Temperature, Effort, and Achievement

Ana Paula Melo
Howard University

Mizuhiro Suzuki*
Univ. of Wisconsin-Madison

March 24, 2023

Abstract

Our paper estimates the effects of temperature on achievement and provides the first empirical evidence on how exam stakes affect the sensitivity of exam performance to temperature. Using data on millions of exam takers in Brazil, we explore a unique context where the stakes of a large-scale standardized exam change from relatively low to high. We find that the higher the stakes, the smaller the effects of temperature on exam performance. Our results suggest that effort is an important channel: in a high-stakes environment, exam takers exert more effort, counterbalancing an otherwise adverse and economically relevant temperature effect.

JEL Codes: I21, Q54, O18
Keywords: achievement, temperature, effort, exam stakes

*Contact information: anapaula.melodasilv@howard.edu and msuzuki7@wisc.edu. We would like to thank Laura Schechter for her extensive feedback and support, Lydia Ashton, Bradford Barham, Joshua Deutschmann, Paul Dower, Jeremy Foltz, Priya Mukherjee, and Jeffrey Smith for helpful comments and suggestions. Any errors in this draft are our sole responsibility.

1 Introduction

Standardized exams are a popular measure of education quality and a widely used resource allocation criterion, including for government financial transfers to schools, teacher compensation, college seats, and financial aid. Worldwide, standardized test scores are commonly used for comparing test-takers across time and context. Standardized tests can also be a cheap and effective signal of ability, especially for high-achieving, low-income students (Hyman, 2017). However, standardized tests have received increasing push-back due to scores being sensitive to various demographics, psychological, and context-specific characteristics such as gender, socioemotional skills, pollution, and temperature (Borghans et al., 2016; Ebenstein et al., 2016; Graff Zivin et al., 2018; Reardon et al., 2018).

The perceived importance of an exam for individuals' outcomes, i.e., the exam stakes, directly determines students' incentives to exert effort. Although there is increasing academic and public attention on the effects of temperature on exam performance and human capital accumulation, the role of exam stakes (high vs. low) as a potential mechanism remains an open question. Differential incentives to perform induced by varying degrees of stakes may affect test takers' ability to respond to unexpected external shocks, such as temperature.

Our paper provides the first evidence of how individual effort mediates the effects of temperature on achievement. We build on standard models of how heat increases the disutility of effort (Park, 2020) and provide testable hypotheses on the interaction of exam stakes and temperature, highlighting the potential role of effort in mitigating the adverse effects of temperature on performance. Our model predicts that, in a low-stakes environment, a significant portion of the reduction in achievement can be attributed to reduced effort. In a high-stakes environment, test takers have incentives to maximize effort, and the remaining effects of temperature on scores are more likely due to the direct impact on cognitive performance.

Using individual-level data on millions of exam takers in a national high school exam administered yearly in Brazil, we estimate the effects of transitory temperature shocks on exam scores and how they interact with the exam stakes. Our identification strategy leverages the fact that individuals take different subject exams on two consecutive days to identify the effects of temperature on performance. This allows us to control for time-invariant individual characteristics such as exam preparation and general ability. We identify the causal impact of transitory temperature shocks on exam scores. We then economically quantify our results by showing how these temperature shocks translate into changes in applicants' potential choices of college majors.

Our empirical strategy can distinguish between effort and cognitive effects by exploring a unique context in which the stakes of a standardized exam change gradually. We use temporal and geographical variation in the number of universities adopting a centralized admissions system, which induces exogenous variation in exam stakes. Since 2010, the national high school exam has gained importance as institutions gradually moved to this centralized admissions system. Before this system was available, the exam was also used for college admissions, providing bonus points, but less commonly used as a necessary criterion for admissions. Universities across the country joined the centralized system at different times. Once a university joins, this national exam becomes a necessary (or exclusive) admission criterion.

Our findings suggest that the effects of temperature are economically sizable. Our baseline results show a negative average impact of high temperature on exam scores – a one standard deviation increase in temperature decreases exam scores by 0.036 s.d. These adverse effects are non-linear, with estimates varying up to 0.12 s.d. More importantly, these effects are economically relevant. Using data on college-major cut-offs, we calculate the number of majors for which the cut-offs fall within the interval between the estimated effects and the average cut-off for all majors in a college. Taking our linear estimated effects of temperature on scores, a one standard deviation in temperature can affect the number of applicant's college-major options by 8.2 percent on average. These changes in the number of applicants' attainable majors can vary from minus 10 to 30 percent if we take into account the estimated effects from the non-linear specification. This evidence supports the increasing scientific and public claims that climate comfort in educational settings is an important policy priority, especially in selective contexts.

