Sweating bullets: Heat, high-stakes evaluations, and the role of incentives

Jose María Martínez, Michigan State University, mart2388@msu.edu
Victor Zuluaga, University of Houston, vhzuluaga@central.uh.edu
Alexander Buriticá, Alliance of Bioversity International and CIAT, a.buritica@cgiar.org

Selected Paper prepared for presentation at the 2024 Agricultural & Applied Economics Association Annual Meeting, New Orleans, LA; July 28-30, 2024

Sweating bullets: Heat, high-stakes evaluations, and the role of incentives*

Jose María Martínez†[1], Victor Zuluaga[2], and Alexander Buriticá[3]

[1]*Michigan State University. East Lansing, MI.*
[2]*University of Houston. Houston, TX.*
[3]*Alliance of Bioversity International and CIAT. Cali, Colombia.*

First Draft: August 16, 2023.
This version: January 19, 2024.

Abstract

We study the effect of temperature on students' performance and how changes in the incentives to study alter this impact in the context of high-school exit exams in Colombia. We show that temperature increases have a negative impact on exam scores, particularly among urban students. Conversely, rural students exhibit slightly positive effects. Leveraging time-use data, we find evidence of individuals in rural areas responding to increases in temperature by reallocating time towards off-farm activities, which are human capital intensive. Additionally, the announcement of a national scholarship program, which introduced exogenous variation in exam stakes, reveals that heightened student effort exacerbates the temperature's impact on scores. In particular, an interquartile change in the exposure to this program increases the impact of temperature on exam scores by 12.4%. This underscores the intricate relationship between incentive-based policies and the challenge of rising temperatures. As global temperatures continue to rise, understanding this dynamic is crucial for informing effective educational policy.

Keywords: Temperature, human capital, climate change, adaptation.
JEL: I21, I26, J24, Q51, Q54, Q56.

***Competing interests**: The authors claim that they have no conflict of interest to report.
Acknowledgments: We are grateful to Songqing Jin, Chinhui Juhn, Mywish Maredia, David Ortega, James Sears, Laura Castro, and all the participants at the University of Houston and Michigan State University Student Brown Bag Seminars for their helpful comments. We also thank the Data Center of Universidad de los Andes for their help in accessing the data from the Colombian Longitudinal Survey (ELCA). All errors that remain are our own.

†**Corresponding author**: mart2388@msu.edu. Address: 446 W Circle Drive, Room 202. East Lansing, MI 48824. ORCID: 0000-0003-4464-2374.

1 Introduction

Although climate change effects are widely studied with a focus on the long run and on monetary measurements of its impacts (Carleton & Hsiang, 2016; Lenton et al., 2023), temperature and environmental changes are becoming increasingly evident in the short run. Hence, analyses that properly measure how harsher weather conditions currently affect diverse social and economic dimensions become of the highest priority (Carleton et al., 2022; Dell, Jones, & Olken, 2014; Hsiang & Kopp, 2018). Human capital accumulation has recently become a central topic in this regard (Garg, Jagnani, & Taraz, 2020; Park, Behrer, & Goodman, 2021; Park, Goodman, Hurwitz, & Smith, 2020). Nevertheless, most of these findings focus on cases of temperate or subtropical weather such as the United States or China, where temperature shocks can be highly seasonal, and thus have a higher relevance if they occur specifically during the period in which students are being evaluated. The case of teenagers and young adults in hot tropical countries, where exposure to increasingly high temperatures throughout the year is becoming the rule instead of a more frequent seasonal event, remains poorly explored.

What are the impacts of long-term exposure to increasing temperatures on students' performance in high-stakes evaluations? How does a hotter environment affect the cost of exerting effort in these high-stakes tests? Using a model of human capital accumulation with temperature stress, we derive three main testable hypotheses: (1) increases in heat exposures negatively affect human capital accumulation, (2) positive shocks to expected income favor the accumulation of human capital, and (3) exerting effort is costly when experiencing hotter temperatures. Combining five years of information from a national semi-annual high school exit exam in Colombia (Saber 11) and daily weather station data on temperature and precipitation, we use fixed-effects estimation to measure the effect of temperature on exam scores. Also, we exploit a large-scale scholarship program (i.e., *Ser Pilo Paga* (SPP)) as a source of exogenous variation in the stakes of the exam, allowing us to set up a quasi-experimental design to test whether temperature increases make it costly to exert effort in preparation for a high-stakes exam.

We find that temperature increases have significant effects on Saber 11 scores, especially for students in urban areas. Our estimates suggest that a 1°C increase in the average daily maximum temperature experienced in the year before the exam reduces scores by at least 2% of a standard deviation. Relative to days with temperatures between 20-25°C, each additional day of exposure to temperatures between 30-35°C causes a math score reduction of 0.05% of a standard deviation, while an additional day over 35°C reduces this score by 0.11% of a standard deviation. Slightly lower effects are found for Spanish and total scores. Interestingly, we find that higher temperatures slightly *improve* students' performance in rural areas. Leveraging detailed time-use data, we provide suggestive evidence that higher temperatures lead to an increase in the time spent by youths on activities complementary to studying. We also find that temperature increases lead to a reallocation in the time spent by adults on off-farm labor, an activity that is more human-capital intensive.

Combining variation in temperatures and in the exam stakes resulting from SPP, we find that for each additional percentage point of exposure to the program, average math scores increase between 2.12% and 6.18% of a standard deviation, with no noticeable effects on Spanish scores. Among students in urban areas, increases in SPP exposure imply lower math score gains for each additional increase in the average maximum temperature experienced in the year before the exam. Putting these effects into perspective, our results suggest that an interquartile change in the exposure to SPP leads to a 12.4% increase in absolute terms in the impact of temperature on exam performance. Therefore, our findings suggest that the additional effort exerted by students as a response to the program led to a stronger negative effect of temperature on scores.

While this subject is still an emerging literature, and most studies focus on the short-term effects of temperature shocks (e.g., those occurring the day of the exam or the week before it) on score performance, some well-defined patterns are worth highlighting. Graff-Zivin, Hsiang, and Neidell (2018) uses same-day weather variation to measure the effect of temperature in the performance of reading and math tasks in the National

Longitudinal Survey of the Young - 1979 cohort, with evidence suggesting that only "high-complexity" tasks like math are affected by these shocks. Similar results for low-stakes examinations were found for China in Zhang, Chen, and Zhang (2024).

In the context of high-stakes examinations in Korea, Cho (2017) also finds a significant effect of same-day temperature shock on math scores but not on reading. Higher temperatures during the day of the exam have also been found to decrease exam scores in China (Graff-Zivin, Song, Tang, & Zhang, 2020), the United States (Park, 2022), and at the world level based on PISA tests (Park et al., 2021). In the context of South America, Li and Patel (2021) find for Brazil that the effects of higher temperatures on high-stakes exam scores may not be substantive. Conversely, Hoffmann, Pulido, and Vera-Cossio (2023) finds that exposure to high temperatures in Colombia during the week before a Saber 11 test day negatively affects scores; on average, an additional hour of exposition to temperatures above 32°C reduces total scores by 0.11% of a standard deviation. To our knowledge, only the works in Park et al. (2020), for the United States, and Garg et al. (2020), for India, exploit the temperature in the year before the exam as potentially affecting final scores, finding that days above 30°C can reduce scores by 0.05% and 0.3% of a standard deviation compared to days with temperatures between 15-17°C, respectively.

Our contribution is, hence, threefold. First, we contribute to the literature on the effects of heat on high-stakes exam performance showing that longer-term exposure to increasing temperatures has negative significant effects, thus suggesting that heat affects not only exam performance but also human capital accumulation (Garg et al., 2020; Graff-Zivin et al., 2018; Park et al., 2021). Second, we study the effect of temperature on students' scores in a tropical country, where the range of variation of temperature is particularly narrow and, therefore, the adaptation strategies available to students may differ (Helo Sarmiento, 2023). Understanding these dynamics is of the utmost importance for countries for which, as in the case of Colombia, the use of adaptation strategies like air conditioning is severely low (under 5% of buildings (Statista Research Department, 2023)). Third, and more importantly, using an exogenous shock to the stakes of the exam, we provide quasi-experimental evidence revealing that exerting effort becomes increasingly costly as individuals are exposed to increasing temperature levels and, therefore, provide evidence of an adaptation mechanism elusive so far in the literature (Burke et al., 2016). Moreover, we provide correlational evidence of a mechanism of opportunity costs that supports both the predictions of our theoretical model and our causal analysis.

The remainder of our paper proceeds as follows. In Section 2, we set up a theoretical model revealing the structural relationship between exposure to heat and the accumulation of human capital and the role of incentives in this relationship. Then, in Section 3, we briefly discuss *Ser Pilo Paga*, its relevance for our analysis, and we summarize our data. Section 4 presents our identification strategies in detail. We present and discuss our main results and robustness checks in Section 5. We conclude with Section 6, where we also provide directions for future research.

2 Theoretical Framework

We start by setting a static model where human capital accumulation is sensitive to temperature stress. We have two goals with our model. First, we want to show how changes in the incentives to exert effort may interact with temperature and, as a result, affect the accumulation of human capital. Second, we use the framework to provide an interpretation of the effects being recovered by our empirical analysis and the mechanisms through which they take place[1]

We begin by defining a human capital production function h that follows:

$$h = zf(e, t), \tag{1}$$

[1]We closely follow Park (2022), although we have two main differences with his model: (a) we add cognitive skills, which are not only important determinants of academic performance but also useful for interpretations in the presence of heterogeneous effects by socioeconomic status; and (b), given our focus on the incentives to accumulate human capital and their interaction with temperature stress, we derive comparative statics not only for temperature but also for the returns to human capital.

where z is the individual's cognitive skill or ability, e is their effort level, t is the temperature, and f is a function assumed to be increasing in effort ($f_e > 0$), decreasing in temperature ($f_t < 0$), and with negative second and cross derivatives ($f_{ee} < 0$, $f_{tt} < 0$, and $f_{et} < 0$). Individuals take temperature and their cognitive skills as given, with the latter being potentially affected by exogenous parental investments not included in our model. Conversely, their effort is chosen endogenously to solve a maximization problem (described below). In our empirical analysis, we proxy human capital with students' performance in the Colombian high-school exit exam to be described below.

Individuals derive instant utility according to $U(c, g(e, t))$, which depends positively on the consumption of a composite good c, whose price we normalize to one, and negatively on the effective effort g. The latter is a function increasing in all of its arguments ($g_e > 0, g_t > 0$), and with positive second and cross derivatives ($g_{ee} > 0$, $g_{tt} > 0$, and $g_{et} > 0$). This last assumption is consistent with the medical literature suggesting that effort is more physically taxing at higher temperatures (Wendt, Van Loon, & Marken Lichtenbelt, 2007).

Given that there is only one period, individuals consume all their labor income (i.e., there are no savings). Assuming no unemployment and that wages w are determined in a perfectly competitive labor market, we can write the individual's problem as follows:

$$\max_{c,e} \quad U(c, g(e,t)) \quad \text{s.t.} \quad wh = c, \tag{2}$$

where the restriction corresponds to the standard budget constraint. The latter means that the agent chooses its consumption and effort levels to maximize utility subject to its labor income. Plugging equation (1) in the above problem, we can derive the first-order condition

$$w\frac{\partial U}{\partial h}\frac{\partial h}{\partial e^*} = wz\frac{\partial U}{\partial f}\frac{\partial f}{\partial e^*} = -\frac{\partial U}{\partial g}\frac{\partial g}{\partial e^*},$$

which states that, at the optimum, the agent equalizes the marginal benefit of effort (increase in disposable income) to its marginal cost (increase in disutility).

Setting up our analysis with specific functions f, g, and U, we can derive closed-form solutions for the optimal levels of effort and human capital. In particular, we assume the following:

$$f(e,t) = \frac{e^{\alpha}}{t^{\beta}}, \quad 0 < \alpha < 1, \quad \beta > 0;$$

$$g(e,t) = e^{\gamma} t^{\zeta}, \quad \gamma > 1, \quad 0 < \zeta < 1;$$

$$U(c, g(e,t)) = c - g(e,t) = wzf(e,t) - g(e,t);$$

with all of these functional forms satisfying the assumptions above. The stated restrictions ensure the objective function in (2) is well-behaved (first- and second-order conditions hold).[2] Solving the maximization problem, it is straightforward to show that the optimal level of effort takes the following form:

$$e^{*} = \left(\frac{\alpha w z}{\gamma t^{\beta+\zeta}} \right)^{\frac{1}{\gamma-\alpha}}. \tag{3}$$

From this expression is clear that effort is increasing in wages and cognitive skills, while it decreases with temperature. The first two results are standard in the economics literature (see, for example, Chadi, De Pinto, and Schultze (2019)). The third one is consistent with recent empirical evidence, including our empirical work with time-use data below, suggesting that temperature affects the time allocated to working and studying (Alberto, Jiao, & Zhang, 2021; Graff-Zivin & Neidell, 2014).

Plugging the expression for the optimal effort in the human capital production function and linearizing, we can get an equation akin to what is usually estimated in the empirical literature, including our analysis below.[3] Concretely,

$$\begin{aligned} \hat{h} = &\left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha} \right] \hat{t} + \left[\frac{\alpha}{\gamma-\alpha} \right] \hat{w} + \left[\frac{\gamma}{\gamma-\alpha} \right] \hat{z} \\ &+ \frac{1}{2} \left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha} \right] \left[\frac{\alpha}{\gamma-\alpha} \right] \hat{t}\hat{w} + \frac{1}{2} \left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha} \right] \left[\frac{\gamma}{\gamma-\alpha} \right] \hat{t}\hat{z} + R, \end{aligned} \tag{4}$$

[2] One of the key assumptions required for the objective function to be concave in the effort is that $\gamma > \alpha$, this is, the disutility cost of effort increases faster than its effect on human capital (see Appendix A).

[3] The non-linearized expression for human capital and the full mathematical derivation of the optimization problem can be found in Appendix A.

where $\hat{x}$ is the percent deviation of x with respect to the population (or historical) mean and R is a residual term composed, among others, of the second-order own effects of each of the variables.[4] As can be seen, a positive deviation in cognitive skills or wages is expected to increase human capital. Conversely, higher-than-average temperatures are expected to decrease it. As in Park (2022), the effect of temperature operates through two channels: a direct effect summarized by the (structural) elasticity of human capital to temperature (β) and an indirect one operating through endogenous changes in effort resulting from temperature increases, namely $\alpha(\beta + \zeta)/(\gamma - \alpha)$. In most empirical setups, individuals' efforts are unobserved or poorly measured. Nevertheless, to identify this effect in (4) all we require is a shock to (expected) wages, namely $\hat{w}$, which in our empirical application corresponds to a policy intervention to be explained in the following section.

Given our interest in the interaction between incentives and changes in temperature stress, our expression for optimal human capital includes the cross-effect of wages and temperature. Now, notice that in our model an increase in expected wages incentivizes individuals' efforts, which is expected to increase the accumulation of human capital (first-order effect). However, exerting effort also exposes individuals to the effects of temperature and associated reductions in human capital (interaction effect). This last quantity is precisely what is captured by the fourth term in (4), which is indeed negative. As we will show in our empirical analysis, we find evidence in favor of this prediction.

Another effect of interest corresponds to the interaction between cognitive ability and temperature (fifth term in (4)). Our framework implies that this interaction should be negative as long as individuals' cognitive skills incentivize the investment in human capital. This is, given that individuals from better socioeconomic backgrounds end up exerting more effort, they are further exposed to temperature stress. We provide suggestive empirical evidence in favor of this implication.

[4]By writing the variables in deviations we account for controls such as geography or time fixed-effects, which are usual in empirical work, including ours.

3 Background and Data

3.1 Policy intervention: Ser Pilo Paga

In 2014, the Colombian Ministry of Education launched the *Ser Pilo Paga* program ("Being a Good Student Pays Off," henceforth SPP) to promote access to higher education. SPP was a scholarship initiative that provided financial support to high-achieving students from low-income backgrounds, enabling them to attend top universities in the country. The program covered the total cost of tuition to attend a four-year (or five-year) degree-granting program at any high-quality university in Colombia. The program provided individual funding to about 40,000 students, with 10,000 new students chosen each year from Colombia's top decile of the population in the high-school exit exam (Bernal & Penney, 2019; Londoño-Vélez, Rodríguez, & Sánchez, 2020; Londoño-Vélez, Rodriguez, Sánchez, & Álvarez-Arango, 2023; Medina et al., 2018).

Given the program's objectives, applicants had to meet two criteria based on merit and need to be eligible. First, they must score at or above the 90th percentile on the national standardized high-school exit test (Saber 11). Second, they must come from disadvantaged households, defined as those below the median of the wealth distribution in Colombia according to their SISBEN score, the socioeconomic index used by the Colombian Government to target social programs. Finally, to receive the award, students needed to secure admission to a high-quality accredited university in the country.

There is now abundant literature evaluating the SPP program and its effects on access to higher education, graduation, learning, and wages in the short-and the medium-run (Bernal & Penney, 2019; Laajaj, Moya, & Sánchez, 2022; Londoño-Vélez, 2022; Londoño-Vélez et al., 2020, 2023). For instance, SPP increased the likelihood of immediate access to higher education among eligible students by eliminating the gap in access by socioeconomic strata among the top performers in the Saber 11 (Londoño-Vélez et al., 2020). Relatedly, Bernal and Penney (2019) and Laajaj et al. (2022) argue that SPP increased effort as measured by improvements in the test scores of students with higher exposures to the program. On the negative side, other studies highlight that SPP allowed the concentration of resources in private universities, the financial weakening of public universities, and the reproduction of regional and social inequalities (Mora & Ruiz, 2019).

