.e., the total academic-year temperature effect for Progresa recipients. The sample is restricted to grades 6–9, where Progresa receipt is directly observed. All specifications include student, municipality, year, and grade fixed effects, state-specific linear time trends, and control for mean daily precipitation. Standard errors clustered at the municipality level in parentheses.

We note two limitations of this analysis. First, our data do not allow us to disentangle the specific channel through which Progresa attenuates the chronic effect—whether through the income transfer itself or through the attendance conditionality. In the framework, these operate through distinct objects: the income transfer expands the household resource bundle that offsets daily heat erosion, raising κ, while the attendance conditionality keeps students in school, raising the effective number of instructional days D and thus accumulated human capital H directly. Both attenuate the chronic effect, but they have different policy implications: the first suggests that un-

[25]This positive effect in temperate areas may reflect two reinforcing channels through which Progresa operates. First, where chronic heat erosion is already low, the protective channel is less binding and the direct effect of the transfer on human capital—through better nutrition, attendance, and compensatory investments—can dominate. Second, this pattern is consistent with Arceo-Gomez and López-Feldman (2024), who find that heat improves test scores among Progresa students in colder climates, and echoes the positive middle-school academic-year coefficients in temperate municipalities (Table 3). Arceo-Gomez and López-Feldman (2024) uses a subsample of the ENLACE data that focuses only on Progresa recipients.

conditional transfers could also be protective, while the second points specifically to the value of attendance incentives.

6 Conclusion

This paper examines how heat affects student learning in Mexico, a country whose climatic and socioeconomic diversity allows us to identify which channel of heat exposure matters most and for whom. We use a nationwide individual-level panel of more than eighteen million students linked to daily municipal-level weather data, which provides the scale and geographic breadth needed to jointly estimate chronic and acute channels across heterogeneous settings. We develop a simple conceptual framework in which chronic exposure erodes learning accumulated over the school year and acute exposure impairs performance on the exam day, with a student's protective capacity, mediating how much chronic erosion occurs during the academic year. The framework's implications align with the empirical patterns we document.

The central finding is that the channel through which heat affects test scores is not uniform, but depends systematically on climate region and student vulnerability. In warm and humid municipalities, chronic exposure over the academic year drives the negative effect, consistent with persistent heat eroding protective capacity throughout the school year. In temperate and dry municipalities, acute exam-day exposure drives the effect instead, consistent with students retaining enough protective capacity that only transitory shocks matter. Within climate regions, the chronic channel is strongest among more vulnerable populations (high marginalization, rural areas, younger students), while the acute channel is strongest among less vulnerable populations. These patterns are not easily visible in pooled specifications or single-channel designs.

We also provide suggestive evidence that social protection can attenuate the chronic channel. Progresa receipt nearly fully offsets the negative effect of academic-year temperature for beneficiaries, consistent with the framework's implication 3 that raising protective capacity should weaken chronic heat's impact. Whether this attenuation operates primarily through the income transfer or through attendance conditionalities remains an open question with distinct policy implications: the first would suggest that unconditional transfers could also be protective, while the second points to the specific value of attendance incentives in hot climates.

These results open several directions for future work. First, social protection is only one of several levers that could raise protective capacity. Testing whether others, such as school infrastructure, air conditioning, or schedule changes can attenuate the chronic and acute channels differently would show which policies fit which settings. Second, what drives the chronic-channel effects remains an open question. Our data identify the channel but cannot separate the underlying mechanisms: lost attendance, health shocks from vector-borne disease, or child labor. A clear decomposition of this channel would help to explain its geographic concentration and which policies can address it. Third, understanding whether heat-induced learning losses persist into later schooling and labor market outcomes is important for quantifying the long-run human capital cost of climate change. Such an analysis would require linking individual students to their later schooling and labor-market records, which administrative systems in many developing countries rarely connect across institutions or over time.

References

Agarwal, S., Ghosh, P., Scarazzato, F., and Waltl, S. R. (2025). The association between adverse temperature shocks and schooling outcomes in india: Impact quantification and mitigation potentials. Working Paper 2025-06, University of Cambridge, Real Estate Research Centre, Cambridge, UK.

Agüero, J. M. and Beleche, T. (2013). Test-mex: Estimating the effects of school year length on student performance in mexico. *Journal of Development Economics*, 103:353–361.

Arceo-Gomez, E. O. and López-Feldman, A. (2024). Extreme temperatures and school performance of the poor: Evidence from mexico. *Economics Letters*, 238:111700.

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., and Shapiro, J. S. (2016). Adapting to climate change: The remarkable decline in the us temperature-mortality relationship over the twentieth century. *Journal of Political Economy*, 124(1):105–159.

Behrman, J. R., Parker, S. W., Todd, P., and Zhang, W. (2025). Prospering through prospera: A dynamic model of cct impacts on educational attainment and achievement in mexico. *Quantitative Economics*, 16(1):133–183.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):352–365.