We also show that effort is an important channel through which temperature affects exam performance. We interact temperature with the proportion of universities in a locality using the centralized system for admissions, which increases the exam stakes by making it a mandatory criterion. We find that the higher the stakes, the smaller the effects of temperature on exam performance. A one standard deviation increase in exam stakes leads to a 58 percent reduction in the average effect of temperature on exam scores. When the stakes are the highest, the temperature effect decreases by over 80 percent. Other heterogeneity analysis also provides suggestive - but inconclusive - evidence that males are less affected by temperature shocks than females, especially when the exam stakes are sufficiently high. Taken together, our results support the hypothesis that, in a high-stakes environment, exam takers exert more effort, counterbalancing an otherwise substantial effect of temperature if the stakes were lower.

As robustness exercises, we investigate the possibility that our results are affected by potential endogenous adoption of the centralized system; selection bias (e.g., the composition of test-takers changing in response to the centralized system); and omitted time-varying factors. Our results are robust to strategies to deal with these potential threats, with minor changes in estimated coefficients of interest in response to controlling for the relevant interaction terms. We also provide evidence that our results are robust to different measures of exam stakes, including yearly adoption of the centralized system at the intensive margin and an alternative large-scale affirmative action policy that also affected exam stakes.

Our paper contributes to our understanding of how effort can mitigate the harmful effects of external factors on performance. Using unique variation, we are the first to show that individual perception of the exam's importance affects how students react to external shocks, such as temperature. Previous work has focused on the short-term, contemporaneous effects of temperature on performance when stakes remain fixed.[1] Park (2020) finds a negative effect of temperature on test performance in the US, with a persistent longer-term impact on educational attainment. Using high-stakes college entrance exams, Graff Zivin et al. (2020) exploit temperature shocks during the exam day and find negative effects on college entrance exam scores and the probability of joining first-tier colleges in China. Our results on the direct effects of temperature during the exam on tests scores corroborate their findings while also avoiding common issues such as grade manipulation, as discussed by Park (2020), or analysis restricted to top achieving students accepted at universities (Graff Zivin et al., 2020).

In the Brazilian context, our paper directly relates to Li and Patel (2021), who also estimate temperature effects on exam performance in the same context and use the same data as ours, but the results differ substantially. While our paper finds a negative, statistically, and economically significant effect, their paper finds negligible and insignificant results. We compare research design choices and discuss why our study likely provides more precise estimates. Mainly, our design uses high-frequency temperature data and precisely isolates exposure during the exam, focuses only on high-school seniors taking the test for the first time (a more homogeneous group), and restricts analysis to multiple-choice questions for cross-subject and temporal comparability.

Finally, our paper particularly informs policy using standardized test scores to allocate

[1] Another branch of this literature estimates the effects of *prolonged* exposure to heat. In the US context, Park et al. (2020) find a negative impact of a hot year on learning. Based on the total number of hot days in the year before the exam, Garg et al. (2020) finds a negative impact of temperature on performance mediated by an agricultural mechanism.

resources, especially when performance is more likely affected by external factors. For places with higher temperature variability, allocating resources based on individual performance on standardized exams can exacerbate inequality. Our findings suggest a role for investment in infrastructure, such as air conditioning, to mitigate a potentially important source of inequality affecting exam performance and college access.

This paper proceeds as follows. In section 2, we describe the institutional background in Brazil. Section 3 formulates the conceptual model, and in section 4 we describe the data. We discuss our identification strategy and results in Sections 5 and 6. Section 7 provides robustness checks of our results. Section 8 concludes.

2 Context

Admissions to public universities in Brazil rely exclusively on entrance exam scores. Until 2009, universities had their specific admissions process and entrance exams (the *Vestibular*), and students applied directly to the institutions of interest. Institutions often provided bonus points based on performance on the *Exame Nacional do Ensino Mèdio* (ENEM, National High School Evaluation Exam), marginally increasing one's chance of acceptance.

ENEM is a non-mandatory national exam initially created as a high school evaluation and mostly taken by people interested in college.[2] The exam is administered once a year in about 1,800 municipalities across all states. Registration costs 68 *reais* ($\approx$ 18 USD), and a fee waiver is available for low-income applicants. Anyone can take the exam, from high school seniors to adults of any age pursuing tertiary education or interested in obtaining a certificate equivalent to a high-school diploma. This test is a self-assessment tool; students' scores reveal their chances of getting into college and specific majors. Applicants can use

[2] In the socio-economic survey administered to all exam takers, 88 percent ranked "college application" as the most important reason for taking the ENEM on a scale of 1 to 5. About 80 percent also listed "obtaining financial aid for college" as a relevant factor.

their scores to apply to public universities and to qualify for federal financial aid to access private institutions (scholarships or student credit). From its creation in 1998 until the 2009 reformulation, ENEM was considered less relevant to public university admissions than the universities' exams, the *Vestibulares*.