Instead of evaluating the effects of SPP on students' performance in the Saber 11 test, which others have already done (see Bernal and Penney (2019) and Laajaj et al. (2022)), we ask the following question: Did the increased effort resulting from SPP lead to a larger effect of temperature on students' test scores? This question speaks to two broad and important topics in the education and climate change literature. First, our results provide evidence of the interaction between environmental factors and education policies, in our case, a large and generous scholarship program. Second, we show that temperature stress increases the cost of effort and, therefore, that further increases in temperature from climate change may end up reducing effort levels in equilibrium. To the best of our knowledge, we are the first to provide causal evidence on this adaptation behavior.

3.2 Data

Our outcome of interest is the students' results in the Colombian high school exit exam known as Saber 11. This test has been mandatory since 1980 (compliance rate is around 97%) and is designed to support universities and all other higher education institutions in their admission processes. Saber 11 is administered semiannually on mostly fixed dates[5] (usually in March and August) by the Colombian Institute for the Evaluation of Education (ICFES). Because of methodological changes in the test that limit the comparability of the scores across time, we use the information from the period between 2014 (August) to 2019 (March), with each calendar year having information on several hundreds of thousands of students. Since Saber 11 serves as a comprehensive evaluation of knowledge and competencies acquired in school, we take it as a proxy variable of human capital accumulation.

[5]Although the dates of the test change from year to year, each of the two annual tests occurs on a single day, within the same time frame.

For weather, we use station-level data on daily maximum temperatures and total precipitation from the Colombian Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM)–the national official meteorological data. This information has been available since 1970, but we restrict our analysis to the period 2013-2019. Following Auffhammer and Kellogg (2011), we impute missing daily observations with those from the five nearest weather stations with non-missing data. Furthermore, we assign to each Municipality the data for the nearest weather station with information for the full period, which allows us to address endogeneity in the temporal coverage and quality of the data (Auffhammer, Hsiang, Schlenker, & Sobel, 2013).

As will be clear below, our empirical results point to important differences between students in urban and rural areas. To explore the mechanisms behind these findings, we use information on time use from the Colombian Longitudinal Survey (ELCA), available for the years 2010, 2013, and 2016.[6] In particular, we employ information for youths between 10 and 16 years of age for both urban and rural areas, available in the last two waves of the survey. To consider potential mechanisms from the side of rural non-farm employment, we also consider time use data on the household heads and their spouses for the three waves of ELCA. In all our analyses, we focus on a repeated cross-sectional approach to keep consistency with our main results using students' scores.[7]

Figures 1 and 2 show the average daily maximum temperature for the period 1981-2010, which is a measure of climate (Dell et al., 2014), and the average standardized Saber 11 score in math for the exams between 2014-II and 2019-I, respectively. Three results stand out: Firstly, there are high cross-sectional variations in the maximum temperature across counties, with an important number having temperatures as low as 12°C and some as high as 32°C. Secondly, the results for the math section of the Saber 11 test also show important spatial

[6]ELCA data comes from a probabilistic, stratified, multistage, and clustered sample, with information on 10,800 households (6,000 in urban areas and 4,800 in rural areas). The urban sample is representative of poor households at the national level and for five regions (Bogota, Central, Eastern, Atlantic, and Pacific). The rural sample is representative of small farmers in four micro-regions (Mid-Atlantic, Coffee growing region, *Cundiboyacense* Plateau, and Center-East).

[7]The data on time use in the ELCA has been used previously in the economics literature. In particular, Fernández, Ibáñez, and Peña (2014) study how the Colombian armed conflict affects time-use patterns among rural households using the first wave of the ELCA.

variation. And thirdly, as also evidenced in Figure 3, there is a remarkable correlation between temperatures and test scores, with most high-performing counties concentrated in the relatively colder parts of the country.

We present the descriptive statistics of the data for our analysis in Table 1.[8] Each specific area evaluated by Saber 11 allows for a score between 0 and 100,[9] while the total score (as defined by ICFES) will be between 0 and 500. On average, students in urban areas consistently outscore those in rural areas by 4.8, 4.4, and 22.9 points in math, Spanish, and total scores, respectively. Based on the test theoretical standard deviations for each area (10 points) and the total score (50 points), these amount to a difference roughly equivalent to 50% of a standard deviation. Most of the evaluated students (55%) are female, while 6% are reported as part of an ethnic minority. In terms of the education of the student's mother, 29% have completed only primary education, while 33% and 23% have completed their studies in secondary and postsecondary education, respectively.

Consistent with the climate normal in Figure 1, the average daily maximum temperature experienced at the Municipalities in the calendar year before the exam date was 26.6°C, while the median was at 27.8°C. These are close to the measures of maximum temperature on the days when the exam occurred, which have a mean of 26.87 and 28 degrees Celsius in the mean and median, respectively. In terms of temperature experienced by weather bins, on average, municipalities experienced 6 days with maximum temperatures of 15°C or below, while the number of days with temperatures over 35°C was 30.7. There is a considerable level of exposition to heat at the Municipality level, with an average of 115.2 days experiencing peak temperatures between 30 and 35°C. In addition, the average daily precipitation during the calendar year before the exam date was 5.52 millimeters per day, while it was 4.66 millimeters during the exam days. Finally, in terms of plausibly exogenous policy shocks, there were between 8 and 9 beneficiaries of the first round of SPP per Municipality, although 50 percent of these Municipalities only experienced up to two beneficiaries.

[8] The descriptive statistics for ELCA data are available in Appendix B, Table B6.

[9] Saber 11 includes five areas of evaluation: math, Spanish, natural sciences, social competencies, and English (as a foreign language). We limit our main analysis to math and Spanish (i.e., reading) since these areas are the most widely studied in the literature, thus allowing for a direct comparison of effects relative to the ones in previous work.

4 Methods

4.1 Identification Strategy

To study the average effect of heat on students' test scores, we follow a linear unobserved effects model:

$$100 \times y_{ijct} = \beta_T T_{ct} + \beta_P P_{ct} + X_{ijct}\theta + a_j + a_{st} + u_{ijct}, \quad (5)$$

where y_{ijct} is the standardized score in math (or Spanish, or total) in the Saber 11 test for student i in school j and Municipality c at time t (semester in our case). Here X_{ijct} is a vector of student characteristics (e.g., age and gender), while a_j and a_{st} are fixed effects that capture unobserved heterogeneities at the school and State-time levels, and e_{ijct} is a set of idiosyncratic shocks that affect students performance in the test. The variable of interest is T_{ct}, which is the average daily maximum temperature (in °C) experienced during the year before the test date. As an additional control, we include P_{ct}, which measures the precipitation level in the same period as T_{ct}. With this variable, we aim to control for disruptions in learning caused by above-average rain levels. The coefficient of interest is β_T, which measures the effect on the student's scores of experiencing an increase in temperature. Specifically, a 1°C increase from the school average would change scores by β_T% of a standard deviation. We follow a fixed effects framework, thus allowing for arbitrary dependence between the unobserved effects (a_j, a_{st}) and the observed explanatory variables (Wooldridge, 2010). Given that our temperature variable varies at the weather-station level, following the recommendations in Angrist and Pischke (2008), MacKinnon, Nielsen, and Webb (2022), and Abadie, Athey, Imbens, and Wooldridge (2023), we cluster our standard errors at that level to allow for arbitrary correlation between schools (and students) within the area covered by the station.

The identification assumption behind equation (5) is that conditional on students' demographic characteristics and school and State-time fixed effects, experiencing a hotter calendar year before the test is not systematically correlated with other determinants of the student's scores (Angrist & Pischke, 2008). In other words, temperatures are as good as random conditional on the controls (Dell et al., 2014). Following Hsiang (2016), we implement a series of measures to rule out several threats to the identification: First, we control for precipitations to account for other potentially correlated geophysical processes. Second, we compute temperature and precipitation using a constant group of weather stations, which rules out any correlation between our weather data and time-varying socioeconomic conditions. Third, we provide evidence suggesting that temperature does not seem to have any noticeable effects on our sample composition (see next subsection).

Following Hsiang (2016), we also estimate the following flexible specification using our daily weather data:

$$100 \times y_{ijct} = \sum_{k=1}^{K} \beta_{T,k} DD_{ct}^{k} + \beta_P P_{ct} + X_{ijct}\theta + a_j + a_{st} + u_{ijct}, \quad (6)$$

where DD_{ct}^{k} is the number of days in the period of interest where the maximum temperature in °C was in the interval k. When estimating this specification, we use the following temperature intervals: $[\text{minimum}, 15)$, $[15, 20)$, $[20, 25)$, $[25, 30)$, $[30, 35)$, and $[35, \text{maximum})$. The baseline category throughout the analysis is the number of days between 20 and 25°C; thus, $\beta_{T,k}$ measures the effect of being exposed to one additional day with a maximum temperature in the kth bin with respect to the omitted category.

One of the main goals of this paper is to investigate how changes in the incentives to achieve higher scores in the Saber 11 interact with higher temperatures. For that, we evaluate how the effect of heat varies for students exposed to SPP. In particular, we estimate the following specification

$$\begin{aligned} 100 \times y_{ijct} = {} & \beta_T T_{ct} + \beta_{T,SPP}(T_{ct} \times SPP_c \times Post_t) + \beta_{SPP}(SPP_c \times Post_t) \\ & + \beta_P P_{ct} + X_{ijct}\theta + a_j + a_{st} + u_{ijct}, \quad (7) \end{aligned}$$

where SPP_c is the share of test-takers in Municipality c that were awarded SPP in the first wave of the program, and $Post_t$ corresponds to a dummy for the post-SPP years. Notice then the interaction between SPP_c and $Post_t$ captures the exposure to SPP originated only by the first cohort of winners, which were selected based on pre-determined household characteristics (SISBEN wealth index) and realized performance in the test at the moment of the program's announcement. The coefficient of interest is $\beta_{T,SPP}$, which corresponds to the differential effect of temperature for students with higher exposures to SPP.[10] From our conceptual framework, we expect this coefficient to be negative as it maps directly to the fourth term of equation (4).

In addition to the identifying assumption for the temperature's effect previously discussed, equation (7) requires that the evolution in the students' scores for different levels of exposure to SPP would have been the same in the counterfactual without the program, this is, that there are parallel trends for the groups defined in terms of SPP_c (Angrist & Pischke, 2008). As we already mentioned, we compute this exposure variable using only the first cohort of the program, which ensures we do not capture endogenous responses in the number of winners. Now, to explore the possibility that the SPP's exposure is capturing the impact of other policies or social programs, the next subsection provides evidence on pre-trends for several dimensions arguably relevant to students' performance in the test.

4.2 Testing for identification assumptions

In our analysis, we attempt to exploit two sources of exogenous variation, namely the variations in Municipality-level temperature and the exposure to SPP. Thus, we need to evaluate whether (a) the weather correlates with attributes from students,[11] and (b) the shock from SPP has some correlation with pre-intervention attributes at the Municipality level. For the first one, we set up a linear model:

$$X_{ijct} = \eta_T T_{ct} + \eta_P P_{ct} + a_j + a_{st} + u_{ijct},$$

from which we would expect that fixed-effect estimates for η_T and η_P are statistical zeroes. In other words, we are checking if weather shocks are directly connected to attributes of individuals that can also affect their

[10] Notice that any average differences in students' scores associated with SPP_c are already captured by the school fixed effects (a_j). Similarly, the post-SPP indicator ($Post_t$) is captured by the State-time fixed effects (a_{st}).

[11] A large literature in economics has relied on the assumption that temperature is exogenous conditional of unit and time fixed effects (see Dell et al. (2014) and Hsiang (2016)). For the sake of completeness, we provide evidence in favor of it in our context.

performance in the test. We estimate this specification for six attributes: age, female student, dummies of mother's educational attainment–primary, secondary, or postsecondary-, and whether the student is from an ethnic minority. Our results suggest that there is no evidence of any statistically significant correlation between weather shocks (either from temperature or precipitation) with students' attributes (see Appendix B, Table B1). Moreover, the estimated coefficients are very close to zero, further providing support for using these weather shocks as exogenous variations directly affecting students' learning and performance.

We now consider the implicit assumption of parallel trends underlying our empirical strategy of using the announcement of SPP. The assumption of parallel trends is theoretically untestable and the practice of pretesting has recently faced considerable scrutiny (Roth, 2022). Nonetheless, results from pretesting can provide relevant information on potential pre-treatment differences that can be empirically considered to still achieve meaningful difference-in-difference estimates (Bilinski & Hatfield, 2018). As we have only one pre-treatment period (2014-II), any pretesting at the student level is unfeasible. However, following the literature on the economics of education, beyond student-specific attributes, we focus on school characteristics that have been thought to be important in explaining students' performance.[12] Given that these variables are also potentially endogenous to policy changes, controlling for them allows us to account for the possibility that SPP assignment rules can turn other social programs into confounders. Therefore, we proceed to analyze any potential differences in pre-trends at the Municipality level based on the exposure to SPP, using pre-treatment information available from the Municipality panel data from the Center of Studies on Economic Development (CEDE) of Universidad de los Andes (Colombia).

With yearly information for the 2010-2019 period on six key variables (i.e., number of schools, population of high-school age, the number of high-school students, total Saber 11 test takers, total educational admin-

[12] As discussed in Hanushek (2006) and Handel and Hanushek (2022), the effect of school resources on students' test scores seems to be modest at best. However, recent papers have found evidence of positive effects for specific inputs such as teachers (Chetty, Friedman, & Rockoff, 2014a, 2014b) or even school resources when using exogenous variation for identification (Jackson, Johnson, & Persico, 2016).

istrative personnel, and the total number of teachers), we follow an event-study approach. We regress each of these variables on a saturated model where we interact all year dummies (except 2014, the base year of our analysis) with the exposure to SPP in the first round. Figure B2 summarizes the point estimates and 95% confidence intervals. We find evidence of a potential violation of the parallel trends assumption at the Municipality level on variables like the total number of teachers and total high-school students. Meanwhile, there is a weakly significant pre-trend correlation between larger exposures to SPP and the number of Saber 11 takers in Municipalities.[13]

Based on these findings, we follow an approach in the style of Bilinski and Hatfield (2018), acknowledging the potential differences in pre-trends and setting "one step up" specifications for our main regressions. This is, we further consider specifications that allow for heterogeneous trends based on Municipality-level differences as measured in 2014, and confirm whether there are significant differences between estimates from this more complex specification to those of simple fixed-effects estimates. Controlling for trends along these dimensions, we expect that our estimates will account for other factors or policies that may affect the evolution of students' scores with different exposures to SPP, thus strengthening our causal interpretation.

5 Results and discussion

5.1 Temperature and test scores

We present the results from our baseline regression on weather shocks during the year before the exam date in Figure 4 (with full regressions' summaries in Appendix B, Table B3). When focusing on the estimates for the total population, evidence points to insignificant correlations between observed scores and increases in the average daily maximum temperature and average precipitation in the year before the Saber 11 exam. The direc-

[13]Notice that finding pre-trends in these Municipality characteristics does not necessarily imply that the same is happening with students' scores, the outcome of interest in our empirical analysis. As was already mentioned, the literature finds weak effects of school resources on students' performance (see Hanushek (2006) and Handel and Hanushek (2022)).

tion of the coefficient estimates is similar in math, Spanish, or total scores, but there are important differences in estimations when comparing urban and rural areas.

While exposure to a 1°C increase in the average maximum temperature reduces the average math score by 2.81% of a standard deviation for urban students, the coefficient estimate is positive (but statistically insignificant) for rural students. Similarly, while the same temperature increase has no statistically significant effect on Spanish scores for rural students, it reduces urban students' scores by 2.24% of a standard deviation. In terms of total scores, each additional one-degree increase decreases the result of urban students by 2.52% of standard deviation, but it is connected to an *increase* of 1.3% of a standard deviation in rural settings. On the other hand, each additional 1 mm/day in precipitation leads to a reduction of 0.89%, 0.69%, and 0.65% of a standard deviation in math, Spanish, and overall scores, respectively, for urban students. We do not detect any significant effect from precipitation in rural areas.

As evidenced by the results Figure 5 (Appendix B, Table B4), year-long weather effect estimates are robust to the addition of weather exposure during the day of the exam, suggesting that our estimates capture the effects on learning–thus in human capital accumulation–and not only a worsening on performance during the test. Again, we find no significant shocks at the national level, but marked differences persist between urban and rural areas. For urban students, our calculations still suggest a reduction in scores (-2.61%, -2.1%, and -2.58% of a standard deviation, respectively, in math, Spanish, and total) from a 1°C increase in the average maximum temperature over the previous year, and an additional millimeter of average daily precipitation still points to a statistically significant reduction of scores (-0.88%, -0.71%, and -0.63% of a standard deviation in math, Spanish, and total scores, respectively). Although numerically different, these estimates are statistically equivalent to those of our baseline estimates. Additional specifications including only the effect of test-day weather (Appendix B, Table B2) provide estimates of similar magnitudes and directions to those from Table B4, but only pointing to a significant, negative effect from higher temperatures on math scores. Namely, an additional 1°C in the exam day maximum temperature implies a reduction of 0.49% of a standard deviation.

Next, we study the effects of exposure (by number of days) across different temperature bins over the calendar year before the exam in Figure 6 (Appendix B, Table B5). For urban areas' students, on average, and relative to maximum temperatures between 20 and 25°C, one additional day exposed to temperatures between 30 and 35°C implies a reduction of 0.055% of a standard deviation in math, and reductions of 0.066% and 0.05% of a standard deviation in Spanish and overall scores. Likewise, one additional day exposed to temperatures above 35°C negatively affects scores, implying reductions amounting to 0.11%, 0.09%, and 0.1% of standard deviations in math, Spanish, and total scores, respectively. Interestingly, it turns out that our point estimates for the total score fall between 0.05% of a standard deviation in the United States (Park et al., 2020) and 0.25-0.3% of a standard deviation in India (Garg et al., 2020).