Burke, M., Hsiang, S. M., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239.

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

De Hoyos, R., Estrada, R., and Vargas, M. J. (2018). Predicting individual wellbeing through test scores: evidence from a national assessment in mexico. Policy research working paper, The World Bank.

de Hoyos, R., Estrada, R., and Vargas, M. J. (2021). What do test scores really capture? evidence from a large-scale student assessment in mexico. *World Development*, 146:105524.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic literature*, 52(3):740–798.

Garg, T., Jagnani, M., and Taraz, V. (2020). Temperature and human capital in india. *Journal of the Association of Environmental and Resource Economists*, 7(6):1113–1150.

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.

Kraft, M. A. and Novicoff, S. (2024). Time in school: A conceptual framework, synthesis of the causal research, and empirical exploration. *American Educational Research Journal*, 61(4):724–766.

Park, R. J. (2022). Hot temperature and high-stakes performance. *Journal of Human Resources*, 57(2):400–434.

Park, R. J., Behrer, A. P., and Goodman, J. (2021). Learning is inhibited by heat exposure, both internationally and within the united states. *Nature Human Behaviour*, 5:19–27.

Park, R. J., Goodman, J., Hurwitz, M., and Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy*, 12(2):306–339.

Parker, S. W. and Todd, P. E. (2017). Conditional cash transfers: The case ofprogresa/oportunidades. *Journal of Economic Literature*, 55(3):866–915.

Pereira, P. F. d. C. and Broday, E. E. (2021). Determination of thermal comfort zones through comparative analysis between different characterization methods of thermally dissatisfied people. *Buildings*, 11(8):320.

Rothfusz, L. P. and Headquarters, N. S. R. (1990). The heat index equation (or, more than you ever wanted to know about heat index). *Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology*, 9023:640.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to u.s. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106(37):15594–15598.

Steadman, R. G. (1979). The assessment of sultriness. part i: A temperature-humidity index based on human physiology and clothing science. *Journal of Applied Meteorology and Climatology*, 18(7):861–873.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):F3–F33.

Venegas Marin, S., Schwarz, L., and Sabarwal, S. (2024). Impacts of extreme weather events on education outcomes: A review of evidence. *The World Bank Research Observer*, 39(2):177–226.

Vu, T. M. (2022). Effects of heat on mathematics test performance in vietnam. *Asian Economic Journal*, 36(1):72–94.

Wargocki, P., Porras-Salazar, J. A., and Contreras-Espinoza, S. (2019). The relationship between classroom temperature and children's performance in school. *Building and Environment*, 157:197–204.

Yeganeh, A. J., Reichard, G., McCoy, A. P., Bulbul, T., and Jazizadeh, F. (2018). Correlation of ambient air temperature and cognitive performance: A systematic review and meta-analysis. *Building and Environment*, 143:701–716.

Zhang, X., Chen, X., and Zhang, X. (2024). Temperature and low-stakes cognitive performance. *Journal of the Association of Environmental and Resource Economists*, 11(1):75–96.

Appendix

"The Unequal Burden of Heat on Learning"

A Data Appendix

1.1 Weather Data Construction

This section documents the construction of the weather variables used in the analysis.

Data sources. Temperature data come from CHIRTS-ERA5 v3 (Climate Hazards Center, UC Santa Barbara), which provides daily maximum and minimum temperature at 0.05° spatial resolution, covering 60°S–70°N from 1980 to 2023. Mean temperature is computed as $T_{mean} = (T_{max} + T_{min})/2$ at the pixel level before spatial aggregation. Precipitation data come from CHIRPS v3.0, which provides daily rainfall at the same resolution, covering 60°S–60°N from 1981 to the present. Both datasets are accessed via GDAL virtual filesystem from the Climate Hazards Center data server.

Municipality boundaries are defined using INEGI's 2018 municipal official boundaries (2,463 municipalities). The municipality identifier is id_mun $= 1000 \times$ state code$+$ municipality code, matching the identifier constructed from ENLACE microdata.

Spatial aggregation. We construct a cell-municipality weight matrix following Barreca et al. (2016). For each municipality, we identify all CHIRTS-ERA5 raster cells whose footprint overlaps the municipality polygon. The overlap fraction—defined as the cell area intersecting the municipality divided by total cell area—is computed using exact geometric extraction. Fractions are normalized to weights summing to 1.0 within each municipality.

We verify grid stability across three dates spanning the study period (2006-09-01, 2010-06-15, 2015-04-01): row count, column count, resolution, and extent are identical across dates, confirming that cell indices are fixed over time. The same weight matrix is applied to both temperature and precipitation, as both products share the 0.05° grid geometry.