In 2008, the federal government conducted a comprehensive college admissions reform by reformulating the ENEM and creating a centralized university admission system (SISU, *Sistema de Seleção Unificada*). ENEM was reformulated to be more rigorous, and its content aimed to reflect the national mandatory high school curriculum. It became a two-day exam consisting of four modules, totaling 180 items, plus one essay. Final scores are calculated based on Item Response Theory, which allows score comparisons over time. October of 2009 was the first time the new ENEM was administered. Students could use 2009 ENEM scores to apply for the first SISU edition in January 2010. The exam repeats once every year.

As of January 2010, colleges participating in SISU could offer seats to students taking the ENEM score as the only criteria, assigning their preferred weights to each exam module and essay. All state and federal institutions were allowed to join the system. Although adoption was not mandatory, universities were incentivized to lower their costs by transferring their admissions process to the federal government. Voluntary college adhesion to this centralized system increased over time. Universities could offer all or partial seats through SISU. In specific cases, universities adopted SISU as one admissions criterion, with additional college-specific exams. Participation increased from 25 out of 170 federal and state universities in 2010 to 92 out of 192 in 2017. As a result, the introduction of SISU was a significant push to establish ENEM as a high-stakes exam, becoming an important criterion for granting or denying admission to college among participating institutions.

Applicants from all over the country can apply to a university through SISU. However, individuals in Brazil have high mobility costs for college purposes. When comparing the location of residence during the ENEM exam and college of attendance, only 10 percent of college students nationwide attend out-of-state colleges, and about half are from the same municipality the university is located (Machado and Szerman, 2021). The inter-state migration averages are stable during the period of our study. However, causal estimates from Machado and Szerman (2021) show that SISU significantly affected interstate migration and the quality of enrolled students as measured by ENEM.

Even though SISU induce people to apply to universities outside their residence locality, mobility costs are still high, and individuals living closer to campus are likely more affected by the policy than individuals living further away (Card, 1995). In our empirical strategy, we address this differential treatment effect by weighting the SISU treatment by the distance between the municipality of residence and all federal and public universities in the country.

3 Theoretical Framework

We develop a model that captures the role of exam stakes, effort, and temperature on exam scores. We build on the model from Park (2020). We modify it to incorporate exam stakes explicitly. We derive testable hypotheses of how the stakes alter the temperature effect on exam performance and the potential mediating role of effort.

Suppose that exam takers gain utility $U(w, e, a)$, where w is future wages, e is the effort made during the exam, and a is the temperature during the exam. We assume that the disutility from the effort and temperature during the exam and the utility from future wages are separable: $U(w, e, a) = u_1(e, a) + u_2(w)$. This assumption is plausible since future wages are not realized on the exam dates but later in their lives. It implies that the effort level or temperature during the exam do not affect how an increase in future wages improves utility, reflected in $\frac{\partial^2 U}{\partial w \partial e} = 0$ and $\frac{\partial^2 U}{\partial w \partial a} = 0$. We further assume that (i) higher future wages increase utility $\left(\frac{\partial u_2}{\partial w} > 0\right)$, (ii) exerting effort is costly $\left(\frac{\partial u_1}{\partial e} < 0\right)$, (iii) a higher temperature gives discomfort and decreases utility $\left(\frac{\partial u_1}{\partial a} < 0\right)$, (iv) marginal returns to future wages diminish

$\left(\frac{\partial^2 u_2}{\partial w^2} < 0\right)$, and (v) the cost of effort to utility is convex $\left(\frac{\partial^2 u_1}{\partial e^2} < 0\right)$. We also assume a higher effort cost under a hotter environment, $\frac{\partial^2 u_1}{\partial e \partial a} < 0$, which is consistent with findings in previous studies (reviewed in Lim et al., 2008).

Future wages are determined by exam score y and exam stakes s as $w = w(y, s)$. We assume a positive relationship between exam score and future wages, $\frac{\partial w}{\partial y} > 0$. The exam score is a function of effort and temperature during the exam: $y = y(e, a)$. For its derivatives, we assume that (i) effort increases scores $\left(\frac{\partial y}{\partial e} > 0\right)$, (ii) the effort effect diminishes $\left(\frac{\partial^2 y}{\partial e^2} < 0\right)$, (iii) effort is less effective in improving the test score when temperature is higher due to cognitive impairment $\left(\frac{\partial^2 y}{\partial e \partial a} < 0\right)$, and (iv) a higher temperature has an adverse impact on cognitive performance and hence on exam scores $\left(\frac{\partial y}{\partial a} < 0\right)$.[3]

Given the above notation, we express the utility maximization problem as

$$\max_{e} \quad u_1(e, a) + u_2(w(y(e, a), s)).$$

The first order condition is

$$\frac{\partial u_2}{\partial w} \frac{\partial w}{\partial y} \frac{\partial y}{\partial e^*} + \frac{\partial u_1}{\partial e^*} = 0,$$

which captures the trade-off between the benefit and cost of increasing effort. While making more effort increases exam scores and future wages, it exhausts the exam taker, decreasing utility. At the optimal effort level, these two counteracting effects are balanced. To guarantee the existence and the uniqueness of the solution in this maximization problem, we assume that the objective function is globally concave in the effort level: $\frac{\partial u_2}{\partial w} \frac{\partial^2 w}{\partial y^2} \left(\frac{\partial y}{\partial e}\right)^2 + \frac{\partial u_2}{\partial w} \frac{\partial w}{\partial y} \frac{\partial^2 y}{\partial e^2} + \frac{\partial^2 u_1}{\partial e^2} < 0$.[4]

We can derive the effect of temperature on exam scores from $y = y(e, a)$, evaluated at

[3]Notice that this assumption is about the effect of temperature without the effort adjustment. Below, we show that the heat worsens the exam scores even after adjusting efforts.