Overall, our findings–specifically those for students in urban areas–provide results in line with those in the literature. Nevertheless, these results highlight an important empirical difference between studies in this line depending on the weather variable used in the analysis. For example, studies like Graff-Zivin et al. (2020), Graff-Zivin et al. (2018), and Park (2022) focus their attention on the exposure of test takers to high temperatures during exam days. Meanwhile, our findings are more in line with those like Park et al. (2020) and Garg et al. (2020), who analyze the effects of exposure to high temperatures on test scores focusing on long-run (e.g., past-year) measures of exposure.[14]

Effects of extreme temperature events on the same day of (or week leading to) high-stakes exams are nonetheless relevant to evidence the need for investments in adaptation strategies (e.g., change of exam timing in the season). However, our findings, when taken together with Garg et al. (2020), suggest that devel-

[14]Consistently with the literature, we further consider a specification that allows for differences in the effect of temperature conditional on the long-run historical average maximum temperature (see Appendix B, Figure B1). We find that temperature increases have a significant effect on the scores of students in Municipalities with long-run temperatures below 30°C. Conversely, we do not detect any statistically significant effects for relatively high-heat Municipalities (above 30°C). All this points to adaptation playing a role in this setting (Barreca, Clay, Deschenes, Greenstone, & Shapiro, 2016; Burke et al., 2016).

oping countries with mostly hot tropical weather are more profoundly affected in an increasingly warming world. Countries like Colombia and India, where the adoption of air conditioning remains under 5% of buildings (Statista Research Department, 2023; Sung, 2022), and that will potentially experience disproportionally harsher climatic conditions (Lenton et al., 2023), are bound to experience considerable losses in human capital unless bold policy interventions take place.

Conversely, our estimates for past-year temperature effects on the performance of students in rural areas suggest that different mechanisms might play a significant role. For instance, there is a high incidence of child labor in rural Colombia, which is estimated up to 20% for children of ages between 15-17 by 2017 (Ramoni-Perazzi, Orlandoni-Merli, Castillo-Paredes, & Peña Guillén, 2021). Hence, increases in average temperatures could be connected to adjustments in time allocations that could–in theory, following Alberto et al. (2021) and Graff-Zivin and Neidell (2014)–lead these rural students to allocate more time to study, since they would likely dedicate more time to indoors activities.[15] Below, we provide suggestive evidence in favor of this explanation. While the estimates from Hoffmann et al. (2023) for short-term weather exposure to temperature are qualitatively similar to our estimates of exposure to exam-day weather shocks,[16] our findings support our main hypothesis: increases in heat lead to exposures that negatively affect the learning process and ultimately affect students' performance in high-stakes evaluations.

5.1.1 Temperature and time use

From our analysis so far, it follows that the effect of temperature on students' performance in Saber 11 varies markedly between urban and rural areas–negative for the former and slightly positive for the latter. Here we at-

[15] Another related explanation is that higher temperatures may decrease agricultural productivity and, therefore, reduce the opportunity cost of studying (see Shah and Steinberg (2021) for related evidence from India).

[16] Point estimates from Hoffmann et al. (2023), however, are statistically different from zero for total scores, unlike ours (see Appendix B, Table B2), that suggests the presence of an impact from heat only on the performance in math for urban students. It should be noted, however, that our results are not directly comparable to theirs given that we focus on the period 2014-2019 to ensure test scores are fully comparable given changes in the methodology of the test as discussed previously. Meanwhile, Hoffmann et al. (2023) use data for the period 2009-2019 instead.

tempt to uncover the mechanisms behind these heterogeneous effects. In particular, we argue that temperature increases may incentivize students in rural areas to spend more time in activities complementary to studying, with the underlying force being a reallocation of labor towards off-farm labor where the returns to human capital are potentially larger (Agrawal & Agrawal, 2019; Jolliffe, 2004). Rural non-farm employment accounts for over 50% of rural households in Asia and Latin America, and it has been identified to work as a response from these households to external shocks on production (Davis, Winters, Reardon, & Stamoulis, 2009; Haggblade, Hazell, & Reardon, 2007; Reardon, Stamoulis, & Pingali, 2007). Hence, increases in temperature would potentially push for a reduction of time allocated to on-farm labor, and instead transitioning into rural non-farm employment. Moreover, evidence from Africa and Asia suggests that as food production and processing tasks become further laborious or comparatively less efficient, there is an additional push on spouses to allocate more time in rural non-farm (or even urban) jobs (Reardon et al., 2019).

To test these hypotheses, we use data from ELCA to estimate specifications that follow:

$$y_{ict} = \beta_T T_{ct} + \beta_P P_{ct} + X_{ict}\theta + a_c + a_m + u_{ict},$$

where y_{ict} is a time-use indicator for individual i in Municipality c at period t, X_{ict} is a vector of individual characteristics (age and gender), a_i and a_m are Municipality and month-by-year fixed-effects, respectively, and e_{ict} is a set of unobservables that affect time-use patterns. We estimate the above expression with two subsamples: First, we employ data for youths 10 to 16 years of age in the last two waves of the ELCA. For them, we know how much time they spend on activities such as watching TV, reading for fun, or doing homework on a typical weekday.[17]. Given that these variables are reported in bins (nothing, less than one hour, etc.), we construct indicators equal to one if the youth spends one hour or more on each of the activities of interest. Second, we use data from the household heads and their spouses for all the years in the survey. In this case, we have information about the time (hours and minutes) spent on activities such as working on farms owned

[17] None of the options corresponds to on-farm or off-farm labor, which means that we are not able to test for changes in labor supply for this population group.

by the household, working on other households' farms in agricultural and non-agricultural activities (off-farm labor), and doing housework. With this information, we can compute the share of time each household member spends on different activities.

As before, we are interested in the effect of temperature on the outcome y_{ict}, which here corresponds to time-use indicators. Given the reference period of the time-use variables (typical weekday in the week before the survey), we recompute our temperature variable (T_{ct}) for each household to be the average daily maximum temperature in the reference week excluding weekends.[18] For consistency, we also recompute our precipitation variable (P_{ct}) to match this reference period. Hence, the identification assumption in the current setup is that temperature is as good as random conditional on the controls, thus the specification above includes Municipality and month-by-year fixed effects (Dell et al., 2014).

Table 2 shows the results using information for the youths, reporting separately the cases of rural and urban samples in Panels A and B, respectively. As can be seen, we find that a higher temperature leads to a statistically significant increase in the percentage of individuals in rural areas that spend one hour or more reading. We do not find evidence for changes in any of the other activities such as watching TV, spending time with parents or siblings, or doing homework. Interestingly, we do not find any statistically significant effects for individuals in the urban areas, including reading. Overall, our results suggest that students in rural and urban areas react differently to temperature stress, with the latter increasing the time spent on an activity potentially complementary to studying.

Even though the results so far point to changes in time use consistent with improvements in students' academic performance in rural areas, it is unclear what mechanisms are behind our findings. One possibility is that temperature stress incentivizes off-farm work, which may have higher returns to education overall. To

[18] Notice that the reference week is potentially household-specific. To simplify the notation, the time index t can be thought to correspond to the reference week instead of the semester as in equations (5)-(7).

test for this possibility, we estimate the model above with time-use data for adults in rural areas. We report the results for this exercise in Table 3, with panel A having the estimates for the household heads and panel B for their spouses.[19] As can be seen, our results suggest that an increase in temperature leads to a reallocation *towards* off-farm labor. In particular, we find that household heads (spouses) decrease (increase) the time spent working on farms owned by the household (other households) when temperatures are higher. The latter goes in line with findings across Africa and Asia, pointing to exogenous shocks in production and productive efficiency, including temperature, as a push for reallocating from on-farm labor out to non-farm activities (Colmer, 2021; Haggblade et al., 2007; Reardon et al., 2019). On the whole, we interpret these findings as suggestive of a shift in the allocation of labor in favor of activities potentially more intensive in human capital.

5.2 Incentives and temperature stress

Following our conceptual framework, exerting effort is costly as it exposes individuals to the effects of increasing temperatures. We use the exposure to SPP as an exogenous shock that increases expected wages, thus leading to potential increases in test scores, but with negative interaction effects with temperature.

In table 4, we summarize our estimates for equation (7). Besides a few small changes in point estimates related to precipitation, estimates of the first-order effects from the average maximum temperature in the year before the exam remain statistically equivalent to those of our baseline specifications (Appendix B, Tables B3 and B4). Meanwhile, the first-order effects are consistently positive across all scores, yet only statistically different from zero for math (all populations) and total scores (national level and urban students). On average, an additional 1 percentage point increase in the exposure to SPP at the Municipality level led to an increase of 2.96% of a standard deviation in math scores across all students. We estimate this effect to be 6.018% and 2.12% of a standard deviation for urban and rural students, respectively. For total scores, we detect an average

[19] We estimate different models for these two groups given the stark differences in the activities each of them seems to engage in (see table B6).

increase of 1.43% and 3.19% of a standard deviation for all students and urban students, respectively, due to an additional 1 percentage point exposure to SPP. These coefficients are aligned with the evidence in Bernal and Penney (2019) and Laajaj et al. (2022).

All but one of the estimated interaction effects follow the expected sign from our theoretical model, which is a piece of initial evidence in favor of our hypothesis of unobserved costs of exerting effort in a warming climate. Nonetheless, we only find this effect to be statistically significant in the case of math scores of urban students, which is precisely where SPP had the strongest effect on performance. On average, when a 1 percentage point increase in exposure to SPP takes place along with a 1°C increase in the average maximum temperature over the year before the Saber 11 test, there is a reduction of 0.10% of a standard deviation in the math section.[20] To put these numbers into perspective, our point estimates in column (2) suggest the temperature effect for students in municipalities with exposures in the 25th and 75th percentiles are -2.124 and -2.366, a 12.4% increase when going from the former to the latter (i.e., an interquartile change).[21]

As discussed in the previous section, our estimation includes interactions of yearly dummies with the Municipality-level dummies of quartiles of (a) the population of high-school age in the municipality, (b) the total administrative personnel, and (c) the total number of teachers, at 2014 values (our available pre-treatment period). We evaluate the robustness of our findings by comparing this "one step up" fixed effects estimation (Bilinski & Hatfield, 2018), with a standard fixed-effects estimation excluding said interactions. Estimates from usual fixed-effects estimation are statistically equivalent, although slightly less precise, under this alternate specification (see Appendix B, Table B7), thus providing evidence in favor of our overall empirical strategy.

[20] Based on the findings of Graff-Zivin et al. (2018), effects from heat shocks are more likely to be of larger magnitudes in math tests, as thermal insults primarily affect the performance of brain functions for tasks that are more complex than the average–which is arguably the case of math vis-à-vis other fields in standardized evaluations.

[21] This number can be calculated as [-0.105×(2.5-0)]/[-2.124]=12.4%, where 2.5 and 0 are the 75th and 25th percentiles of SPP exposure, respectively.

5.2.1 SPP targeted population as a source of heterogeneity

Based on the results above, and following our discussion about SPP, it is important to explore whether these interaction effects between temperature and incentives on effort vary across specific populations. While the previous findings indicate that SPP had larger direct and second-order (interaction) effects among urban students, we now focus our attention on who among those urban students were more largely impacted by the increased exposure to changes in temperature. Based on the eligibility criteria, we explore two main heterogeneities of interest, namely (a) the students' mothers' educational attainment, and (b) the reported strata of the house where a student's family lives.[22]

Parents' educational attainments are well known to be significant determinants of their offspring's own educational achievements and human capital accumulation (Björklund & Salvanes, 2011). It is such gaps that programs like SPP attempt to potentially address, by providing a window of opportunity to those in less favored backgrounds. In this sense, and in line with our conceptual framework, we would predict two stark differences: (a) exposition to SPP should have a larger effect among exam takers whose mothers have lower levels of education–i.e., a more significant change in incentives; and, as a result of this (b) the interaction effects of temperature and SPP exposure should also be larger (in absolute value) for students whose mothers have lower educational levels. To this purpose, we estimate our model in equation (7) on two sub-samples of urban exam-takers: those whose mothers have up to primary education, and those with secondary or post-secondary education.[23]

We summarize this first heterogeneity analysis in Panel A of Table 5. As suggested, we observe that there are larger effects from exposure to SPP for exam-takers with mothers that have lower educational attainment.

[22] Our focus on urban students is also motivated by the fact that the socioeconomic stratum has more variation in urban settings where population density is higher.

[23] We also do these estimations for both the overall population and for rural students alone (see Appendix B, Tables B8 and B9) but here we focus on urban students following our previous discussion on the potentially differentiated mechanisms at play between urban and rural populations.

Ultimately, as this effect is connected to a change in incentives, our finding implies that SPP successfully promoted additional effort among those from less favored backgrounds. Likewise, we see that most interaction effects between temperature and SPP exposure are of the expected sign, although only two of them are statistically significant: for math and total scores, for students with mothers that have lower levels of education, as hypothesized. In addition, notice that there are some interesting differences in the effect of temperature increases. While all estimates are of the expected sign, they seem larger for those with better household educational backgrounds–again consistent with our framework. We will return to this in the next subsection for our mechanisms analysis.

Since the enforcement of Colombian Law 142 of 1994, the country has followed a system of socioeconomic stratification that assigns a stratum number from 1 to 6 for a residence, which increases with the quality of the residence. The system is widely used as a proxy income, although it has also been found to serve as a potential source of social discrimination (Bogliacino, Jiménez Lozano, & Reyes, 2018). We implement it in our analysis to separate the sample of exam takers between two groups: those coming from households with potentially less-favored backgrounds (strata 1, 2, and 3) and those of potentially more favorable backgrounds (strata 4, 5, and 6). As seen in Panel B of Table 5, we find that exposure to SPP had statistically significant effects among exam takers from potentially less-favored backgrounds, which again follows the line of our conceptual framework. Similarly, we find that the interaction between temperature and exposure to SPP is only significant in the case of math for students from less-favored backgrounds.

Overall, we observe a consistent result among urban exam-takers. For starters, the effects of temperature and exposure to SPP are systematically larger and statistically significant in Saber 11 math scores than in other sections. On the one hand, these findings are in line with the literature suggesting that heat insults have larger effects on test performance and knowledge accumulation related to high-complexity tasks, which is arguably the case with math. On the other hand, given that admissions to high-return post-secondary programs (like

STEM or STEM-related fields (Kinsler & Pavan, 2015)) usually give a larger weight to math scores, the shock on incentives from SPP was expected to have a larger impact on math scores than in others sections, which is what we find. Finally, this combination of shocks leads to important interaction effects again concentrated in math, mainly changing the behavior of eligible students.

5.2.2 Potential Mechanism

Up to this point, we exploit the exogenous shock of SPP since, on average, it is tentatively increasing the expected wages or, equivalently, reducing the opportunity costs of accessing tertiary education (Londoño-Vélez et al., 2020). Consistent with this interpretation, we should expect that temperature increases have a larger effect on those who are likely to have lower opportunity costs for exerting effort. Access and completion of higher education are strongly correlated with a student's mother's educational attainment (Guzman Ruiz et al., 2009) and household income (Londoño-Vélez et al., 2020). Based on our available information, we can test whether temperature increases have increasingly negative effects on students' outcomes as their mother's educational attainment increases or as the reported stratum of the house is higher, which would be consistent with our theoretical model and empirical results so far.

Based on a specification that follows

$$100 \times y_{ijct} = \sum_{k=1}^{4} \beta_{EM(k),T}(T_{ct} \times EM_k) + \sum_{k=1}^{4} \beta_{EM(k)} EM_k + \beta_P P_{ct} + a_j + a_{st} + u_{ijct},$$

where y_{ijct} is the standardized score of interest, and EM_k is an index variable of whether the student's mother has maximum education attainment at level k. Namely, we have that k goes from 1 to 4, following if the mother has no educational attainment, primary education, secondary education, or postsecondary education. We present our fixed-effect estimates in Table 6, taking the no educational attainment cases as the base category (for results including coefficients for dummy variables alone see Appendix B, Table B10).

All estimates are consistent with our rationale. To begin with, the first-order coefficient estimates for the effect of temperature increases on test performance and those of our baseline estimates remain in the same ballpark. Also, on average, when mothers' educational attainments are larger, so are their children's test results–up to 51% of a standard deviation when compared to students whose mothers report no educational attainments. Finally, we find that all estimates of the second-order correlations of temperature increase relative to mothers' schooling are of the expected sign (negative), and all are increasingly negative with said education level. Moreover, this correlation is found across all subjects, at the aggregate level, and in urban and rural settings.

In addition, we also estimate an alternative specification following

$$100 \times y_{ijct} = \sum_{l=1}^{6} \beta_{HS(l),T}(T_{ct} \times HS_l) + \sum_{l=1}^{6} \beta_{HS(l)} HS_l + \beta_P P_{ct} + a_j + a_{st} + u_{ijct},$$

where HS_l is an index variable of whether the exam-taker dwelling is classified as stratum l, according to the Colombian residential strata system. In Table 7, we summarize fixed-effects estimates under this alternate specification, with strata 1 (lowest in the scale of dwelling quality) as the base category.[24] In line with all our results so far, we observe that temperature increases have a larger impact on the scores of students coming from potentially more favored backgrounds. For instance, take again the case of math scores of urban exam-takers. While an increase of 1°C in the average maximum temperature during the year leading to the exam day is connected to a reduction of 2.08% of a standard deviation for those living in a stratum 1 residence, the magnitude will increase by up to 1.38 percentage points–i.e., a rough total effect of 3.45% of a standard deviation–for those in a stratum 6 dwelling.

Ultimately, we see that although the direct effects of heat increases may suggest potentially different dynamics between rural and urban settings, the differentiated effects of increased heat suggest a highly consistent mechanism across all individuals in terms of opportunity cost. For those coming from a less favored back-

[24] Also see Appendix B, Table B11 for a detailed summary table including coefficients for dummies of each stratum.

ground (lower mothers' educational attainment or lower dwelling stratum), the opportunity cost of exerting effort in preparation for an exam like Saber 11 is comparatively high since their baseline expectations to access higher education are low. Conversely, those from more favored backgrounds will perceive a lower opportunity cost for exerting additional effort in studying, leading to a larger exposure to the effects of weather on average. It should be stressed that we do not expect the last two specifications to have a causal interpretation. For that, we would need exogenous variation in the mother's education attainments and residence stratum, which is beyond the scope of this paper. However, our estimates of $\beta_{EM(k),T}$ and $\beta_{HS(l)}$ are informative of the mechanisms behind our findings using the variation created by SPP.