For each municipality-day, the aggregation procedure is:

1. Extract pixel values using the pre-computed cell indices from the weight matrix.

2. Apply outlier filters: precipitation values outside $[-1, 700]$ mm and temperature values outside $[-20, 65]$°C are set to missing. For temperature, a pixel is dropped if either T_{max} or T_{min} is missing.

3. Renormalize weights after dropping invalid pixels, ensuring weights sum to 1 using only valid cells.

4. Compute the area-weighted mean: $\bar{x}_{m,d} = \sum_i w_{i,m} \cdot x_{i,d}$, where $w_{i,m}$ is the normalized weight of cell i in municipality m.

A municipality-day is flagged as low-coverage if the count of valid cells falls below 70% of a baseline count (the number of valid cells on a reference date: January 1, 2000 for CHIRPS, January 1, 1980 for CHIRTS). A month is flagged as incomplete if it has fewer than 27 days with non-missing data.

Academic-year averages. We compute cumulative weather exposure over the academic year (August through May or June), excluding official school breaks:

- Winter break (Christmas): variable dates per year, typically December 19–22 through January 2–7.

- Spring break (Easter): variable dates per year, typically two weeks in March–April.

Vacation dates are verified from official SEP calendars for each year. Four nested truncation points allow testing whether cumulative versus recent exposure drives effects: (i) full academic year, (ii) up to the day before the exam, (iii) up to the day before the week-before-exam window, and (iv) up to the day before the month-before-exam window.

For each truncation, we compute realized means of T_{max}, T_{mean}, and precipitation, as well as long-term standardized deviations. The long-term baseline is constructed by matching calendar-day combinations across all available years in the daily panel

(2001–2018), with vacation exclusion applied to both the realized and baseline calculations. The standardized deviation is $z_{m,y} = (\bar{x}_{m,y} - \mu_m)/\sigma_m$, where μ_m and σ_m are the municipality-specific long-run mean and standard deviation for the same calendar-day window.

Exam-window averages. We construct three additional time windows defined relative to the ENLACE exam dates: the exam window itself (five test days), the week before the exam, and the month before the exam. ENLACE administration dates, verified from official SEP records, are:

Academic year	Exam start	Exam end
2007–2008	Apr 14, 2008	Apr 18, 2008
2008–2009	Apr 23, 2009	Apr 29, 2009
2009–2010	Apr 19, 2010	Apr 23, 2010
2010–2011	May 23, 2011	May 27, 2011
2011–2012	Jun 4, 2012	Jun 8, 2012
2012–2013	Jun 3, 2013	Jun 7, 2013

The shift from April to June across years provides additional temporal variation in the weather conditions under which students are tested.

Temperature bins. To capture nonlinear temperature effects, we follow Schlenker and Roberts (2009) and classify temperatures into bins at the grid-cell level before aggregating to municipalities. This avoids the bias that arises from first averaging temperatures to the municipality level and then binning, which attenuates exposure to extreme temperatures within spatially heterogeneous municipalities.

For each school day and each raster cell: (i) fetch daily T_{max}; (ii) drop values below $-20°C$; (iii) classify the temperature into a bin using left-closed intervals; (iv) renormalize weights after dropping invalid cells; (v) sum weights per municipality-bin combination, yielding fractional days of exposure. Exposures accumulate across all school days in the academic year. Bin totals within each scheme sum to total school days (validated in the data).

Our primary bin scheme (Bin 1) uses 4°C-wide intervals:

Bin	Range	Base category
1	<18°C	
2	18–22°C	✓
3	22–26°C	
4	26–30°C	
5	30–34°C	
6	≥34°C	

As robustness checks, we also estimate models using three alternative schemes: Bin 2 with 3°C-wide intervals (base 21–24°C), Bin 3 with 2°C-wide intervals over [18, 34]°C, and Bin 4 with 2°C-wide intervals over [20, 32]°C.

Climate zone classification. We classify municipalities into climate zones using INEGI's official Köppen-García climatological raster, which maps long-run climate normals covering 1902–2011 onto a fine spatial grid for all of Mexico. Each municipality is assigned the modal (most frequent) climate category across all raster cells overlapping its polygon, using exact geometric extraction with boundary-touching cells included.

We construct two independent binary classifications from the Köppen-García codes. For the *temperature* dimension: Köppen group A (tropical, *cálido*) and the warm subtypes of group B—identified by the suffix "h" in the Köppen notation (e.g., BSh, BWh)—are classified as *warm*; group C (*templado*) and the cool subtypes of group B (suffix "k", e.g., BSk, BWk) are classified as *temperate*. Group E (cold/polar) municipalities exist in the data but contain negligible student populations. For the *moisture* dimension: all group B municipalities (*seco*, including both BS semi-arid and BW arid subtypes) are classified as *dry*; all remaining municipalities—encompassing humid (suffix "f" and "m" variants) and sub-humid (suffix "w" variants) precipitation regimes—are classified as *non-dry*.