[4]This assumption holds if $\frac{\partial^2 w}{\partial y^2}$ is (i) negative or (ii) positive but sufficiently small. In other words, the condition holds if a high exam score does not result in excessively high future income.

the optimal effort e^*:

$$\frac{dy}{da} = \frac{\partial y}{\partial e^*}\frac{\partial e^*}{\partial a} + \frac{\partial y}{\partial a}.$$

This equation shows two paths from temperature to exam scores: The first path is through a change in effort, and the second path is the direct effect on performance. Note that the sign of $\frac{\partial e^*}{\partial a}$ is undetermined. Under a higher temperature, while effort costs may decrease effort level (due to $\frac{\partial^2 u_1}{\partial e \partial a} < 0$), exam takers might increase their effort level to compensate for the negative heat effect on performance. We provide proof that the total effect of temperature on exam scores is negative ($\frac{dy}{da} < 0$).[5] This result provides us with the following testable hypothesis:

Hypothesis 1 *An increase in temperature negatively impacts exam scores.*

We now derive the effect of a change in stakes on the response of exam scores to temperature, evaluated at e^*:

$$\begin{aligned}\frac{d}{ds}\frac{dy}{da} &= \frac{\partial^2 y}{\partial e^{*2}}\frac{\partial e^*}{\partial s}\frac{\partial e^*}{\partial a} + \frac{\partial y}{\partial e^*}\frac{\partial^2 e^*}{\partial s \partial a} + \frac{\partial y^2}{\partial a \partial e^*}\frac{\partial e^*}{\partial s} \\ &= \left(\frac{\partial^2 y}{\partial e^{*2}}\frac{\partial e^*}{\partial a} + \frac{\partial y^2}{\partial a \partial e^*}\right)\frac{\partial e^*}{\partial s} + \frac{\partial y}{\partial e^*}\frac{\partial^2 e^*}{\partial s \partial a}.\end{aligned}$$

The increase in s affects $\frac{dy}{da}$ through two channels. The first channel is the change in the level of e^*: for example, students may exert different levels of effort at a mock exam and a college entrance exam given that the latter is strongly related to the future income. The size of this effect depends on temperature through the cognitive effect and effort costs. The second channel is the change in the temperature effect on the effort level. This reflects the compensatory effort by students to counterbalance the effect of temperature. Exam takers may make more effort to mitigate the negative impact of heat when the exam is more important.

[5]The proof is provided in Appendix A.

The sign of $\frac{d}{ds}\frac{dy}{da}$ is undetermined. Note that, if stakes are sufficiently high, exam takers might sufficiently compensate for the temperature effect. That is, $\frac{d}{ds}\frac{dy}{da} > 0$ if $\frac{\partial^2 e^*}{\partial s \partial a}$ is sufficiently large and positive. We derive the following hypothesis:

Hypothesis 2 *An increase in exam stakes mitigates the negative effect of temperature on exam scores.*

We empirically test these two hypotheses and provide estimates of the temperature effects on exam scores and the mitigating effects of increasing exam stakes.

4 Data description

In this section, we provide information on datasets and descriptive statistics. We use three datasets: (i) individual-level data on the national exam (ENEM); (ii) university-campus-level information on the adoption of the centralized system, SISU; (iii) municipal-level data on weather.

4.1 Exam scores and exam stakes data

ENEM (*Exame Nacional do Ensino Médio* - National High School Exam) is the primary outcome data, covering the universe of exam takers in Brazil from 2010 to 2016.

The Ministry of Education maintains a publicly available database.[6] It contains information on exam takers collected at registration and their subsequent exam scores. The data includes information on IRT-based final scores in the four subjects - natural sciences, social sciences, Portuguese (language), and mathematics. It also contains demographic and socioeconomic information on exam takers. The data provides information on the municipality where each exam taker took ENEM. We use this geographic information to link the exam outcome data to the weather data described in the following subsection.

[6]INEP - *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*

We restrict the population of exam takers to students in their last year of high school, who represent only about 20 percent of all exam takers. We keep applicants that were present and not eliminated from the exam.[7] We also restrict the population to 16 to 20 years old applicants. The resulting data (Table F.1) covers about 8 million high-school seniors distributed taking the national exam from 2010 to 2016 in about 1,800 municipalities in Brazil (out of $\approx$ 5,600 municipalities total).[8] Figure E.1 shows the distribution of exam locations across the country, which are more concentrated in populated areas.