6 Conclusion

We set up a theoretical model for human capital accumulation that is sensible to exogenous weather shocks, which suggests that heat has an overall negative effect on students' academic performance. Also, our model states that increases in effort should improve academic performance, but at the cost of higher exposure to thermal insults that can–as an interaction effect–reduce the performance of students. We use information from a National high-school exit exam in Colombia (Saber 11) between 2014-2019, precipitation and temperature data at the weather station level, and information on a policy intervention of a scholarship program in 2015 that serves as an exogenous shock to test these hypotheses.

Based on linear unobserved effect models and using fixed-effect regression methods, we demonstrate that 1°C increases in the average maximum temperature during the year before taking a high-stakes exam significantly reduce math, Spanish, and total scores. Our findings of negative shocks from increases in average temperature are consistent with the theoretical propositions and empirical results in the literature (Garg et al., 2020; Graff-Zivin et al., 2018; Park et al., 2021, 2020). Nevertheless, we highlight that longer-term exposures (e.g., over a calendar year) to increasing temperatures, may be of higher relevance than short-term exposures (e.g., exam day or week before the exam) for countries with predominantly tropical weather.

We provide quasi-experimental evidence supporting the hypothesis that exerting effort in high-stakes evaluations is increasingly costly for students as they are further exposed to weather shocks. While government programs like *Ser Pilo Paga* create incentives that increase students' outcomes, the efficacy of these interventions may be compromised under higher temperatures. We also evaluate a potential underlying mechanism supporting these findings, namely from the side of opportunity costs to access tertiary education (and likely expected wages). Students whose mothers achieved greater educational attainments, or who resided in houses of higher socioeconomic strata, have better scores in high-stakes exams but exhibit larger negative shocks from exposition to increased temperatures.

Among the limitations of our research, we encounter unexpected results from the effects of increasing temperatures on rural areas–namely, a positive effect in scores. We hypothesize that time reallocation in response to higher temperatures (Alberto et al., 2021; Graff-Zivin & Neidell, 2014) could play an important role in settings where child labor is high. We test this hypothesis with time-use data for youths and their parents in Colombia, finding a reallocation towards activities complementary to study (for youths) and non-farm work (for household heads and spouses). While these results follow our suggested mechanism, future research should explore these findings in more detail and for other contexts. Furthermore, we encourage research trying to understand the interaction between weather and social policies, which we believe will open the door to a more realistic evaluation of the impacts of climate change.

References

Abadie, A., Athey, S., Imbens, G., & Wooldridge, J. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35. doi: 10.1093/qje/qjac038

Agrawal, T., & Agrawal, A. (2019). Who gains more from education? A comparative analysis of business, farm and wage workers in India. *The Journal of Development Studies*, *55*(6), 1081–1098. doi: 10.1080/00220388.2018.1443209

Alberto, I. C., Jiao, Y., & Zhang, X. (2021). Too hot or too cold to study? The effect of temperature on student time allocation. *Economics of Education Review*, *84*, 102152. doi: 10.1016/j.econedurev.2021.102152

Angrist, J., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Auffhammer, M., Hsiang, S., Schlenker, W., & Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, *7*(2), 181–198. doi: 10.1093/reep/ret016

Auffhammer, M., & Kellogg, R. (2011). Clearing the air? The effects of gasoline content regulation on air quality. *American Economic Review*, *101*(6), 2687–2722. doi: 10.1257/aer.101.6.2687

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., & Shapiro, J. (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. doi: 10.1086/684582

Bernal, G., & Penney, J. (2019). Scholarships and student effort: Evidence from Colombia's Ser Pilo Paga program. *Economics of Education Review*, *72*, 121–130. doi: 10.1016/j.econedurev.2019.04.008

Bilinski, A., & Hatfield, L. A. (2018). *Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions.* doi: 10.48550/ARXIV.1805.03273

Björklund, A., & Salvanes, K. G. (2011). Education and Family Background. In *Handbook of the Economics of Education* (Vol. 3, pp. 201–247). Elsevier. doi: 10.1016/B978-0-444-53429-3.00003-X

Bogliacino, F., Jiménez Lozano, L., & Reyes, D. (2018). Socioeconomic stratification and stereotyping: lab-in-the-field evidence from Colombia. *International Review of Economics*, *65*(1), 77–118. doi: 10.1007/s12232-017-0285-4

Burke, M., Craxton, M., Kolstad, C., Onda, C., Allcott, H., Baker, E., … others (2016). Opportunities for advances in climate change economics. *Science*, *352*(6283), 292–293. doi: 10.1126/science.aad9634

Carleton, T., & Hsiang, S. (2016). Social and economic impacts of climate. *Science*, *353*(6304), aad9837. doi: 10.1126/science.aad9837

Carleton, T., Jina, A., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., … others (2022). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *The Quarterly Journal of Economics*, *137*(4), 2037–2105. doi: 10.1093/qje/qjac020

Chadi, A., De Pinto, M., & Schultze, G. (2019). Young, gifted and lazy? The role of ability and labor market prospects in student effort decisions. *Economics of Education Review*, *72*, 66–79. doi: 10.1016/j.econedurev.2019.04.004

Chetty, R., Friedman, J., & Rockoff, J. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. doi: 10.1257/

aer.104.9.2593
Chetty, R., Friedman, J., & Rockoff, J. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. doi: 10.1257/aer.104.9.2633
Cho, H. (2017). The effects of summer heat on academic achievement: A cohort analysis. *Journal of Environmental Economics and Management*, *83*, 185–196. doi: 10.1016/j.jeem.2017.03.005
Colmer, J. (2021). Temperature, labor reallocation, and industrial production: Evidence from India. *American Economic Journal: Applied Economics*, *13*(4), 101–124. doi: 10.2139/ssrn.3900866
Davis, B., Winters, P., Reardon, T., & Stamoulis, K. (2009). Rural nonfarm employment and farming: household-level linkages. *Agricultural Economics*, *40*(2), 119–123. doi: 10.1111/j.1574-0862.2009.00374.x
Dell, M., Jones, B., & Olken, B. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, *52*(3), 740–98. doi: 10.1257/jel.52.3.740
Fernández, M., Ibáñez, A. M., & Peña, X. (2014). Adjusting the labour supply to mitigate violent shocks: Evidence from rural Colombia. *The Journal of Development Studies*, *50*(8), 1135–1155. doi: 10.1080/00220388.2014.919384
Garg, T., Jagnani, M., & Taraz, V. (2020). Temperature and human capital in India. *Journal of the Association of Environmental and Resource Economists*, *7*(6), 1113–1150. doi: 10.1086/710066
Graff-Zivin, J., Hsiang, S., & 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. doi: 10.1086/694177
Graff-Zivin, J., & Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, *32*(1), 1–26. doi: 10.1086/671766
Graff-Zivin, J., Song, Y., Tang, Q., & Zhang, P. (2020). Temperature and high-stakes cognitive performance: Evidence from the national college entrance examination in China. *Journal of Environmental Economics and Management*, *104*, 102365. doi: 10.1016/j.jeem.2020.102365
Guzman Ruiz, C., Duran Muriel, D. M., Franco Gallego, J., Castaño Velez, E., Gallon Gomez, S., & Gomez Portilla, K. (2009). *Desercion estudiantil en la educacion superior colombiana: Metodologia de seguimiento, diagnóstico y elementos para su prevención*. Bogotá: Ministerio de Educacion Nacional. (OCLC: 777064247)
Haggblade, S., Hazell, P. B., & Reardon, T. (2007). *Transforming the rural nonfarm economy: Opportunities and threats in the developing world*. International Food Policy Research Institute.
Handel, D., & Hanushek, E. (2022). US school finance: Resources and outcomes. *NBER Working Paper*, *w30769*. doi: 10.3386/w30769
Hanushek, E. (2006). School resources. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 865–908). Elsevier.
Helo Sarmiento, J. (2023). Into the tropics: Temperature, mortality, and access to health care in Colombia. *Journal of Environmental Economics and Management*, *119*, 102796. doi: 10.1016/j.jeem.2023.102796
Hoffmann, B., Pulido, X., & Vera-Cossio, D. A. (2023). *The unequal effect of temperature on test scores: Evidence from Colombia.* (Inter-American Development Bank. Discussion Paper IDB-DP-1000.) doi: 10.18235/0004832
Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics*, *8*, 43–75. doi: 10.3386/w22181
Hsiang, S., & Kopp, R. (2018). An economist's guide to climate change science. *Journal of Economic*

Perspectives, *32*(4), 3–32. doi: 10.3386/w25189

Jackson, K., Johnson, R., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics*, *131*(1), 157–218. doi: 10.1093/qje/qjv036

Jolliffe, D. (2004). The impact of education in rural Ghana: Examining household labor allocation and returns on and off the farm. *Journal of Development Economics*, *73*(1), 287–314. doi: 10.1016/j.jdeveco.2003.02.002

Kinsler, J., & Pavan, R. (2015). The Specificity of General Human Capital: Evidence from College Major Choice. *Journal of Labor Economics*, *33*(4), 933–972. doi: 10.1086/681206

Laajaj, R., Moya, A., & Sánchez, F. (2022). Equality of opportunity and human capital accumulation: Motivational effect of a nationwide scholarship in Colombia. *Journal of Development Economics*, *154*, 102754. doi: 10.1016/j.jdeveco.2021.102754

Lenton, T. M., Xu, C., Abrams, J. F., Ghadiali, A., Loriani, S., Sakschewski, B., … Scheffer, M. (2023). Quantifying the human cost of global warming. *Nature Sustainability*, *6*, 1237–1247. doi: 10.1038/s41893-023-01132-6

Li, X., & Patel, P. (2021). Weather and high-stakes exam performance: Evidence from student-level administrative data in Brazil. *Economics Letters*, *199*, 109698. doi: 10.1016/j.econlet.2020.109698

Londoño-Vélez, J. (2022). The impact of diversity on perceptions of income distribution and preferences for redistribution. *Journal of Public Economics*, *214*, 104732. doi: 10.1016/j.jpubeco.2022.104732

Londoño-Vélez, J., Rodríguez, C., & Sánchez, F. (2020). Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser Pilo Paga in Colombia. *American Economic Journal: Economic Policy*, *12*(2), 193–227. doi: 10.1257/pol.20180131

Londoño-Vélez, J., Rodriguez, C., Sánchez, F., & Álvarez-Arango, L. (2023). Financial aid and social mobility: Evidence from Colombia's Ser Pilo Paga. *NBER Working Paper*, *31737*. doi: 10.3386/w31737

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. (2022). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, *232*(2), 272–299. doi: 10.1016/j.jeconom.2022.04.001

Medina, P., Ariza, N., Navas, P., Rojas, F., Parody, G., Valdivia, J. A., … Penagos, J. F. (2018). An unintended effect of financing the university education of the most brilliant and poorest Colombian students: The case of the intervention of the Ser Pilo Paga program. *Complexity*, *2018*, 1–9. doi: 10.1155/2018/3528206

Mora, A., & Ruiz, L. M. (2019). "Ser pilo no paga": Privatización, desigualdad y desfinanciamiento de la universidad pública en Colombia. *Ciencia Política*, *14*(27), 115–142. doi: 10.15446/cp.v14n27.73369

Park, J. (2022). Hot temperature and high stakes performance. *Journal of Human Resources*, *57*(2), 400–434. doi: 10.3368/jhr.57.2.0618-9535R3

Park, J., Behrer, P., & Goodman, J. (2021). Learning is inhibited by heat exposure, both internationally and within the United States. *Nature Human Behaviour*, *5*(1), 19–27. doi: 10.1038/s41562-020-00959-9

Park, J., Goodman, J., Hurwitz, M., & Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy*, *12*(2), 306–39. doi: 10.1257/pol.20180612

Ramoni-Perazzi, J., Orlandoni-Merli, G., Castillo-Paredes, L., & Peña Guillén, J. A. (2021). Child labor in Colombia: factors affecting the selection of economic activity. *Revista de Economía del Rosario*, *24*(2). doi: 10.12804/revistas.urosario.edu.co/economia/a.9086

Reardon, T., Echeverria, R., Berdegué, J., Minten, B., Liverpool-Tasie, S., Tschirley, D., & Zilberman, D. (2019). Rapid transformation of food systems in developing regions: Highlighting the role of agricultural research & innovations. *Agricultural Systems, 172*, 47-59. (Agricultural research for rural prosperity: Rethinking the pathways) doi: 10.1016/j.agsy.2018.01.022

Reardon, T., Stamoulis, K., & Pingali, P. (2007). Rural nonfarm employment in developing countries in an era of globalization. *Agricultural Economics, 37*(s1), 173-183. doi: 10.1111/j.1574-0862.2007.00243.x

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights, 4*(3), 305–322. doi: 10.1257/aeri.20210236

Shah, M., & Steinberg, B. (2021). Workfare and human capital investment: Evidence from India. *Journal of Human Resources, 56*(2), 380–405. doi: 10.3368/jhr.56.2.1117-9201R2

Statista Research Department. (2023). *Household penetration rate of selected household appliances in Colombia in 2018.* Statista. (Retrieved from https://www.statista.com/statistics/1007425/household-penetration-appliances-colombia/)

Sung, S. (2022). *Household penetration rate of home appliances in India in 2018.* Statista. (Retrieved from https://www.statista.com/statistics/370635/household-penetration-home-appliances-india/)

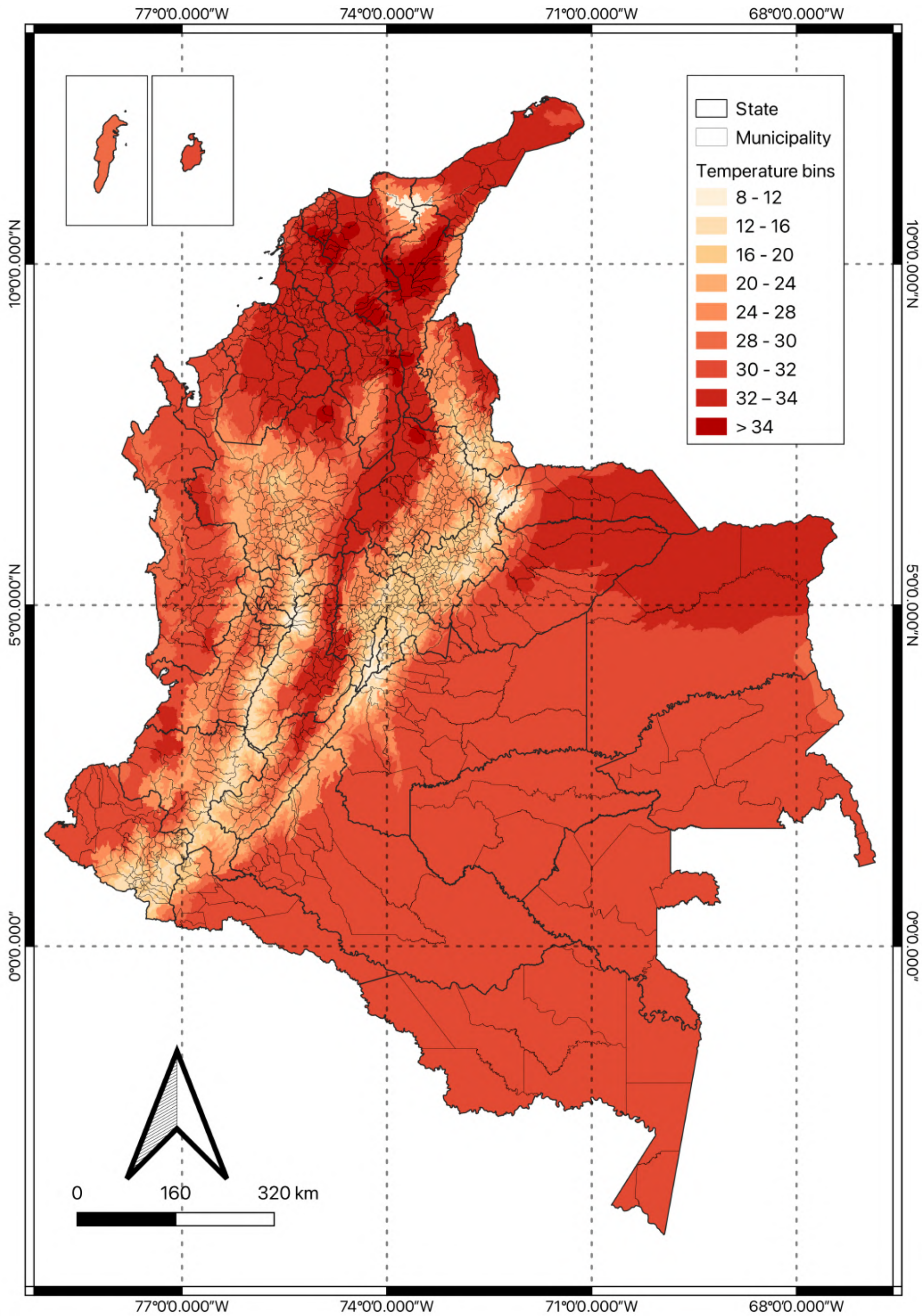
Wendt, D., Van Loon, L., & Marken Lichtenbelt, W. (2007). Thermoregulation during exercise in the heat: Strategies for maintaining health and performance. *Sports Medicine, 37*(8), 669–682. doi: 10.2165/00007256-200737080-00002

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed ed.). Cambridge, Mass: MIT Press.