Both classifications are time-invariant and based on century-scale normals, so they capture structural climate differences rather than weather realizations during our study period. Of the 2,447 municipalities in the estimation sample, 1,489 (53.2% of students) are classified as warm and 657 (34.9% of students) as dry.

1.2 ENLACE Panel Construction

This appendix section the construction of the student-level panel used in the analysis, from raw ENLACE administrative files to the estimation sample.

Raw data structure. SEP released ENLACE micro-data in three file "bands" defined by the grade groups tested in each administrative round: (i) grades 3–5 of primary school (the "345" band), available for all six academic years 2007–08 through 2012–13; (ii) grade 6 of primary plus grade 9 of secondary (the "6p3s" band), also available for all six years; and (iii) grades 7–8 of secondary (the "sec" band), available from 2008–09 onward.[26] Each file contains one row per student-year, with fields for a unique student identifier (derived from the civil registry key, CURP), school code (CCT), grade, raw test scores, and a limited set of student and school characteristics.

Student identification and panel linkage. Students are identified across years by a hashed version of their CURP, which SEP applied consistently across all bands and years. We follow de Hoyos et al. (2021) in using this identifier to link records into a longitudinal panel. The linkage success rate is approximately 91–95 percent for grades 3–5 in most years, with a notable exception in the 2010–11 academic year, when only about half of grade 3–5 records could be linked due to an administrative change in the identifier assignment process. For the grade 6–12 bands, linkage rates are consistently above 95 percent.

Geographic assignment. Raw ENLACE files do not consistently include municipality codes across all grade-year combinations. We construct a school-to-municipality crosswalk by combining three sources: an external school catalog (234,869 schools), the grade 6–12 ENLACE files (72,262 schools), and the 2011 grade 3–5 file (43,187

[26] The band labels reflect SEP's administrative grouping of grade levels for testing logistics. In some years the exact band composition shifts slightly (e.g., the secondary band expands from grades 7–8 to grades 7–9); see the project data dictionary for year-by-year details.

schools). The crosswalk covers 350,318 unique school codes with a conflict rate of 0.8 percent (schools mapping to different municipalities across sources). After merging, 0.16 percent of student-year observations remain unmatched and are dropped from the estimation sample.

Test dates. Table 5 reports the exact ENLACE administration dates for primary and secondary levels in each academic year. The test was administered over three consecutive days, with primary schools tested one to two weeks before secondary schools. The timing shifted across years, ranging from early April (2010–11) to late June (2011–12), which generates useful variation in exam-day temperature.

Table 5: ENLACE examination dates by academic year and school level

Academic year	Primary (grades 3–6)	Secondary (grades 7–9)
2007–08	April 14–16, 2008	April 21–23, 2008
2008–09	April 20–22, 2009	April 27–29, 2009
2009–10	April 20–22, 2010	April 27–29, 2010
2010–11	April 5–7, 2011	April 12–14, 2011
2011–12	June 19–21, 2012	June 26–28, 2012
2012–13	May 14–16, 2013	May 21–23, 2013

Cleaning and sample restrictions. Starting from 71.7 million stacked student-year records, the cleaning pipeline applies the following steps: (i) drop within-year duplicates (3,833 rows, per SEP's Nota Técnica guidance that duplicate records cannot be verified); (ii) flag out-of-range scores outside the 200–800 IRT bounds (2.4 percent of observations, predominantly ceiling effects rather than coding errors); (iii) flag records with suspected answer copying using SEP's K-index (4.7 percent of observations, with rates reaching 12.9 percent in indigenous-language schools); and (iv) flag unmatched school codes (0.16 percent). Flagged observations are retained in the clean panel; restrictions are applied at the analysis stage.

Score standardization. Raw scores are reported on a 200–800 IRT scale calibrated within each grade–subject pair. Because scaling procedures differ across grades and years, raw scores are not comparable along either dimension. We standardize each score by subtracting the grade–year mean and dividing by the grade–year standard deviation, yielding z-scores with mean zero and unit variance within each grade–year cell. Our primary outcome is the average of the standardized mathematics and Spanish scores.

Estimation sample. The estimation sample is constructed from the clean panel by: (i) dropping observations with missing standardized scores (zero observations lost, since scores are available for all cleaned records); (ii) dropping observations with missing municipality codes (114,853 observations); and (iii) iteratively removing singleton observations for the fixed-effect combination used in the main specifications (student, year, municipality, grade, and state-specific year trends). Singleton detection follows an iterative algorithm as follows: removing an observation that is unique in one fixed-effect dimension can create new singletons in other dimensions, so the procedure repeats until convergence. In practice, convergence occurs in two iterations, with 7.5 million observations dropped—almost entirely students observed in only one year, who contribute no within-student variation to the estimates. The final estimation sample contains approximately 64 million student-year observations from roughly 18 million unique students across 2,273 municipalities.