Exams are administered on two consecutive days, Saturday and Sunday. Table 1 summarizes the types of exams by day and the amount of time exam takers have available. Each multiple-choice exam is paper-based and has 45 items.

Table 1: Details on the structure of the exam

	Exams	Exam start	Max. duration
Day 1	Social Sciences, Natural Sciences	1pm*	4h30min
Day 2	Portuguese, Mathematics and Essay	1pm*	5h30min

Note: (*) The start time refers to the Brasilia timezone. During the time of the exam, Brazil is under four different time zones. Students start the exam at 10 am, 11 am, or 12 pm local time, depending on the area. We adjust the temperature at the time of the exam for each municipality to reflect the hours the students are taking the exam. All exam takers need to be at the exam location at least one hour before the exam starts, strictly enforced. We exclude from the sample exam takers who cannot start the exam until the evening for religious reasons. These individuals arrive at the exam location at the same time as everyone else, and they wait in a room with no external communication until they can start the exam.

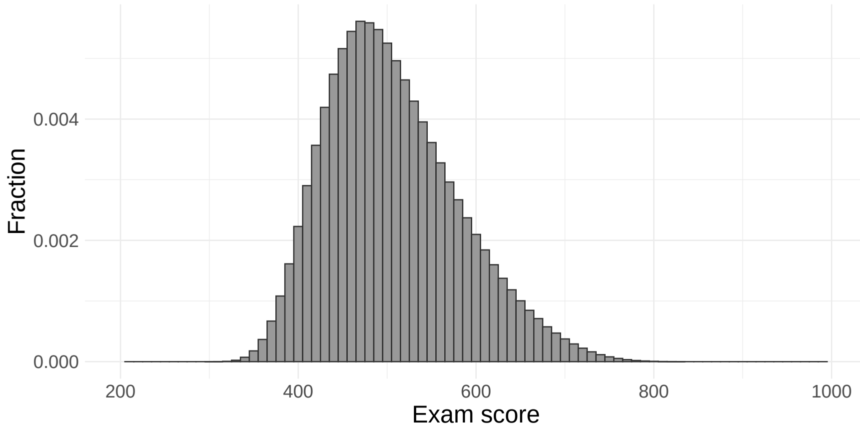
The ENEM is composed of four multiple-choice exams and an essay. We focus on the scores from the four exams - mathematics, natural sciences, social sciences, and Portuguese. The government computes the scores based on Item Response Theory, and thus the exam does not have a universal minimum or maximum. Scores are officially normalized to have a mean of 500 and a standard deviation of 100 for comparison over time. Figure 1 shows the distribution of scores in the four exams.

[7] For example, students can be eliminated from the exam if they are caught cheating.

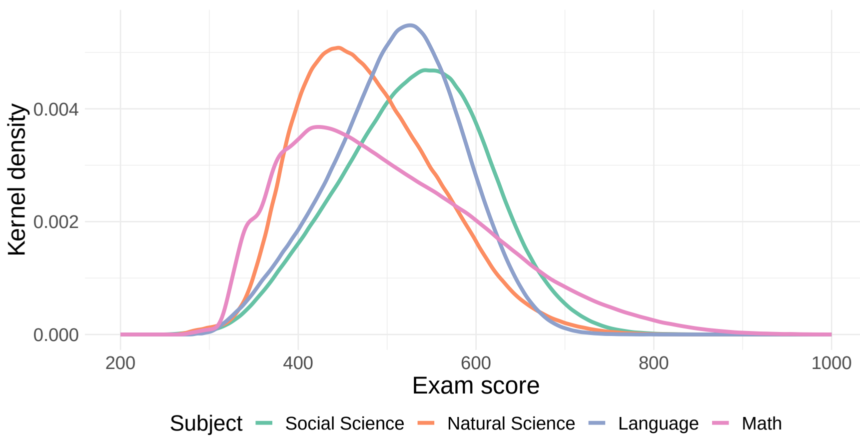
[8] The exam is not administered in every municipality, and students living in other places often take the exam in the nearest available municipality.