Zhang, X., Chen, X., & Zhang, X. (2024). Temperature and low-stakes cognitive performance. *Journal of the Association of Environmental and Resource Economists, 11*(1), 75-96. doi: 10.1086/726007

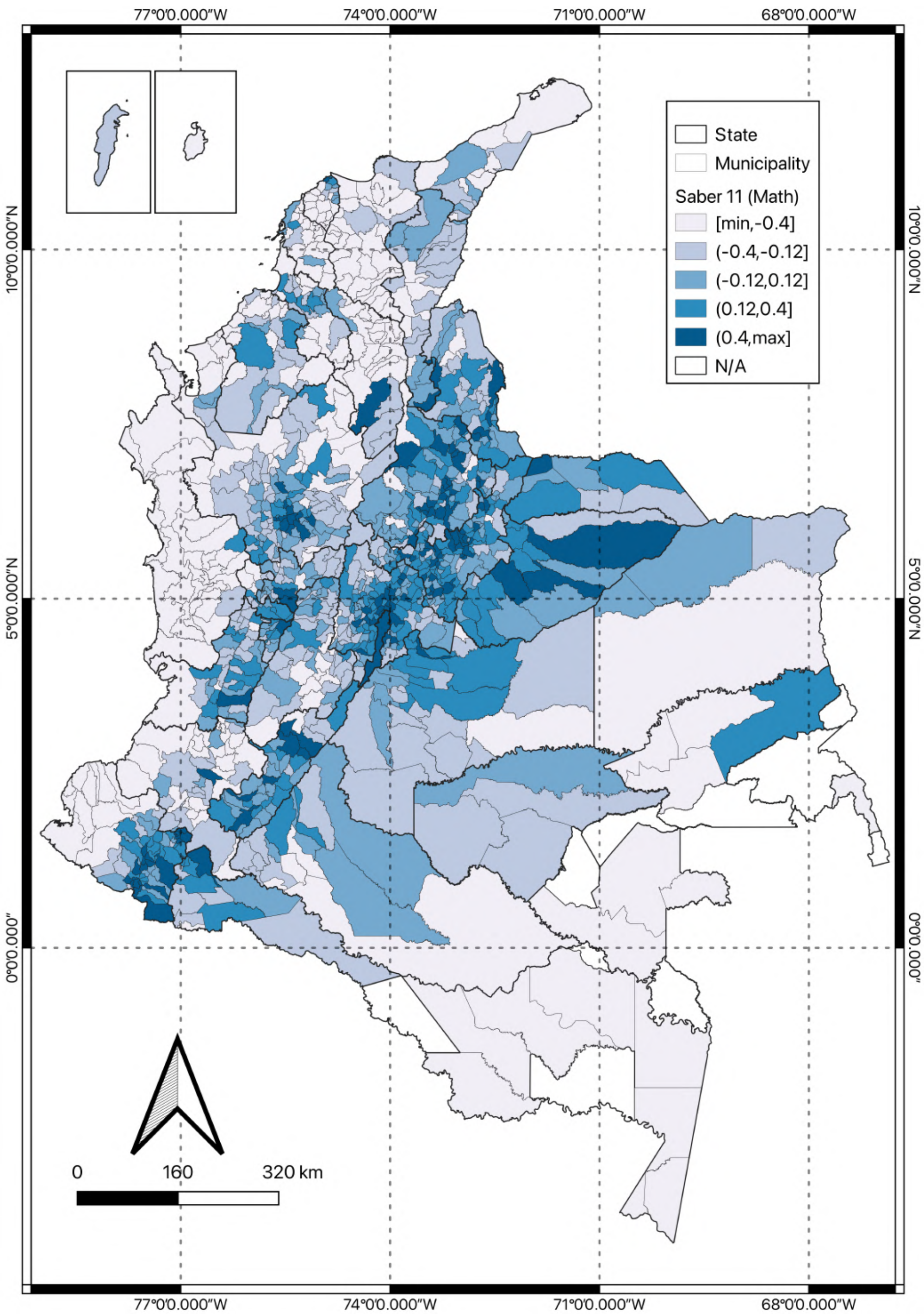
Figures and Tables

Figure 1: Municipality-level historical average of daily maximum temperature, 1981–2010.



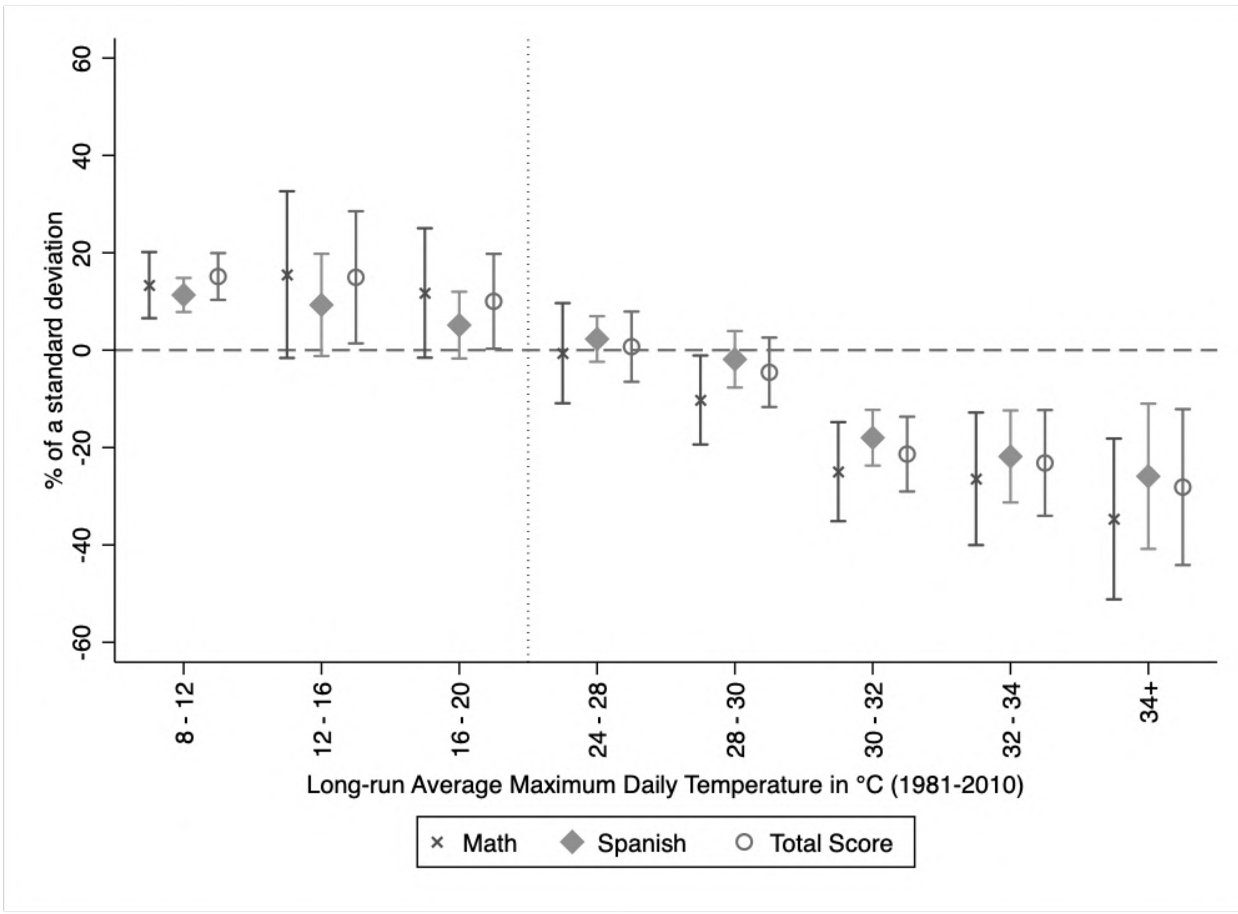
Notes — Figure displays the average of the daily maximum temperature (°C) at the Municipality level for the period 1981-2010 (i.e., climate normal). Raw weather data is publicly available from the Colombian Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM).

Figure 2: Municipality-level average Saber 11 standardized score in math, 2014-II–2019-I.



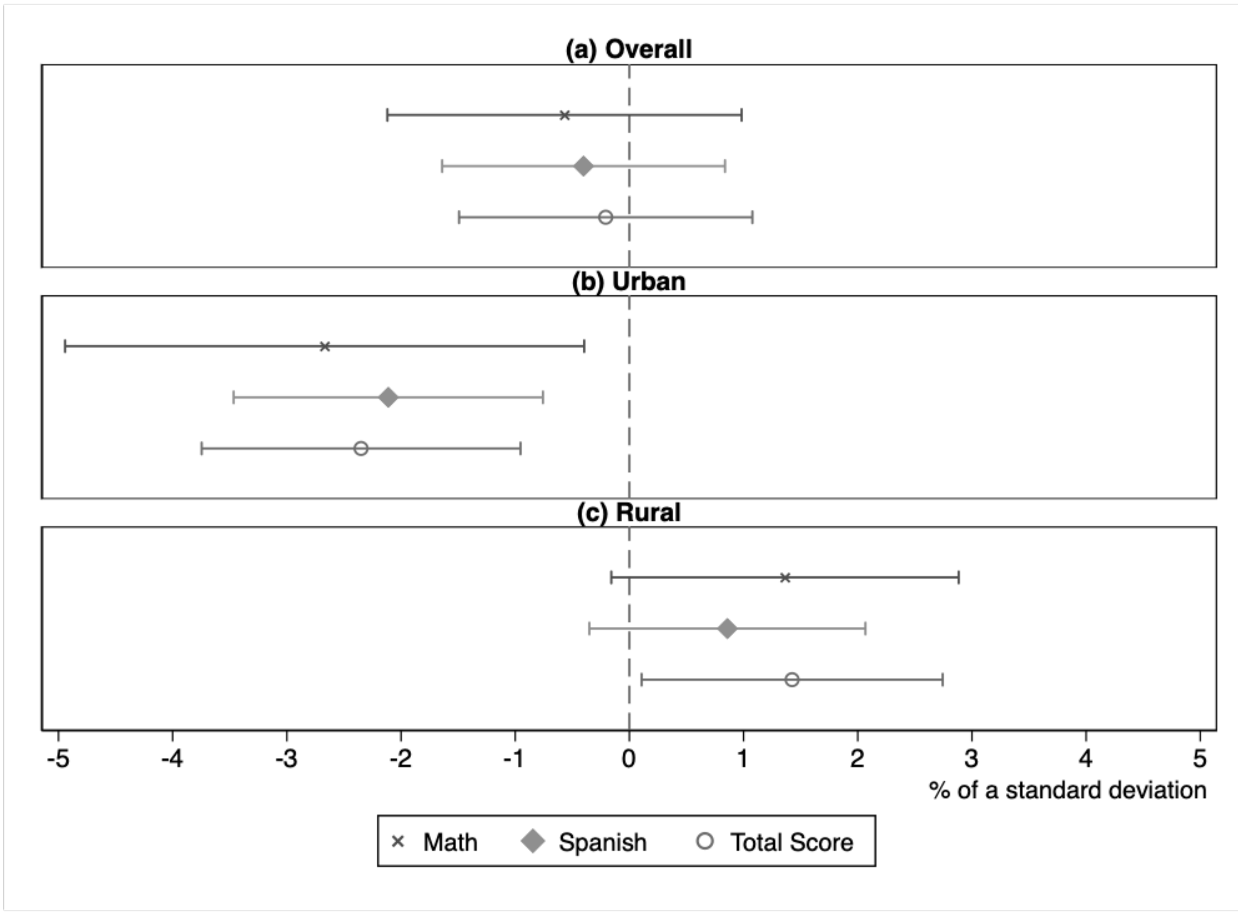
Notes — The figure shows the Municipality-level average standardized Saber 11 score in math for the tests between 2014-II and 2019-I. Student-level information is publicly available from the Colombian Institute for the Evaluation of Education (ICFES).

Figure 3: Correlation of Municipality-level historical average of daily maximum temperature (1981–2010) and Saber 11 standardized scores (2014-II–2019-I)



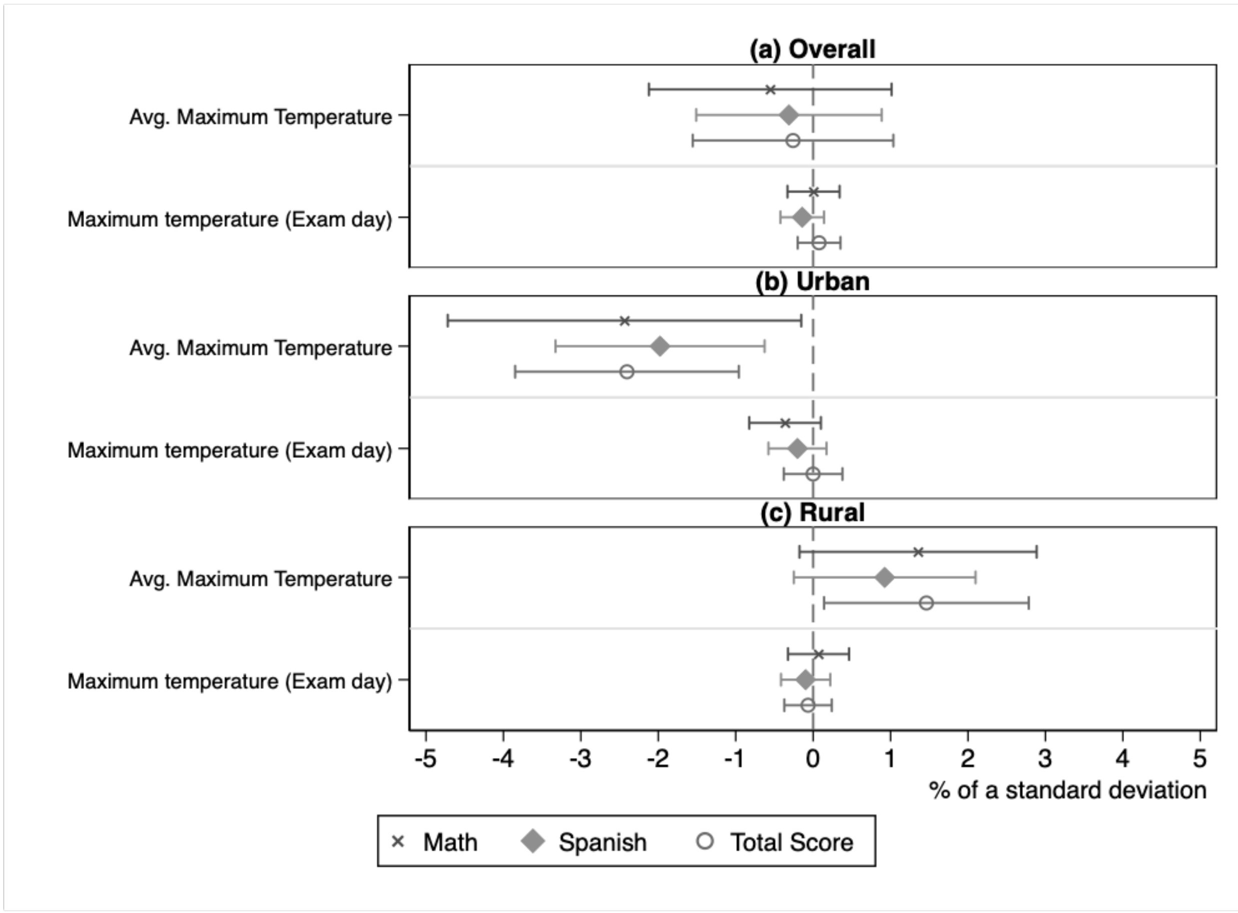
Notes — Reporting Pooled-OLS coefficient estimates (and 95% confidence intervals) for dummy variables of the long-run average maximum temperature (AMT) experienced in the Municipality, captured in temperature bins. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Estimates capture the difference in scores for Municipalities with a long-run AMT in the bin relative to Municipalities that experienced a long-run AMT between 20 and 24°C. Exam-taker attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Figure 4: Effect of the past-year average maximum temperature on Saber 11 scores, 2014-II–2019-I.



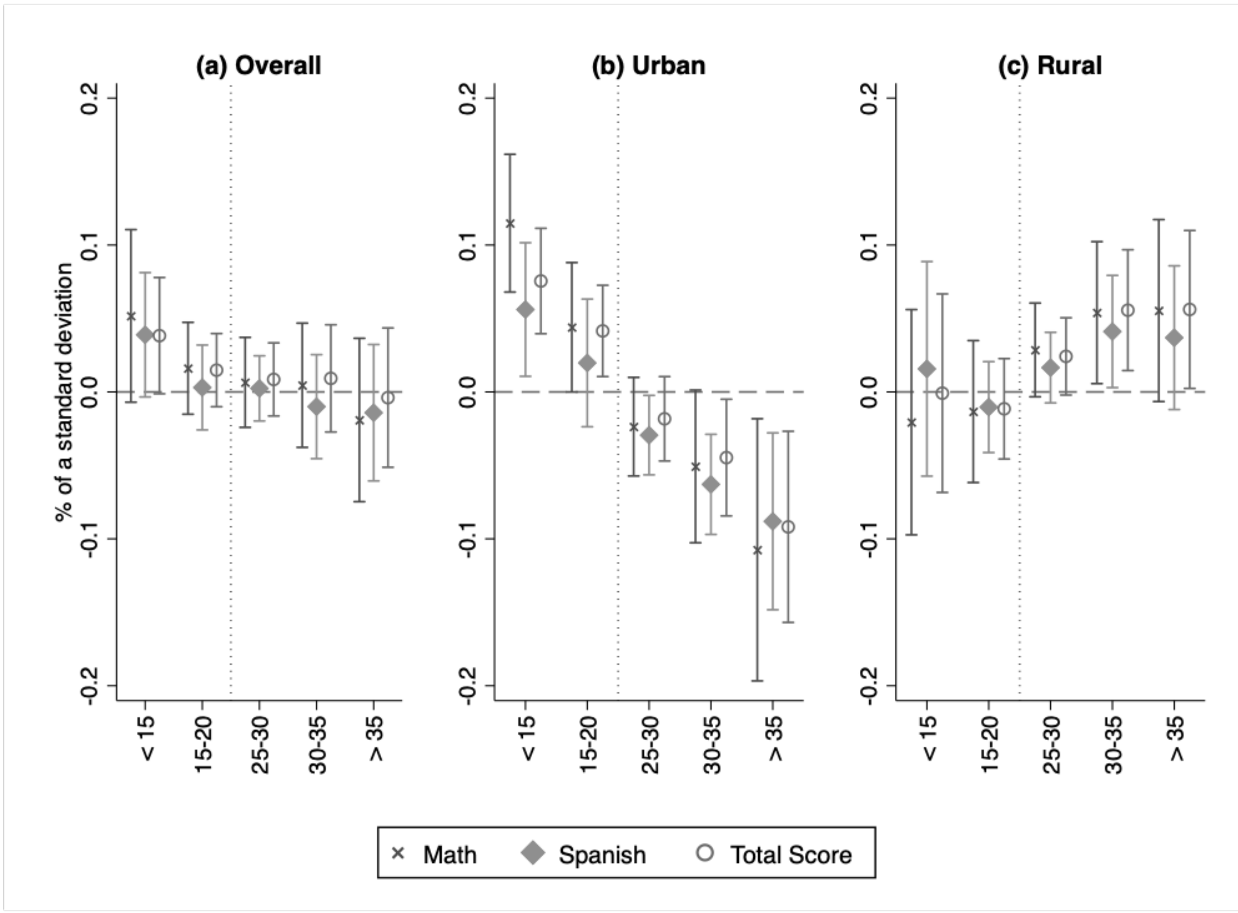
Notes — Reporting Fixed-Effects estimation of the effect (and 95% confidence interval) from a 1°C increase in the average daily maximum temperature at the Municipality level during the calendar before the Saber 11 test on the scores achieved by students between 2014-II and 2019-I. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Regressions include controls of precipitation at the Municipality level for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Figure 5: Effect of the past-year average maximum and exam date maximum temperatures on Saber 11 scores, 2014-II–2019-I.