Non-participation and structural gaps. Two geographic gaps in ENLACE coverage warrant mention. First, the state of Oaxaca did not administer ENLACE from 2009–10 through 2011–12 due to a teacher-union boycott. Second, in 2010–11, grade 3–5 records are available only for states with entity codes 1–15 (roughly the northern half of the country), representing a loss of approximately half the primary-school sample in that year. Both gaps are absorbed by the state–year fixed effects included in all specifications. The four complete cohorts documented in Table C.1 are unaffected by the 2011 grade 3–5 truncation, since by that year all four cohorts had progressed beyond grade 5 into the secondary-school bands.

Cohort structure and attrition. Table C.1 in the main text identifies four cohorts that can be followed longitudinally. These cohorts entered grade 3 in successive years beginning in 2007–08, progressing along a diagonal through the grade–year matrix. Not all students appear in every expected year: panel attrition arises from dropout, grade repetition, school transfer to non-participating institutions, and imperfect identifier linkage. Among the longest cohort (grade 3 in 2007–08, expected through grade 8 in 2012–13), approximately 58 percent of initial students are still observed at the end of the window. Students observed in only one year—the primary source of singleton observations—are predominantly from the "Other" category (students whose first appearance does not fall on any of the four cohort diagonals).

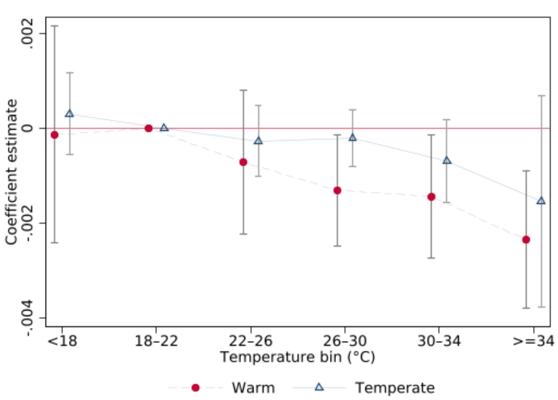
B Binned Specification

To further characterize the nonlinear relationship between temperature and academic performance, we also estimate a binned specification that replaces the continuous temperature measure with the number of days in different temperature ranges during the academic year:

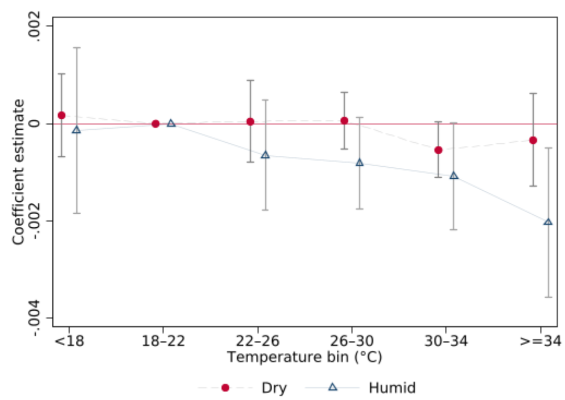
$$test_{ismgy} = \beta_1 \text{DaysAbove33}_{my} + \beta_2 \text{Days30-33}_{my} + \beta_3 \text{Days26-30}_{my} + \beta_4 \text{Days22-26}_{my} + \beta_5 \text{DaysBelow18}_{my} + \delta\, P_{my} + \eta_i + \rho_m + \gamma_y + \phi_g + \tau_{sy} + \epsilon_{ismgy} \quad (1)$$

where the omitted category is the 18–22°C range, which approximates the thermal comfort zone (Pereira and Broday, 2021). Each coefficient measures the change in test scores associated with one additional day in a given temperature range relative to one additional day in the baseline range.

Figure 1: Nonlinear Effects of Temperature on Test Scores by Climate Type



(a) Warm and Temperate Weather



(b) Dry and Humid Weather

Notes: The figure plots coefficients from the binned specification in equation (1), estimated separately by climate type. The omitted category is 18–22°C. Panel (a) reports estimates for warm and temperate municipalities; Panel (b) reports estimates for dry and humid municipalities. Bars represent 95% confidence intervals. All specifications include student, municipality, year, and grade fixed effects, state-specific linear time trends, and control for mean daily precipitation. Standard errors clustered at the municipality level.

Figure 1 shows that, in warm and humid municipalities, the negative effects concentrate sharply above 30°C, with bins below the thermal comfort zone showing no detectable effect. In temperate and dry municipalities, the gradient is flatter and less precisely estimated, consistent with their lower exposure to extreme heat days.