Figure 1: Distribution of exam scores

(a) Average across exams



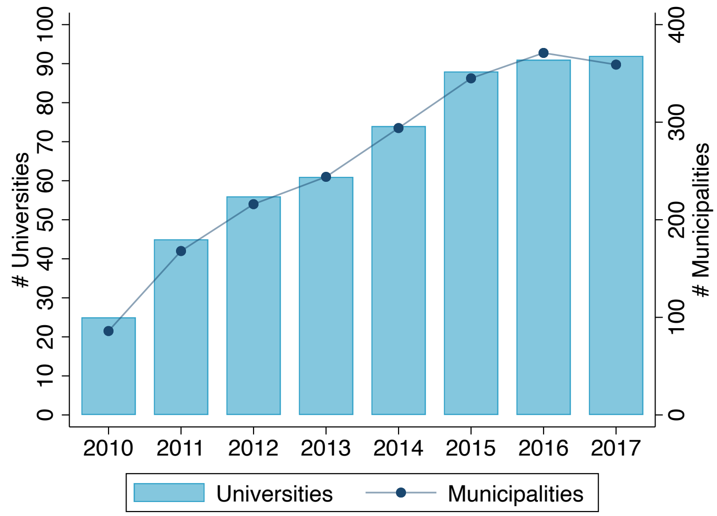
(b) By exam subject



Note: This figure shows the histogram of the score in the ENEM for all subjected pooled (Panel (a)) and kernel density (Panel (b)) for each exam - science, social science, language, and math - for 2010-2016 data. The observation is at the student-exam level.

The ministry of education provides publicly available information on the number of universities adopting SISU. The dataset contains yearly major-college level information on the number of seats offered through the system. We merge this information with the Census of Higher Education, which includes the universe of majors and colleges. Figure 2 shows the number of universities adopting SISU (left-axis) and the number of municipalities with at least one campus (right-axis) adopting SISU. As described in detail later in the paper, we use this information as time and geographic variation in the importance (stakes) of ENEM.

Figure 2: Number of universities and municipalities adopting SISU



Notes: By 2017, (i) 92 out of 192 state and federal universities have fully or partially adopted SISU, i.e., ENEM became their main or only criteria for admission; (ii) 359 municipalities out of 628 municipalities with a federal or state university campus had at least one university-campus adopting SISU.

4.2 Weather data

For weather information, we use the Princeton Global Meteorological Forcing Dataset for land surface modeling. Details of the dataset are provided in Sheffield et al. (2006). The Princeton data provides 3-hourly weather information such as temperature, humidity, and daily rainfall on a 0.25-degree global grid. Exploiting its temporal resolution, we create weather variables covering the exam period. In our main analysis, following previous studies in the literature, we focus on the effects of temperature during exams on exam performance.

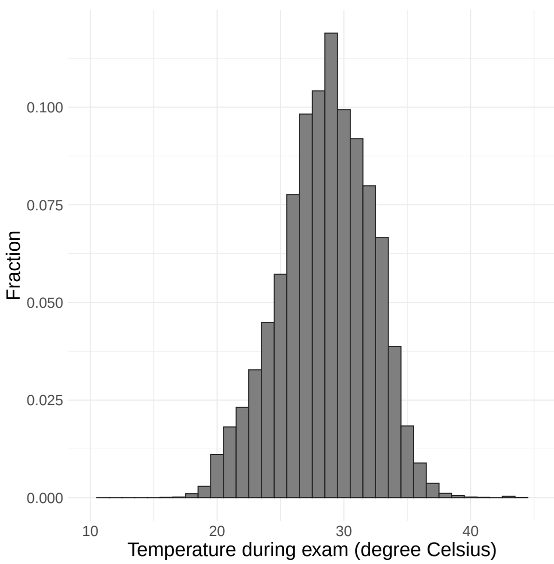
We use two different temperature measures. One is dry-bulb temperature, which we call "temperature" henceforth - which is the temperature one would usually refer to in daily life. The other is wet-bulb (WB) temperature. Wet-bulb temperature captures the interaction effect of temperature and humidity. It is calculated based on dry-bulb temperature, air pressure, and specific humidity. This measure has been used to represent heat stress danger and thermal comfort, for instance, in the climate science and biology fields (Budd, 2008; Liljegren et al., 2008). Several recent economic studies, such as Adhvaryu et al. (2020) and

Geruso and Spears (2018), have used wet-bulb temperature to account for the interactions between temperature and humidity.[9]

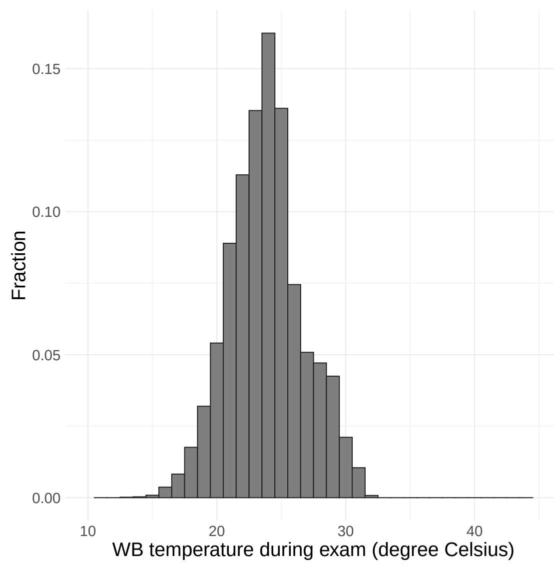
Figure 3 shows the distributions of the two temperature measures. Comparing both graphs, we see that temperatures are, on average, high (28^{o}C), while the wet-bulb measurement is, on average, 5^{o}C lower.