Notes — Reporting Fixed-Effects estimation of the effect (and confidence interval) from 1°C increases in the exam-day maximum temperature and the average daily maximum temperature at the Municipality level during the calendar before the Saber 11 test on the scores achieved by students between 2014-II and 2019-I. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Regressions include controls of precipitation at the Municipality level for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Figure 6: Effect of the past-year number of days exposure to maximum temperatures on Saber 11 scores, 2014-II–2019-I.



Notes — Reporting Fixed-Effects estimation of the effect (and confidence interval) from 1 additional day exposition to maximum temperatures in the range relative to days with maximum temperatures between 20°and 25°C at the Municipality level during the calendar before the Saber 11 test on the scores achieved by students between 2014-II and 2019-I. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Regressions include controls of precipitation at the Municipality level for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table 1: Descriptive statistics of the sample and Municipality-level variables.

	Mean	Median	S.D.
STUDENT-LEVEL INFORMATION (N=2,202,917)			
Saber 11 scores			
Mathematics	52.2	52	11.5
Urban	53.7	52	11.6
Rural	48.9	49	10.8
Spanish	53.1	53	9.7
Urban	54.5	54	9.7
Rural	50.1	50	9.1
Total	261.5	258	47.3
Urban	268.7	266	47.6
Rural	245.8	242	42.6
Independent variables			
Age	17.29	17.08	0.99
Female (1=Yes)	0.55	1	0.5
Mother: Complete primary (1=Yes)	0.29	0	0.45
Mother: Complete secondary (1=Yes)	0.33	0	0.47
Mother: Complete postsecondary (1=Yes)	0.23	0	0.42
Ethnic minority (1=Yes)	0.06	0	0.24
Lives and studies in different Municipalities (1=Yes)	0.03	0	0.17
House Strata: Low (1 = Residence stratum is 1, 2, or 3)	0.925	1	0.26
House Strata: High (1 = Residence stratum is 4, 5, or 6)	0.075	0	0.26
MUNICIPALITY-LEVEL INFORMATION (N=1,052)			
Ser Pilo Paga (SPP) 2015			
Number of beneficiaries	8.61	2	59.27
Exposure to SPP	0.017	0.012	0.02
Weather variables			
Average maximum temperature (°C)	26.66	27.81	6.19
Maximum temperature (°C, exam days)	26.87	28.00	6.77
Long-run average maximum temperature above 30°C (1=Yes)	0.34	0	0.47
Mean precipitation (mm/day)	5.52	4.35	3.97
Precipitation (mm, exam days)	4.66	0.30	9.89
Days per temperature bin			
DD < 15	6.0	0.0	20.5
DD $15 - 20$	72.2	0.0	124.2
DD $20 - 25$	67.1	8.0	104.7
DD $25 - 30$	74.8	23.1	96.2
DD $30 - 35$	115.2	36.1	122.7
DD ≥ 35	30.7	0.0	58.0

Notes — Student-level information is publicly available from the Colombian Institute for the Evaluation of Education (ICFES). The total number of test takers in urban Municipalities is 1,517,000, while the remaining 685,917 are located in rural Municipalities. Age is calculated as the exact number of years from date of birth to date of exam. Female and Ethnic minority dummies are based on self-identification from test takers. Mothers' education attainment dummies are mutually exclusive and only consider schooling levels fully completed. Anonymized information of beneficiaries of *Ser Pilo Paga* (SPP) was provided by the Colombian Ministry of Education, which can be matched at the school level. The number of beneficiaries relates only to the number of assigned scholarships within the Municipality in the first round of SPP. The variable Exposure to SPP is the ratio of beneficiaries in the Municipality over the total number of test-takers within that first round period of SPP. Raw weather data is publicly available from the Colombian Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM).

Table 2: Effect of the past-year average maximum temperature on time use for youths (10 to 16 years of age), ELCA (2013 and 2016).

Panel A: Rural

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	TV	Computer	Parents	Play	Chores	Exercise	Read	Homework
Avg. maximum temperature (AMT)	-0.013	0.002	-0.004	-0.014	0.012	0.011	0.018*	0.001
	(0.012)	(0.007)	(0.013)	(0.009)	(0.011)	(0.009)	(0.009)	(0.009)
Mean precipitation	-0.001	-0.001	0.002	-0.003	-0.004**	-0.002	-0.001	0.000
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Observations	3,197	3,197	3,197	3,197	3,197	3,197	3,197	3,197
R^2	0.046	0.061	0.063	0.070	0.070	0.081	0.037	0.047

Panel B: Urban

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	TV	Computer	Parents	Play	Chores	Exercise	Read	Homework
Avg. maximum temperature (AMT)	0.006	0.018	-0.002	-0.009	-0.004	0.012	-0.007	0.003
	(0.010)	(0.014)	(0.011)	(0.013)	(0.011)	(0.010)	(0.007)	(0.009)
Mean precipitation	0.002	-0.001	0.002	0.001	0.002	0.001	-0.000	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Observations	3,122	3,122	3,122	3,122	3,122	3,122	3,122	3,122
R^2	0.065	0.098	0.060	0.076	0.084	0.109	0.062	0.068
Month-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes — Cluster (Municipality level) robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The dependent variables are dummies equal to one if the respondent spent one hour or more in each of the activities mentioned in the column header. Avg. maximum temperature captures the average of the daily maximum temperature (in °C) during the week before the survey for each household (the reference period for the time-use variables). Mean precipitation is the average daily precipitation (mm/day) also for the same reference period. Attribute controls added to the regression include age dummies and a dummy variable for whether the individual is female.

Table 3: Effect of the past-year average maximum temperature on time use for household heads and spouses, ELCA (2010, 2013, and 2016).

	Panel A: Household Heads				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Own-farm	Off-farm	Commuting	Chores	Leisure
Avg. maximum temperature (AMT)	-0.620*	0.355	0.014	0.253	0.011
	(0.321)	(0.397)	(0.049)	(0.181)	(0.261)
Mean precipitation	-0.014	0.018	-0.013	-0.019	0.023
	(0.049)	(0.055)	(0.008)	(0.026)	(0.051)
OBSERVATIONS	12,082	12,082	12,082	12,082	12,082
R^2	0.127	0.195	0.065	0.575	0.133

	Panel B: Spouses				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Own-farm	Off-farm	Commuting	Chores	Leisure
Avg. maximum temperature (AMT)	-0.104	0.285**	0.033	-0.009	-0.195
	(0.111)	(0.131)	(0.021)	(0.293)	(0.352)
Mean precipitation	-0.013	0.057**	0.000	-0.006	-0.038
	(0.029)	(0.022)	(0.005)	(0.051)	(0.038)
OBSERVATIONS	9,091	9,091	9,091	9,091	9,091
R^2	0.171	0.135	0.067	0.262	0.118
MONTH-BY-YEAR FE	Yes	Yes	Yes	Yes	Yes
MUNICIPALITY FE	Yes	Yes	Yes	Yes	Yes

Notes — Cluster (Municipality level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variables are the percentages of time spent on each activity mentioned in the column header. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the week before the survey for each household (the reference period for the time-use variables). Mean precipitation is the average daily precipitation (mm/day) for the same reference period. Attribute controls added to the regression include age and a dummy variable for whether the individual is female.

Table 4: Heterogeneous effect of the past-year average maximum temperature on Saber 11 scores by exposure to *Ser Pilo Paga*, 2014-II–2019-I.

Variables	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. maximum temperature (AMT)	-0.506	-2.124**	1.407*	-0.558	-1.929***	0.883	-0.268	-2.064***	1.456**
	(0.703)	(0.838)	(0.776)	(0.577)	(0.720)	(0.627)	(0.598)	(0.532)	(0.685)
Exposure to SPP × Post	2.962***	6.178***	2.120*	0.0879	0.425	-0.0544	1.428**	3.196**	0.815
	(0.987)	(1.845)	(1.092)	(0.639)	(1.087)	(0.692)	(0.721)	(1.302)	(0.820)
AMT × Exposure to SPP × Post	-0.0115	-0.105**	-0.0107	-0.00577	0.0248	-0.0231	-0.00858	-0.0332	-0.0135
	(0.0358)	(0.0488)	(0.0397)	(0.0231)	(0.0365)	(0.0238)	(0.0254)	(0.0379)	(0.0281)
Mean precipitation	-0.107	-0.330	0.0793	-0.127	-0.649**	0.0825	-0.102	-0.278	-0.0142
	(0.235)	(0.468)	(0.201)	(0.177)	(0.315)	(0.166)	(0.149)	(0.251)	(0.154)
Mean precipitation × Exposure to SPP × Post	-0.0888	-0.277*	-0.0349	-0.0224	0.00982	-0.0374	-0.0578	-0.183	-0.0234
	(0.0693)	(0.165)	(0.0727)	(0.0436)	(0.100)	(0.0507)	(0.0524)	(0.121)	(0.0562)
Observations	2,175,278	1,506,211	669,066	2,175,278	1,506,211	669,066	2,175,278	1,506,211	669,066
R^2	0.392	0.384	0.337	0.334	0.312	0.284	0.442	0.430	0.371
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attribute Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneous Trends[(a)]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities. (a) Trend heterogeneity is allowed by adding interactions between period (year-semester) dummies and quartiles of Municipality-level variables (namely Total High-School Students, Total Number of School Teachers, and Total Number of School Administrative Personnel).

Table 5: Heterogeneous effect of the past-year average maximum temperature on Saber 11 scores by exposure to *Ser Pilo Paga* in Urban population under specific house strata and mother's educational attainment levels, 2014-II–2019-I.

Panel A: Differences by Mother's Educational Attainment

	Math		Spanish		Total Score	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature (AMT)	-1.319*	-2.557*	-1.014	-2.397**	-0.712	-2.808***
	(0.719)	(1.348)	(1.046)	(0.944)	(0.727)	(0.877)
Exposure to SPP × Post	8.024***	4.303**	2.728*	-1.535	5.730***	0.975
	(2.070)	(2.008)	(1.404)	(1.639)	(1.505)	(1.689)
AMT × Exposure to SPP × Post	-0.154***	-0.0576	-0.0387	0.0770	-0.105**	0.0269
	(0.0530)	(0.0593)	(0.0425)	(0.0540)	(0.0422)	(0.0511)
Mother's Education	Low	High	Low	High	Low	High
Observations	550,045	955,949	550,045	955,949	550,045	955,949
R^2	0.256	0.386	0.205	0.301	0.278	0.427
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneous Trends[(a)]	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Differences by House Strata

	Math		Spanish		Total Score	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature (AMT)	-2.157***	0.230	-1.925***	-1.545	-1.964***	-2.198
	(0.789)	(3.630)	(0.702)	(2.122)	(0.524)	(2.526)
Exposure to SPP × Post	6.584***	0.431	0.850	-2.798	3.722***	-3.052
	(1.835)	(4.352)	(1.029)	(3.027)	(1.286)	(2.922)
AMT × Exposure to SPP × Post	-0.106**	-0.0446	0.0133	0.137	-0.0463	0.149
	(0.0480)	(0.136)	(0.0346)	(0.123)	(0.0373)	(0.108)
House Strata	Low	High	Low	High	Low	High
Observations	1,354,424	150,960	1,354,424	150,960	1,354,424	150,960
R^2	0.320	0.485	0.263	0.370	0.357	0.531
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneous Trends[(a)]	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average of the daily maximum temperature (in °C) during the calendar year before the exam within each round. All regressions include mean precipitation (not reported), measured as the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities. (a) Trend heterogeneity is allowed by adding interactions between period (year-semester) dummies and quartiles of Municipality-level variables (namely Total High-School Students, Total Number of School Teachers, and Total Number of School Administrative Personnel).

Table 6: Heterogeneous effect of past-year average maximum temperature on Saber 11 scores by education of the mother, 2014-II–2019-I.

Variables	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. maximum temperature (AMT)	-0.302	-2.335**	1.575**	-0.143	-1.769**	1.045*	0.0779	-1.971***	1.630**
	(0.776)	(1.118)	(0.769)	(0.616)	(0.680)	(0.613)	(0.640)	(0.690)	(0.667)
AMT × Mother: Complete primary	-0.154***	-0.166***	-0.100*	-0.143***	-0.183***	-0.0701	-0.166***	-0.199***	-0.0971**
	(0.0542)	(0.0595)	(0.0563)	(0.0455)	(0.0477)	(0.0520)	(0.0518)	(0.0570)	(0.0478)
AMT × Mother: Complete secondary	-0.392***	-0.349***	-0.498***	-0.383***	-0.361***	-0.460***	-0.412***	-0.395***	-0.471***
	(0.0604)	(0.0595)	(0.0978)	(0.0529)	(0.0551)	(0.0885)	(0.0540)	(0.0611)	(0.0878)
AMT × Mother: Complete tertiary	-0.526***	-0.494***	-0.718***	-0.512***	-0.499***	-0.671***	-0.575***	-0.559***	-0.730***
	(0.143)	(0.130)	(0.139)	(0.102)	(0.0737)	(0.124)	(0.117)	(0.109)	(0.125)
Observations	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150
R^2	0.392	0.384	0.337	0.334	0.312	0.284	0.442	0.430	0.371
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attribute Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average of the daily maximum temperature (in °C) during the calendar year before the exam within each round. All regressions include mean precipitation (not reported), measured as the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table 7: Heterogeneous effect of past-year average maximum temperature on Saber 11 scores by house strata level, 2014-II–2019-I.

VARIABLES	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. maximum temperature (AMT)	-0.174	-2.079*	1.575**	-0.0681	-1.633**	1.040*	0.157	-1.793**	1.608**
	(0.760)	(1.120)	(0.763)	(0.608)	(0.670)	(0.594)	(0.626)	(0.686)	(0.653)
AMT × House Strata: Level 2	-0.564***	-0.555***	-0.533***	-0.461***	-0.441***	-0.439***	-0.505***	-0.513***	-0.448***
	(0.0572)	(0.0649)	(0.0827)	(0.0571)	(0.0651)	(0.0761)	(0.0599)	(0.0694)	(0.0735)
AMT × House Strata: Level 3	-0.878***	-0.786***	-0.986***	-0.794***	-0.670***	-1.046***	-0.844***	-0.755***	-0.995***
	(0.119)	(0.145)	(0.182)	(0.110)	(0.136)	(0.171)	(0.117)	(0.145)	(0.169)
AMT × House Strata: Level 4	-1.392***	-1.165***	-1.559***	-1.165***	-0.913***	-1.454***	-1.339***	-1.125***	-1.442***
	(0.200)	(0.203)	(0.347)	(0.163)	(0.157)	(0.399)	(0.180)	(0.178)	(0.352)
AMT × House Strata: Level 5	-1.374***	-1.186***	-0.619**	-1.211***	-0.942***	-1.036***	-1.347***	-1.135***	-0.796***
	(0.249)	(0.275)	(0.312)	(0.184)	(0.181)	(0.295)	(0.224)	(0.240)	(0.292)
AMT × House Strata: Level 6	-1.619***	-1.377***	-0.990**	-1.338***	-1.110***	-0.365	-1.550***	-1.325***	-0.627*
	(0.279)	(0.263)	(0.477)	(0.209)	(0.208)	(0.370)	(0.244)	(0.231)	(0.327)
OBSERVATIONS	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150
R^2	0.392	0.384	0.337	0.334	0.312	0.284	0.442	0.430	0.372
SCHOOL FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TIME-STATE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATTRIBUTE CONTROLS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SAMPLE	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average of the daily maximum temperature (in °C) during the calendar year before the exam within each round. All regressions include mean precipitation (not reported), measured as the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam-taker lives and studies in different Municipalities.

Appendix A Optimization Problem

We assume functional forms such that

$$f(e,t) = \frac{e^{\alpha}}{t^{\beta}}, \quad 0 < \alpha < 1, \quad \beta > 0;$$

$$g(e,t) = e^{\gamma} t^{\zeta}, \quad \gamma > 1, \quad 0 < \zeta < 1;$$

$$U(c, g(e,t)) = c - g(e,t) = wzf(e,t) - g(e,t).$$

Hence, the individual's problem is reduced to

$$\max_{e} \ wz\frac{e^{\alpha}}{t^{\beta}} - e^{\gamma} t^{\zeta},$$

where the first-order condition (FOC) implies that

$$\alpha wz\frac{e^{\alpha-1}}{t^{\beta}} - \gamma e^{\gamma-1} t^{\zeta} = 0,$$

while the second-order condition for maximization

$$\alpha(\alpha-1)wz\frac{e^{\alpha-2}}{t^{\beta}} - \gamma(\gamma-1)e^{\gamma-2} t^{\zeta} < 0$$

is satisfied when $0 < \alpha \leq 1$ and $\gamma > 1$, and thus $\gamma > \alpha$; hence, we must assume that the marginal cost of effort increases faster than its marginal benefit.