C Tables

Table C.1: Identifiable student panels by cohort

Academic Year	**3**	**4**	**5**	**6**	**7**	**8**	**9**
2007–2008	C1	C2	C3	C4			
2008–2009	C5	C1	C2	C3	C4		
2009–2010	C6	C5	C1	C2	C3	C4	
2010–2011		C6	C5	C1	C2	C3	C4
2011–2012			C6	C5	C1	C2	C3
2012–2013				C6	C5	C1	C2

Notes: In Mexico, compulsory basic education consists of six years of elementary school (grades 1–6) followed by three years of middle school (grades 7–9). The table summarizes the student cohorts observed across academic years taking the ENLACE test from 2007 to 2013. Labels "C" indicate specific cohorts that can be followed longitudinally, with each cohort representing the same group of students tracked as they progress through grades.

Table C.2: Summary Statistics

	All	Dry	Humid	Warm	Temperate
Panel A: Education & Climate					
Average standardized score	0.008	0.014	0.005	0.007	0.010
Math standardized score	0.008	0.004	0.010	0.019	-0.005
Language standardized score	0.008	0.024	-0.001	-0.007	0.025
Share of scholarship recipients	0.268	0.191	0.310	0.320	0.211
Mean max temperature, full year (C)	26.85	26.17	27.22	29.04	24.42
Mean max temperature, exam window (C)	30.68	30.71	30.67	32.75	28.38
Rainfall deficit index	0.040	0.052	0.034	0.026	0.056
Days >34C per school year	15.8	17.3	15.0	28.7	1.5
Panel B: Demographics & Infrastructure (Census 2010)					
Indigenous language speakers (%)	0.055	0.015	0.077	0.073	0.035
Rural population (%)	0.234	0.182	0.262	0.278	0.185
Illiteracy rate, 15+ (%)	0.070	0.047	0.082	0.083	0.054
Average years of schooling	8.60	8.81	8.48	8.29	8.94
Non-attendance, ages 6–11 (%)	0.029	0.025	0.031	0.031	0.026
Non-attendance, ages 12–14 (%)	0.077	0.073	0.080	0.083	0.070
No access to piped water (%)	0.114	0.064	0.140	0.144	0.080
No health insurance (%)	0.332	0.294	0.353	0.308	0.358
No drainage/sanitation (%)	0.094	0.065	0.109	0.109	0.076
Panel C: School Characteristics					
Air conditioning	0.229	0.318	0.179	0.352	0.100
Potable water	0.921	0.968	0.895	0.889	0.955
High student absenteeism	0.035	0.040	0.033	0.034	0.037
Student absenteeism (ordinal 1–5)	2.36	2.39	2.34	2.30	2.42
High teacher absenteeism	0.015	0.015	0.016	0.015	0.016
Teacher absenteeism (ordinal 1–5)	1.97	1.97	1.97	1.90	2.04
External class disruptions	0.244	0.226	0.255	0.287	0.200
N municipalities	2268	613	1655	1406	862
Share of students (%)	100.0	35.2	64.8	52.8	47.2
Students per year (approx.)	10,739,555	3,777,445	6,962,110	5,666,753	5,072,802

Notes: All statistics weighted by student enrollment to reflect the estimation sample. Panels A–B at municipality level; Panel C at school level. Test scores standardized within grade and year (mean 0, SD 1). Panel B reports census 2010 municipality characteristics; values describe the municipalities where students are tested, not individual student attributes. Climate columns based on INEGI Köppen-García climatological normals (warm/temperate by temperature regime, dry/humid by precipitation regime). High absenteeism equals one if the school director reports "almost always" or "always" on a 1–5 frequency scale; ordinal rows report the raw scale mean. Panel C covers ≈87% of director questionnaire records matched to ENLACE, 2008–2013.

Table C.3: Robustness: Heat Index as Alternative Temperature Measure

	(1) All Sample	(2) Warm	(3) Temperate	(4) Humid	(5) Dry
Panel A: Avg HI (°C)					
Full Academic Year	-0.0258***	-0.0242***	-0.0229***	-0.0343***	-0.0141***
	(0.0046)	(0.0064)	(0.0056)	(0.0070)	(0.0053)
Panel B: Exam Day HI (°C)					
Full Academic Year	-0.0242***	-0.0240***	-0.0141***	-0.0335***	-0.0103*
	(0.0047)	(0.0063)	(0.0053)	(0.0073)	(0.0055)
Exam day	-0.0020	-0.0010	-0.0060***	-0.0009	-0.0062***
	(0.0013)	(0.0014)	(0.0019)	(0.0016)	(0.0015)
Observations	64,005,332	33,576,955	30,251,695	41,240,101	22,513,653
Mean HI	17.2	20.7	13.3	17.7	16.2
SD	0.30	0.33	0.25	0.26	0.32

Notes: This table replicates Table 1 replacing daily maximum temperature with the Heat Index, which combines temperature and relative humidity into a measure of perceived heat stress. All specifications and fixed effects are identical to those in Table 1. Standard errors clustered at the municipality level in parentheses.