Figure 3: Temperature histograms (^{o}C)

(a) Temperature



(b) Wet-bulb (WB) temperature



Note: This figure shows the histograms of the average temperature during the exam over the two exam days. The observation is at the exam-day/municipality/year level.

4.3 Summary statistics

Table 2 shows summary statistics for the sub-population of exam takers and the set of municipalities used in our analysis. High-school seniors taking the exam are, on average, 17-18 years old, and 77 percent attend public high schools (either federal, state, or municipal). Note that the number of high school seniors taking ENEM increased over time. One possibility for this increase is the introduction of SISU, which affects the importance of ENEM,

[9]For more details about the weather data, how to create the weather variables covering the exam period, and how to calculate wet-bulb temperature, refer to Appendix B.

inducing more people to take the exam.[10] In section 7, we discuss selection bias due to exam take-up induced by SISU.

Table 2: Summary Statistics

	Mean	SD	Min	Max
Raw exam score	506.47	90.57	252.90	1,008.30
Temperature (degree C)	27.68	3.68	16.15	43.97
Wet-bulb Temperature (degree C)	23.17	2.83	12.83	32.22
Precipitation (mm/day)	0.03	0.05	0	0.46
Female	0.59	0.49	0	1
Age	17.52	0.83	16	20
High-income HH	0.39	0.49	0	1
High school type				
Federal HS	0.02	0.14	0	1
State HS	0.74	0.44	0	1
Municipal HS	0.01	0.10	0	1
Private HS	0.23	0.42	0	1
Gini coefficient	0.54	0.06	0.33	0.80
Share of poor	12.78	12.78	0.19	74.20
Education Development Indicator	0.66	0.08	0.27	0.81
SISU ratio (weighted, all, km)	0.44	0.22	0.03	0.93

Note: The unit of observation is subject-student. The number of observations is 32,392,992. When we include the variable "type of high schools an exam taker is from", due to a few missing values, the number of observations is 32,392,960. A household is high-income if the household's income (a categorical variable based on multiples of the minimum wage per household) is above the median income category.

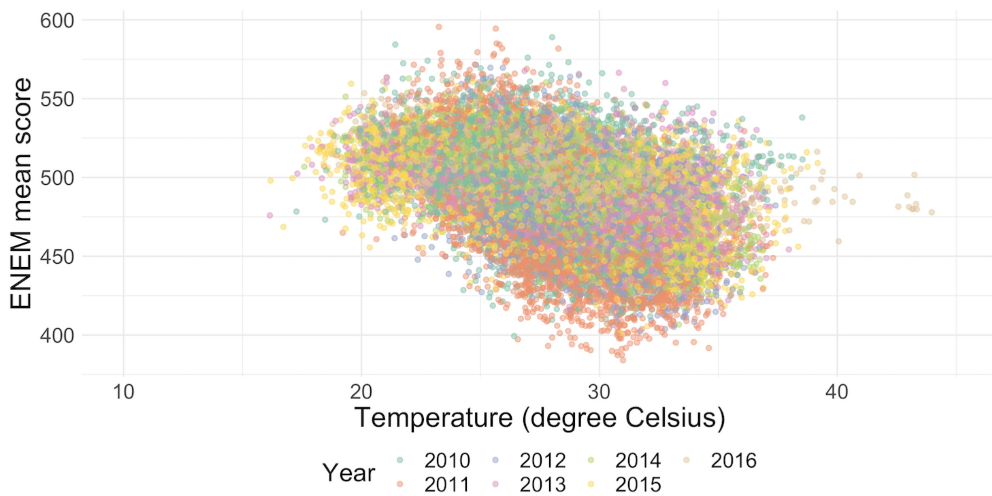
5 Main effects: how do temperature shocks affect test scores?

We start from the descriptive fact that temperature and scores are negatively correlated (Figure 4). Previous findings on the relationship between temperature and economic development suggest that much of this negative correlation is likely due to other indirect

[10]Other reasons are not directly related to this study, such as the introduction of affirmative action and other policies that provided incentives to pursue higher education, plus the potential increases in the returns to schooling, population increase, and others.

channels through which temperature can affect test scores.[11] Our identification strategy aims to identify the direct effects of temperature during the exam on exam performance.

Figure 4: Unconditional correlation between ENEM mean score and temperature.