From the FOC we can recover the optimal level of effort, e^*, which follows

$$e^* = \left(\frac{\alpha wz}{\gamma t^{\beta+\zeta}}\right)^{\frac{1}{\gamma-\alpha}}.$$

Therefore, the production of human capital under an optimal effort level is

$$\begin{aligned} h^*(t,w,z) &= z\frac{(e^*)^{\alpha}}{t^{\beta}} \\ &= t^{-\beta-\frac{\alpha(\beta+\zeta)}{\gamma-\alpha}} w^{\frac{\alpha}{\gamma-\alpha}} z^{\frac{\gamma}{\gamma-\alpha}} \left(\frac{\alpha}{\gamma}\right)^{\frac{\alpha}{\gamma-\alpha}}. \end{aligned}$$

Finally, linearizing this expression with a second-order Taylor expansion, we have that

$$\begin{aligned} h^*(t,w,z) = h^*(t',w',z') &+ \left(\frac{t-t'}{t'}\right) h^*(t',w',z') \left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha}\right] \\ &+ \left(\frac{w-w'}{w'}\right) h^*(t',w',z') \left[\frac{\alpha}{\gamma-\alpha}\right] + \left(\frac{z-z'}{z'}\right) h^*(t',w',z') \left[\frac{\gamma}{\gamma-\alpha}\right] \\ &+ \frac{1}{2}\left(\frac{t-t'}{t'}\right)\left(\frac{w-w'}{w'}\right) h^*(t',w',z') \left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha}\right]\left[\frac{\alpha}{\gamma-\alpha}\right] \\ &+ \frac{1}{2}\left(\frac{t-t'}{t'}\right)\left(\frac{z-z'}{z'}\right) h^*(t',w',z') \left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha}\right]\left[\frac{\gamma}{\gamma-\alpha}\right] + R_2(t,z,w), \end{aligned}$$

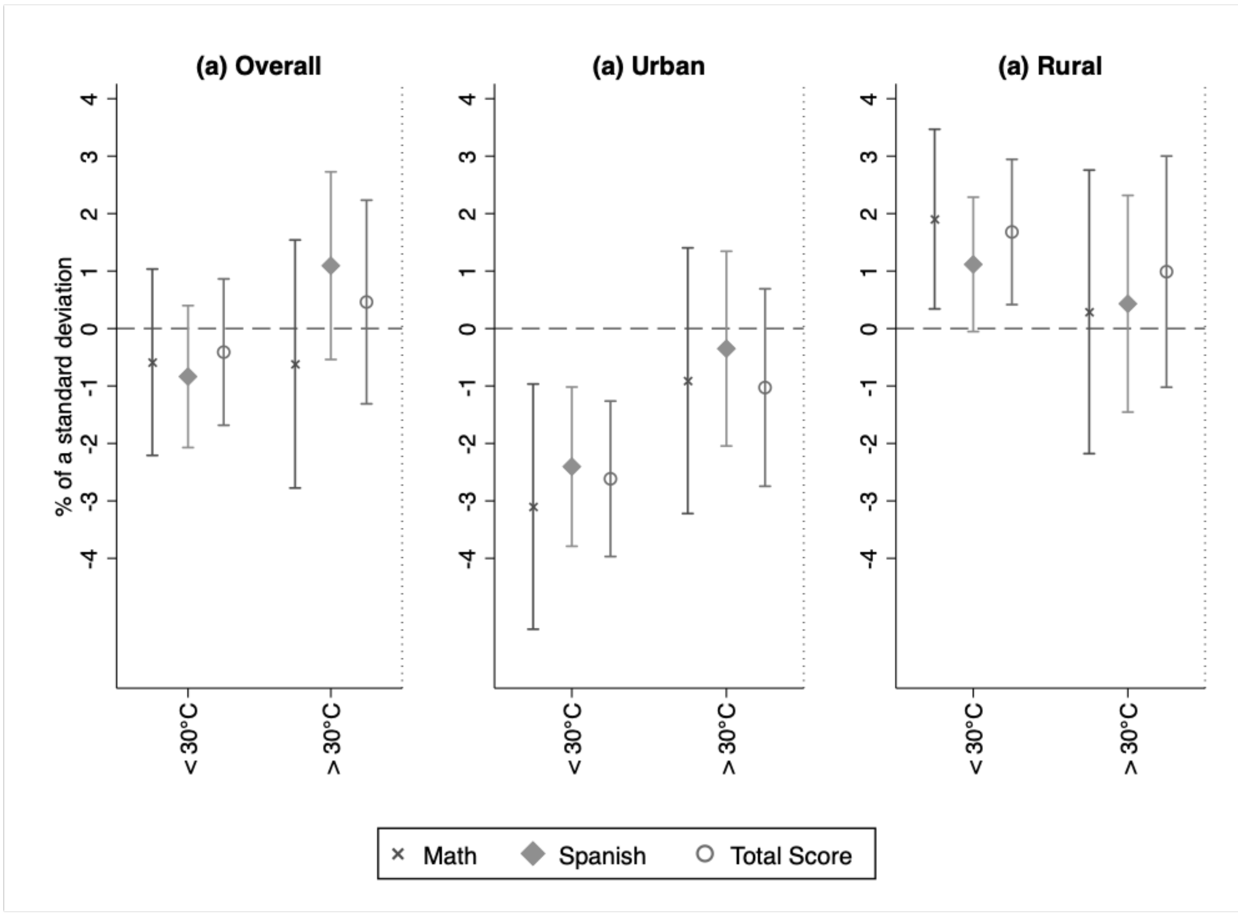
where $R_2(t,z,w)$ includes the cross-effect between wages and cognitive abilities and other own- and cross-effects of second and higher orders that are approximate to zero. Then, the percent deviation of human capital with respect to the mean follows:

$$\begin{aligned} \hat{h} = \left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha}\right]\hat{t} + \left[\frac{\alpha}{\gamma-\alpha}\right]\hat{w} + \left[\frac{\gamma}{\gamma-\alpha}\right]\hat{z} + \frac{1}{2}\left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha}\right]\left[\frac{\alpha}{\gamma-\alpha}\right]\hat{t}\hat{w} \\ + \frac{1}{2}\left[-\beta - \frac{\alpha(\beta+\zeta)}{\gamma-\alpha}\right]\left[\frac{\gamma}{\gamma-\alpha}\right]\hat{t}\hat{z} + R, \end{aligned}$$

where $\hat{x} = (x - x')/x'$, for $x \in \{h, w, t, z\}$.

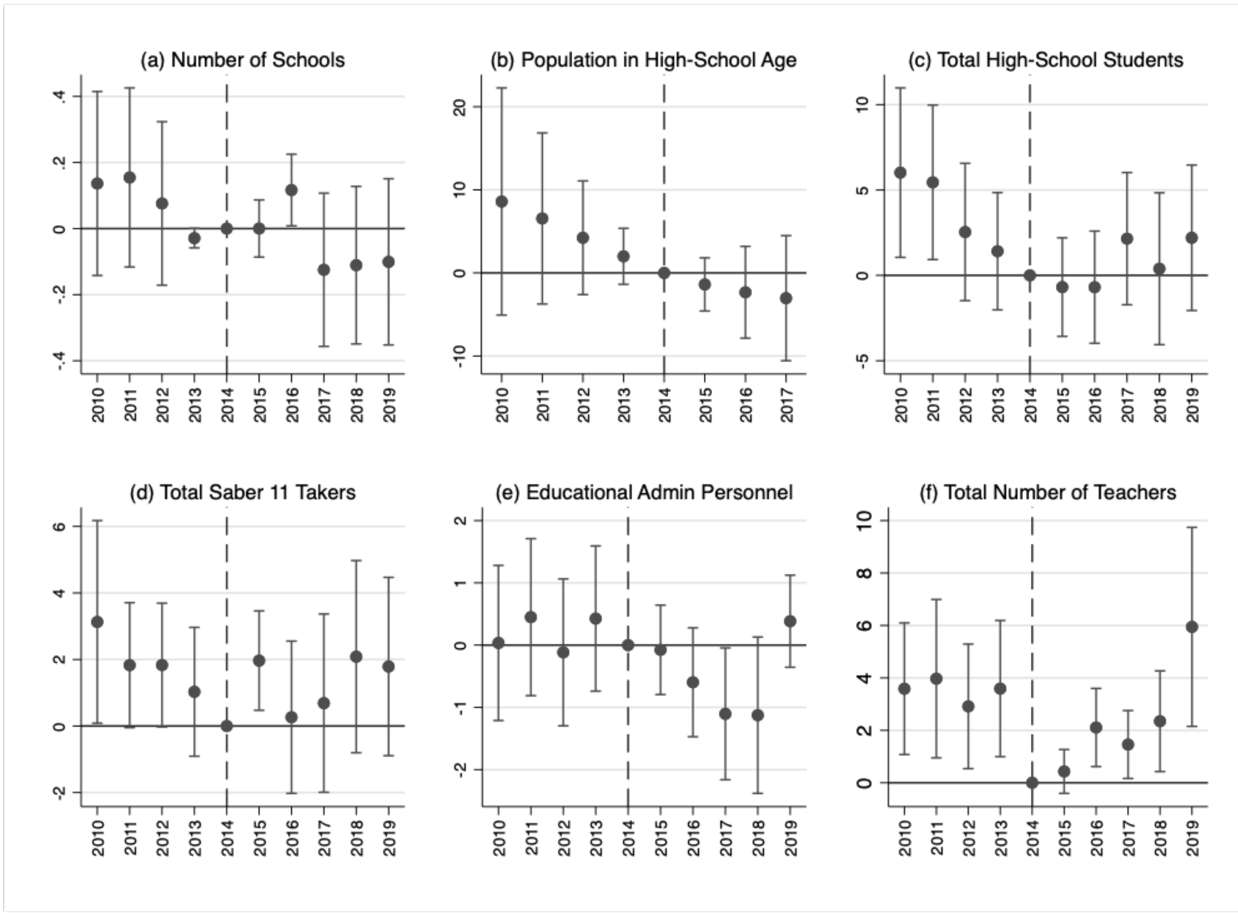
Appendix B Additional figures and tables

Figure B1: Effect of the past-year average maximum temperature on Saber 11 scores by Long-Run Low and High temperatures, 2014-II–2019-I.



Notes — Reporting Fixed-Effects estimation of the effect (and confidence interval) from 1°C increases in the average daily maximum temperature at the Municipality level during the calendar before the Saber 11 test on the scores achieved by students between 2014-II and 2019-I. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Regressions include controls of precipitation at the Municipality level for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Figure B2: Pre-trend testing of exposure to *Ser Pilo Paga* on Municipality-level variables, 2010-2019.



Notes — Municipality-level information is publicly available from the Center for Studies on Economic Development (CEDE) of Universidad de los Andes, Colombia. Graphs report the coefficient estimates (and 95% CI) from the regression of the outcome of interest (e.g., the number of schools in the Municipality) on interactions between the exposure to *Ser Pilo Paga* (SPP) and year dummies. Results are based on linear regression with State-Year and Municipality-level fixed effects. Excluded interaction for the year 2014 (coefficient set at zero) takes into account the baseline of Saber 11 data, for which no Municipality was exposed to SPP.

Table B1: Correlation of student-level attributes and past-year weather variables: average maximum temperature and average precipitation, 2014-II–2019-I.

VARIABLES	(1) Age	(2) Female	(3) Primary[(a)]	(4) Secondary[(a)]	(5) Tertiary[(a)]	(6) Minority[(b)]
Avg. maximum temperature	0.006	-0.000	0.001	-0.000	-0.001	0.001
	(0.007)	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)
Mean precipitation	0.002	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
OBSERVATIONS	2,202,869	2,202,869	2,202,869	2,202,869	2,202,869	2,202,869
R^2	0.119	0.077	0.090	0.064	0.263	0.557
SCHOOL FE	Yes	Yes	Yes	Yes	Yes	Yes
TIME-STATE FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day), also for the calendar year before the exam date. (a) Variables capturing the educational attainment of the test-taker's mother, measured as whether the educational level was completed. (b) Minority is the exam-taker self-identification as part of an ethnic minority.

Table B2: Effect of the exam-day maximum temperature on Saber 11 scores, 2014-II–2019-I.

Variables	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Maximum temperature (exam day)	-0.0146	-0.527**	0.130	-0.150	-0.336	-0.0559	0.0713	-0.166	0.00335
	(0.173)	(0.230)	(0.202)	(0.151)	(0.221)	(0.168)	(0.144)	(0.207)	(0.158)
Precipitation (exam day)	0.0134	0.0310	0.000949	-0.0231	-0.0117	-0.0417*	-0.000882	0.0175	-0.0191
	(0.0244)	(0.0342)	(0.0289)	(0.0218)	(0.0270)	(0.0247)	(0.0186)	(0.0268)	(0.0210)
Observations	2,185,457	1,502,504	682,952	2,185,457	1,502,504	682,952	2,185,457	1,502,504	682,952
R^2	0.395	0.386	0.340	0.337	0.314	0.284	0.446	0.433	0.374
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attribute Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Maximum temperature is the highest recorded temperature (in °C) during the exam day within each round. Precipitation is the total precipitation (mm) during the exam day within each round. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table B3: Effect of the past-year average maximum temperature on Saber 11 scores, 2014-II–2019-I.

VARIABLES	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. maximum temperature	-0.568	-2.669**	1.363*	-0.401	-2.111***	0.858	-0.207	-2.350***	1.425**
	(0.785)	(1.136)	(0.769)	(0.627)	(0.677)	(0.611)	(0.650)	(0.697)	(0.667)
Mean precipitation	-0.307	-0.893**	0.0178	-0.216	-0.704**	0.0405	-0.255	-0.654***	-0.0529
	(0.266)	(0.445)	(0.192)	(0.169)	(0.273)	(0.181)	(0.164)	(0.240)	(0.153)
OBSERVATIONS	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150
R^2	0.392	0.384	0.337	0.334	0.312	0.284	0.442	0.430	0.371
SCHOOL FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TIME-STATE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATTRIBUTE CONTROLS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SAMPLE	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Age-squared control is demeaned, so the coefficient estimate of age is the partial correlation at the sample mean. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table B4: Effect of the past-year average maximum and exam date maximum temperatures on Saber 11 scores, 2014-II–2019-I.

Variables	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. Maximum Temperature	-0.554	-2.436**	1.356*	-0.312	-1.977***	0.924	-0.260	-2.404***	1.463**
	(0.793)	(1.140)	(0.774)	(0.606)	(0.675)	(0.593)	(0.655)	(0.722)	(0.669)
Maximum temperature (exam days)	0.00510	-0.362	0.0695	-0.141	-0.203	-0.0968	0.0764	-0.000580	-0.0656
	(0.170)	(0.231)	(0.199)	(0.142)	(0.187)	(0.161)	(0.140)	(0.189)	(0.155)
Mean precipitation	-0.303	-0.878*	0.0181	-0.237	-0.721**	0.0124	-0.255	-0.639**	-0.0667
	(0.266)	(0.440)	(0.195)	(0.169)	(0.278)	(0.181)	(0.163)	(0.242)	(0.153)
Precipitation (exam day)	0.0101	0.0185	0.00330	-0.0255	-0.0221	-0.0401	-0.00344	0.00923	-0.0170
	(0.0244)	(0.0337)	(0.0291)	(0.0220)	(0.0281)	(0.0246)	(0.0188)	(0.0273)	(0.0212)
Observations	2,158,356	1,492,021	666,334	2,158,356	1,492,021	666,334	2,158,356	1,492,021	666,334
R^2	0.392	0.385	0.337	0.335	0.313	0.284	0.443	0.431	0.371
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attribute Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date within each round. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table B5: Effect of the past-year number of days exposure to maximum temperatures on Saber 11 scores, 2014-II–2019-I.