Table C.4: Sensitivity of Temperature Effects to Fixed Effects Specification and Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Avg Max Temp — Academic Year	0.0017	-0.0155***	-0.0155***	-0.0151***	-0.0151***	-0.0153***	-0.0153***
	(0.0035)	(0.0054)	(0.0054)	(0.0049)	(0.0049)	(0.0054)	(0.0054)
Avg Max Temp — Exam Day	-0.0148***	-0.0135***	-0.0129***	-0.0031**	-0.0032**	-0.0031**	-0.0032**
	(0.0019)	(0.0019)	(0.0019)	(0.0016)	(0.0016)	(0.0015)	(0.0015)
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality FE		Yes	Yes	Yes	Yes	Yes	Yes
Grade FE			Yes	Yes	Yes	Yes	Yes
State × Year				Yes	Yes	Yes	Yes
Pollution (PM2.5)					Yes		Yes
Humidity (RH)						Yes	Yes
Observations	62,957,649	62,957,649	62,957,649	62,957,649	62,957,649	62,957,649	62,957,649

Notes: The dependent variable is the average standardized score between Mathematics and Spanish. Columns (1)–(4) add fixed effects incrementally: column (1) includes student and year fixed effects; column (2) adds municipality fixed effects; column (3) adds grade fixed effects; column (4) adds state-by-year fixed effects. Columns (5)–(7) keep the column (4) fixed effects and add academic-year air pollution (PM2.5), relative humidity, and both as controls. "Academic Year" is the average of daily maximum temperatures (°C) during the school year; "Exam Day" is the daily maximum temperature on the test date. All columns control for mean daily precipitation, and the sample is held fixed across columns. Standard errors clustered at the municipality level in parentheses.

Table C.5: Robustness: Leads and Lags of Academic-Year Temperature

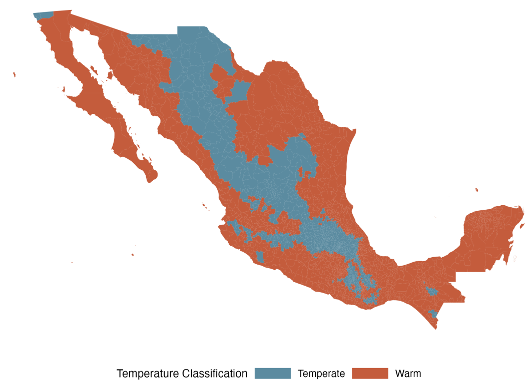
	All	Warm	Temperate	Humid	Dry
Acad-Year Temp $(t-1)$	-0.001	0.006	-0.011	-0.008	0.004
	(0.004)	(0.005)	(0.008)	(0.008)	(0.005)
Acad-Year Temp (t)	-0.011**	-0.024***	-0.005	-0.027**	-0.005
	(0.004)	(0.008)	(0.005)	(0.010)	(0.005)
Exam-Day Temp	-0.004***	0.000	-0.006***	-0.001	-0.006***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Acad-Year Temp $(t+1)$	-0.005	-0.007	-0.003	-0.012	-0.003
	(0.005)	(0.006)	(0.006)	(0.008)	(0.004)

Notes: The dependent variable is the average standardized score between Mathematics and Spanish. Each column estimates equation (5) for the indicated climate group, adding a one-year lag $(t-1)$ and a one-year lead $(t+1)$ of academic-year maximum temperature alongside the contemporaneous (t) and exam-day measures. The lead and lag are statistically insignificant, and the contemporaneous and exam-day effects remain unchanged, so future and past temperatures do not predict current scores. All specifications include student, municipality, year, and grade fixed effects, state-specific linear time trends, and control for mean daily precipitation. Standard errors clustered at the municipality level in parentheses.

D Figures

Figure D.1: Municipalities by Climate Zone

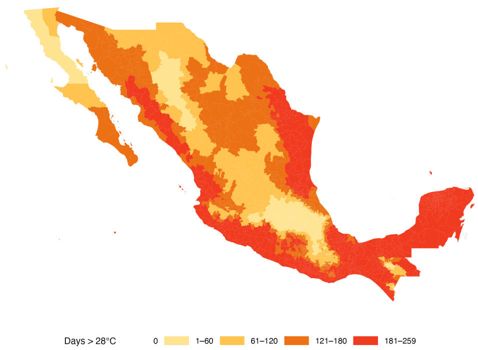
(a) Dry/Humid



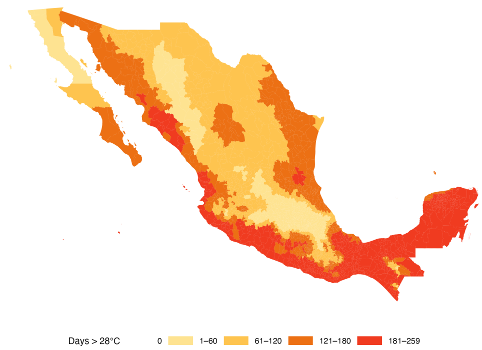
(b) Warm/Temperate

Notes: Classification of municipalities as warm or temperate and as dry or humid based on INEGI's Köppen–García climatological normals. The two dimensions overlap: 40% of dry municipalities are temperate, while the remaining 60% are warm.