Variables	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
DD < 15	0.0518*	0.115***	-0.0206	0.0389*	0.0561**	0.0157	0.0383*	0.0755***	-0.000851
	(0.0297)	(0.0234)	(0.0388)	(0.0214)	(0.0227)	(0.0369)	(0.0200)	(0.0179)	(0.0342)
DD $15-20$	0.0161	0.0440*	-0.0134	0.00303	0.0197	-0.0103	0.0148	0.0415***	-0.0115
	(0.0158)	(0.0220)	(0.0244)	(0.0146)	(0.0217)	(0.0156)	(0.0126)	(0.0155)	(0.0173)
DD $25-30$	0.00652	-0.0237	0.0286*	0.00240	-0.0294**	0.0166	0.00852	-0.0183	0.0242*
	(0.0155)	(0.0168)	(0.0161)	(0.0112)	(0.0135)	(0.0121)	(0.0126)	(0.0144)	(0.0133)
DD $30-35$	0.00456	-0.0507*	0.0540**	-0.0100	-0.0629***	0.0411**	0.00922	-0.0447**	0.0557***
	(0.0214)	(0.0260)	(0.0244)	(0.0179)	(0.0170)	(0.0193)	(0.0185)	(0.0198)	(0.0208)
DD ≥ 35	-0.0191	-0.107**	0.0554*	-0.0141	-0.0881***	0.0369	-0.00387	-0.0919***	0.0562**
	(0.0281)	(0.0446)	(0.0313)	(0.0235)	(0.0301)	(0.0247)	(0.0240)	(0.0325)	(0.0272)
Mean precipitation	-0.301	-0.896*	0.0209	-0.198	-0.704**	0.0306	-0.251	-0.659***	-0.0674
	(0.264)	(0.455)	(0.182)	(0.167)	(0.277)	(0.182)	(0.159)	(0.241)	(0.148)
Observations	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150
R^2	0.392	0.384	0.337	0.334	0.312	0.284	0.442	0.430	0.371
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attribute Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. DD counts the number of days during the calendar year before the exam date for which the maximum temperature (in °C) was within the range. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table B6: Descriptive statistics of the data from ELCA

	Mean	Median	S.D.	Mean	Median	S.D.
YOUTHS	*Rural (N=3,241)*			*Urban (N=3,162)*		
Time-use variables (1 = One hour or more in a week)						
Watch TV	0.56	1.00	0.50	0.63	1.00	0.48
Use computer	0.13	0.00	0.33	0.36	0.00	0.48
Spend time with parents	0.62	1.00	0.49	0.63	1.00	0.48
Play with siblings or friends	0.59	1.00	0.49	0.58	1.00	0.49
Help with household chores	0.43	0.00	0.49	0.37	0.00	0.48
Do exercise	0.36	0.00	0.48	0.45	0.00	0.50
Read for fun	0.23	0.00	0.42	0.20	0.00	0.40
Do homework	0.63	1.00	0.48	0.67	1.00	0.47
Play videogames	0.06	0.00	0.24	0.17	0.00	0.38
Participate in social groups	0.07	0.00	0.25	0.12	0.00	0.33
Independent variables						
Age	11.99	12.00	1.65	12.06	12.00	1.67
Female (Yes=1)	0.49	0.00	0.50	0.50	0.00	0.50
ADULTS	*Household heads (N=12,155)*			*Spouses (N=9,158)*		
Time-use variables (% of reported time)						
Work on household farms	23.17	12.12	26.10	8.09	0.00	15.14
Work on other households' farms	22.02	0.00	27.37	4.46	0.00	13.82
Commuting or looking for jobs	2.66	0.00	5.56	0.70	0.00	3.04
Domestic chores	12.28	0.00	19.43	47.91	50.00	21.69
Leisure and personal care	39.65	36.36	17.39	38.61	36.42	17.45
Other activities	0.23	0.00	2.43	0.23	0.00	2.16
Independent variables						
Age	48.79	48.00	12.59	44.12	43.00	12.88
Female (1=Yes)	0.19	0.00	0.40	0.94	1.00	0.23

Notes — Individual-level information was accessed remotely using the computational infrastructure of the Data Center of the Universidad de los Andes. Age is calculated as the exact number of years from the date of birth to the survey date. Female is a binary variable based on self-identification from the respondents. The time-use variables for the sample of youths are dummies equal to one if the respondent spent one hour or more in a typical day in the reference period in each of the activities mentioned in the row. For the adults, they correspond to the percentage of time spent in the same reference period in each of the activities.

Table B7: Heterogeneous effect of the past-year average maximum temperature on Saber 11 scores by exposure to *Ser Pilo Paga*, 2014-II–2019-I.

VARIABLES	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. maximum temperature (*AMT*)	-0.445	-2.158**	1.436*	-0.348	-2.044***	0.907	-0.145	-2.169***	1.473**
	(0.726)	(0.937)	(0.781)	(0.622)	(0.721)	(0.624)	(0.622)	(0.576)	(0.684)
Exposure to SPP × Post	3.359***	6.013***	2.414**	0.104	0.629	0.124	1.697**	3.068**	1.029
	(1.100)	(1.820)	(1.134)	(0.654)	(1.159)	(0.716)	(0.790)	(1.271)	(0.869)
AMT × Exposure to SPP × Post	-0.00365	-0.0923*	-0.0161	-0.0189	0.0148	-0.0301	-0.00606	-0.0271	-0.0171
	(0.0416)	(0.0511)	(0.0417)	(0.0240)	(0.0388)	(0.0244)	(0.0282)	(0.0388)	(0.0294)
Mean precipitation	-0.0460	-0.326	0.0575	-0.247	-0.658**	0.0714	-0.115	-0.301	-0.0288
	(0.251)	(0.473)	(0.204)	(0.175)	(0.313)	(0.172)	(0.157)	(0.253)	(0.153)
Mean precipitation × Exposure to SPP × Post	-0.138*	-0.250	-0.0334	0.0195	0.0337	-0.0309	-0.0813	-0.153	-0.0214
	(0.0751)	(0.162)	(0.0739)	(0.0435)	(0.102)	(0.0503)	(0.0531)	(0.119)	(0.0557)
Observations	2,202,869	1,516,989	685,879	2,202,869	1,516,989	685,879	2,202,869	1,516,989	685,879
R^2	0.394	0.385	0.340	0.336	0.313	0.284	0.445	0.432	0.374
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attribute Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneous Trends	No	No	No	No	No	No	No	No	No
Sample	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average of the daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table B8: Heterogeneous effect of the past-year average maximum temperature on Saber 11 scores by exposure to *Ser Pilo Paga* under specific mother's educational attainment levels, 2014-II–2019-I.

Panel A: Overall

	Math		Spanish		Total Score	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature	0.569	-1.420	0.111	-1.083	0.719	-1.137
	(0.723)	(0.871)	(0.630)	(0.737)	(0.597)	(0.723)
Exposure to SPP × Post	3.285***	2.669**	0.281	-0.220	1.820**	0.866
	(0.918)	(1.166)	(0.565)	(1.050)	(0.712)	(0.910)
AMT × Exposure to SPP × Post	-0.0382	-0.00576	-0.0141	0.0105	-0.0327	0.0116
	(0.0327)	(0.0403)	(0.0192)	(0.0377)	(0.0237)	(0.0321)
Mother's Education	Low	High	Low	High	Low	High
Observations	965,202	1,209,602	965,202	1,209,602	965,202	1,209,602
R^2	0.281	0.388	0.236	0.313	0.309	0.432

Panel B: Urban

	Math		Spanish		Total Score	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature	-1.319*	-2.557*	-1.014	-2.397**	-0.712	-2.808***
	(0.719)	(1.348)	(1.046)	(0.944)	(0.727)	(0.877)
Exposure to SPP × Post	8.024***	4.303**	2.728*	-1.535	5.730***	0.975
	(2.070)	(2.008)	(1.404)	(1.639)	(1.505)	(1.689)
AMT × Exposure to SPP × Post	-0.154***	-0.0576	-0.0387	0.0770	-0.105**	0.0269
	(0.0530)	(0.0593)	(0.0425)	(0.0540)	(0.0422)	(0.0511)
Mother's Education	Low	High	Low	High	Low	High
Observations	550,045	955,949	550,045	955,949	550,045	955,949
R^2	0.256	0.386	0.205	0.301	0.278	0.427

Panel C: Rural

	Math		Spanish		Total Score	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature	2.070**	0.532	1.190	0.619	1.842***	0.982
	(0.797)	(1.086)	(0.725)	(0.949)	(0.692)	(0.931)
Exposure to SPP × Post	2.268**	1.810	-0.216	0.102	0.897	0.432
	(1.091)	(1.638)	(0.661)	(1.182)	(0.750)	(1.224)
AMT × Exposure to SPP × Post	-0.0303	0.00389	-0.0182	-0.0266	-0.0251	0.00373
	(0.0423)	(0.0554)	(0.0243)	(0.0407)	(0.0272)	(0.0423)
Mother's Education	Low	High	Low	High	Low	High
Observations	415,155	253,649	415,155	253,649	415,155	253,649
R^2	0.290	0.343	0.234	0.281	0.314	0.373
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for dwelling stratum classification, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities. (a) Trend heterogeneity is allowed by adding interactions between period (year-semester) dummies and quartiles of Municipality-level variables (namely Total High-School Students, Total Number of School Teachers, and Total Number of School Administrative Personnel).

Table B9: Heterogeneous effect of the past-year average maximum temperature on Saber 11 scores by exposure to *Ser Pilo Paga* under specific House strata levels, 2014-II–2019-I.

	Math		Spanish		Total Score	
			Panel A: Overall			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature (AMT)	-0.371	-0.781	-0.480	-2.415	-0.139	-2.610
	(0.705)	(2.727)	(0.562)	(1.961)	(0.593)	(2.044)
Exposure to SPP × Post	2.994***	0.860	0.160	-0.251	1.506**	-1.921
	(0.987)	(3.259)	(0.600)	(3.055)	(0.736)	(2.743)
AMT × Exposure to SPP × Post	-0.00986	-0.00580	-0.00593	0.0423	-0.00990	0.116
	(0.0358)	(0.106)	(0.0217)	(0.118)	(0.0260)	(0.0998)
HOUSE STRATA	Low	High	Low	High	Low	High
OBSERVATIONS	2,011,055	162,315	2,011,055	162,315	2,011,055	162,315
R^2	0.337	0.537	0.291	0.438	0.379	0.592
			Panel B: Urban			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature (AMT)	-2.157***	0.230	-1.925***	-1.545	-1.964***	-2.198
	(0.789)	(3.630)	(0.702)	(2.122)	(0.524)	(2.526)
Exposure to SPP × Post	6.584***	0.431	0.850	-2.798	3.722***	-3.052
	(1.835)	(4.352)	(1.029)	(3.027)	(1.286)	(2.922)
AMT × Exposure to SPP × Post	-0.106**	-0.0446	0.0133	0.137	-0.0463	0.149
	(0.0480)	(0.136)	(0.0346)	(0.123)	(0.0373)	(0.108)
HOUSE STRATA	Low	High	Low	High	Low	High
OBSERVATIONS	1,354,424	150,960	1,354,424	150,960	1,354,424	150,960
R^2	0.320	0.485	0.263	0.370	0.357	0.531
			Panel C: Rural			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Avg. Maximum Temperature (AMT)	1.510*	-7.438	0.980	-12.74**	1.527**	-7.159*
	(0.780)	(4.881)	(0.625)	(5.592)	(0.691)	(4.262)
Exposure to SPP × Post	2.136*	21.54*	-0.134	18.00*	0.786	13.70
	(1.094)	(12.17)	(0.688)	(9.826)	(0.821)	(9.869)
AMT × Exposure to SPP × Post	-0.0103	-0.759*	-0.0203	-0.563*	-0.0121	-0.418
	(0.0399)	(0.413)	(0.0236)	(0.294)	(0.0282)	(0.344)
HOUSE STRATA	Low	High	Low	High	Low	High
OBSERVATIONS	656,709	11,335	656,709	11,335	656,709	11,335
R^2	0.325	0.733	0.275	0.672	0.356	0.795
SCHOOL FE	Yes	Yes	Yes	Yes	Yes	Yes
TIME-STATE FE	Yes	Yes	Yes	Yes	Yes	Yes
ADDITIONAL CONTROLS	Yes	Yes	Yes	Yes	Yes	Yes

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities. (a) Trend heterogeneity is allowed by adding interactions between period (year-semester) dummies and quartiles of Municipality-level variables (namely Total High-School Students, Total Number of School Teachers, and Total Number of School Administrative Personnel).

Table B10: Heterogeneous effect of past-year average maximum temperature on Saber 11 scores by education of the mother (reporting education-level dummies), 2014-II–2019-I.

Variables	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. maximum temperature (AMT)	-0.302	-2.335**	1.575**	-0.143	-1.769**	1.045*	0.0779	-1.971***	1.630**
	(0.776)	(1.118)	(0.769)	(0.616)	(0.680)	(0.613)	(0.640)	(0.690)	(0.667)
AMT × Mother: Complete primary	-0.154***	-0.166***	-0.100*	-0.143***	-0.183***	-0.0701	-0.166***	-0.199***	-0.0971**
	(0.0542)	(0.0595)	(0.0563)	(0.0455)	(0.0477)	(0.0520)	(0.0518)	(0.0570)	(0.0478)
AMT × Mother: Complete secondary	-0.392***	-0.349***	-0.498***	-0.383***	-0.361***	-0.460***	-0.412***	-0.395***	-0.471***
	(0.0604)	(0.0595)	(0.0978)	(0.0529)	(0.0551)	(0.0885)	(0.0540)	(0.0611)	(0.0878)
AMT × Mother: Complete tertiary	-0.526***	-0.494***	-0.718***	-0.512***	-0.499***	-0.671***	-0.575***	-0.559***	-0.730***
	(0.143)	(0.130)	(0.139)	(0.102)	(0.0737)	(0.124)	(0.117)	(0.109)	(0.125)
Mother: Complete primary	9.791***	10.66***	7.373***	8.501***	9.754***	5.936***	8.785***	10.09***	6.256***
	(1.654)	(1.715)	(1.659)	(1.398)	(1.427)	(1.535)	(1.624)	(1.747)	(1.418)
Mother: Complete secondary	24.32***	22.66***	28.40***	23.81***	22.61***	26.87***	23.88***	23.03***	26.35***
	(1.652)	(1.607)	(2.707)	(1.460)	(1.481)	(2.543)	(1.476)	(1.665)	(2.480)
Mother: Complete tertiary	45.13***	42.95***	54.04***	43.88***	42.17***	51.72***	46.21***	44.69***	53.51***
	(4.241)	(3.499)	(3.732)	(3.017)	(1.861)	(3.342)	(3.450)	(2.903)	(3.367)
Mean precipitation	-0.306	-0.894**	0.0181	-0.215	-0.705**	0.0405	-0.254	-0.656***	-0.053
	(0.266)	(0.445)	(0.192)	(0.169)	(0.273)	(0.181)	(0.163)	(0.240)	(0.153)
Observations	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150
R^2	0.392	0.384	0.337	0.334	0.312	0.284	0.442	0.430	0.371
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attribute Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average of the daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.

Table B11: Heterogeneous effect of past-year average maximum temperature on Saber 11 scores by house strata (reporting strata-level dummies), 2014-II–2019-I.

VARIABLES	(1) Math	(2) Math	(3) Math	(4) Spanish	(5) Spanish	(6) Spanish	(7) Total	(8) Total	(9) Total
Avg. maximum temperature (AMT)	-0.174	-2.079*	1.575**	-0.0681	-1.633**	1.040*	0.157	-1.793**	1.608**
	(0.760)	(1.120)	(0.763)	(0.608)	(0.670)	(0.594)	(0.626)	(0.686)	(0.653)
AMT × House Strata: Level 2	-0.564***	-0.555***	-0.533***	-0.461***	-0.441***	-0.439***	-0.505***	-0.513***	-0.448***
	(0.0572)	(0.0649)	(0.0827)	(0.0571)	(0.0651)	(0.0761)	(0.0599)	(0.0694)	(0.0735)
AMT × House Strata: Level 3	-0.878***	-0.786***	-0.986***	-0.794***	-0.670***	-1.046***	-0.844***	-0.755***	-0.995***
	(0.119)	(0.145)	(0.182)	(0.110)	(0.136)	(0.171)	(0.117)	(0.145)	(0.169)
AMT × House Strata: Level 4	-1.392***	-1.165***	-1.559***	-1.165***	-0.913***	-1.454***	-1.339***	-1.125***	-1.442***
	(0.200)	(0.203)	(0.347)	(0.163)	(0.157)	(0.399)	(0.180)	(0.178)	(0.352)
AMT × House Strata: Level 5	-1.374***	-1.186***	-0.619**	-1.211***	-0.942***	-1.036***	-1.347***	-1.135***	-0.796***
	(0.249)	(0.275)	(0.312)	(0.184)	(0.181)	(0.295)	(0.224)	(0.240)	(0.292)
AMT × House Strata: Level 6	-1.619***	-1.377***	-0.990**	-1.338***	-1.110***	-0.365	-1.550***	-1.325***	-0.627*
	(0.279)	(0.263)	(0.477)	(0.209)	(0.208)	(0.370)	(0.244)	(0.231)	(0.327)
House Strata: Level 2	15.15***	15.79***	13.83***	14.40***	14.88***	12.95***	14.49***	15.86***	12.10***
	(1.591)	(1.793)	(2.241)	(1.648)	(1.832)	(2.173)	(1.705)	(1.894)	(2.037)
House Strata: Level 3	20.01***	19.86***	17.22***	22.68***	21.48***	24.75***	21.85***	21.90***	20.57***
	(2.951)	(3.555)	(4.464)	(2.781)	(3.334)	(4.404)	(2.900)	(3.514)	(4.234)
House Strata: Level 4	28.24***	26.60***	12.29	29.09***	26.52***	18.00	30.81***	29.65***	13.17
	(4.830)	(4.886)	(10.84)	(3.983)	(3.880)	(12.49)	(4.420)	(4.353)	(11.18)
House Strata: Level 5	23.65***	23.73***	-17.99**	27.02***	24.78***	1.164	27.31***	27.08***	-10.36
	(6.287)	(6.398)	(8.731)	(4.680)	(4.686)	(8.318)	(5.680)	(5.808)	(8.229)
House Strata: Level 6	26.52***	25.56***	-8.817	26.52***	25.68***	-17.97	28.92***	28.88***	-16.82*
	(7.802)	(7.108)	(13.86)	(5.592)	(5.679)	(10.98)	(6.759)	(6.246)	(9.802)
Mean precipitation	-0.321	-0.910**	0.00340	-0.227	-0.717**	0.0271	-0.268*	-0.671***	-0.0660
	(0.262)	(0.447)	(0.190)	(0.165)	(0.270)	(0.180)	(0.159)	(0.240)	(0.152)
OBSERVATIONS	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150	2,175,363	1,506,211	669,150
R^2	0.392	0.384	0.337	0.334	0.312	0.284	0.442	0.430	0.372
SCHOOL FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TIME-STATE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATTRIBUTE CONTROLS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SAMPLE	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

Notes — Cluster (weather station level) robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The response variable is $100y$, where y is the standardized score of the specific subject (math, Spanish) or total test score. Avg. maximum temperature captures the average daily maximum temperature (in °C) during the calendar year before the exam within each round. Mean precipitation is the average daily precipitation (mm/day) for the calendar year before the exam date. Attribute controls added to the regression include age (and its square), dummy variables for the maximum educational attainment of the mother, whether the test taker is female, self-identification as part of an ethnic minority, and whether the exam taker lives and studies in different Municipalities.