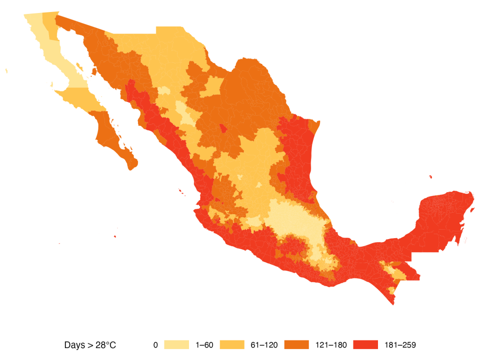
Figure D.2: Geographical variation of temperature during academic year



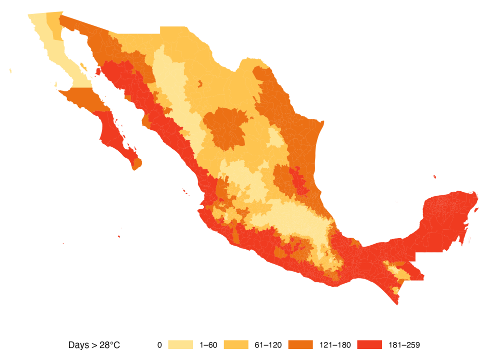
Panel a) 2007-2008



Panel b) 2009-2010



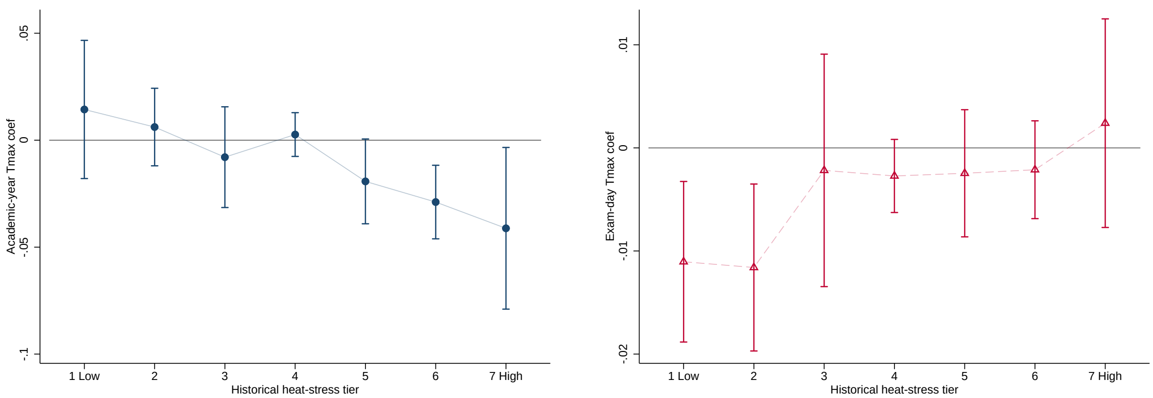
Panel c) 2011-2012



Panel d) 2013-2014

Notes: Figure presents the distribution of the number of days a municipality is above 28 °C during the beginning of the academic year until the date of ENLACE.

Figure D.3: Chronic and acute effects as a function of baseline climate region



(a) Academic year Coefficient

(b) Exam date coefficient

Notes: The figure plots coefficients from equation (5) estimated separately within each historical heat-stress tier of municipalities. Tiers rank municipalities into seven equal-sized groups by the share of school-year days during 1980–2007 on which the daily Heat Index exceeded 22°C. Tier 1 contains the municipalities with the lowest historical exposure to harmful heat and tier 7 those with the highest. We use the Heat Index because effective heat stress depends jointly on temperature and humidity (humidity blocks evaporative cooling and raises the physiological cost of any given temperature), and the Heat Index combines both through the NOAA/Rothfusz (1990) algorithm. The classification window ends in 2007, before the ENLACE estimation sample begins, so tier assignment is exogenous to the outcomes. Panel (a) reports the academic-year (chronic) coefficient and Panel (b) the exam-day (acute) coefficient. Bars are 95% confidence intervals. All specifications include student, municipality, year, and grade fixed effects, state-specific linear time trends, and control for mean daily precipitation. Standard errors clustered at the municipality level.