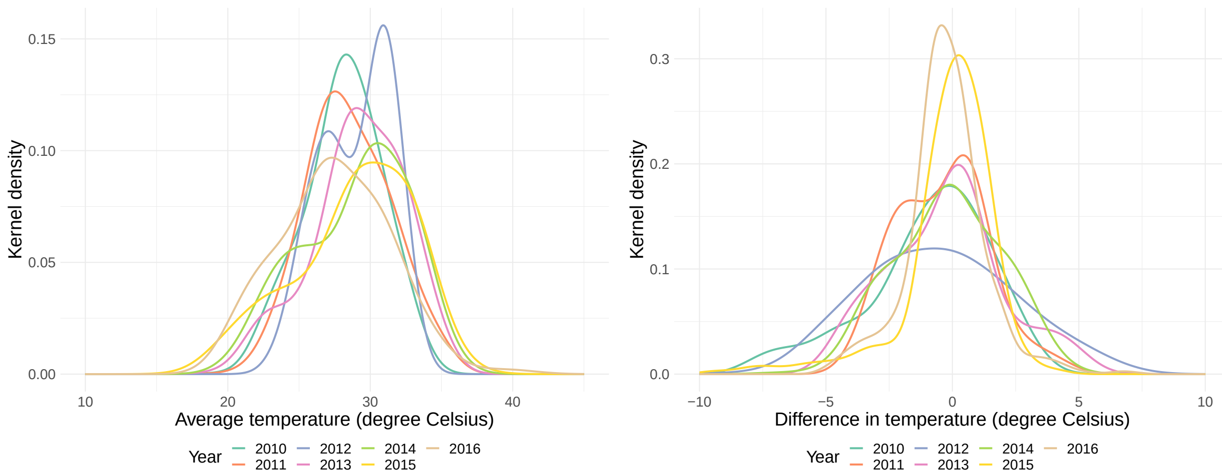
Note: The figure shows the relationship between average temperature (in Celsius) and average mean test scores at the municipality level. Mean scores are calculated as the simple average of the four multiple-choice exams, excluding the essay.

We estimate the impact of temperature on exam performance by exploiting variation in local temperature experienced by the same individual across two exam days. Figure 5 illustrates the yearly variation in temperature from 2010 to 2016. The figure shows that temperature varies across municipalities every year (panel (a)) and the two exam days in a given municipality per year (panel (b)). The cross-day variation in temperature is used to identify the effect of temperature on exam scores while controlling for individual-specific factors.

[11]See Park et al. (2020) for evidence on learning or Dell et al. (2014) for evidence related to institutional capacities.

Figure 5: Distribution of temperatures per year: pooled and difference between day 2 and day 1

(a) Pooled temperature (b) Difference (Day 2 - Day 1)



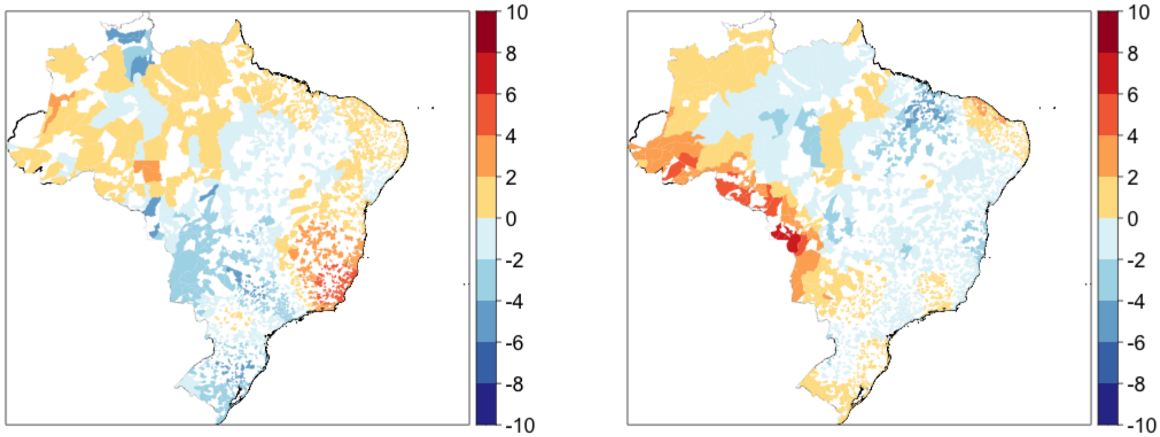
Note: The figure shows the municipality level variation in temperature (a) pooling the two days within a year, (b) the difference from day 1 to day 2 per year in Celsius.

Another important source of variation relies on temperature differences between two exam days across years within the same municipality. Figure 6 explores the within-municipality variation across years for two different years in our period of analysis, 2013 and 2016.[12] These observed yearly municipality-level variation in temperatures across exam days accounts for the possibility that the temperature difference across exam days is correlated with municipal characteristics such as long-run climate.

[12]Maps of the temperature differences across two exam days for all of the years in our data are provided in Figure E.2.

Figure 6: Variation in temperature during the exam from day one to day two, for 2013 and 2016

(a) 2013 (b) 2016



Note: The figure shows municipality-level variation in temperature from day 1 to day 2 (difference = day 2 − day 1) for 2013 and 2016. Cross-day variations for all years are shown in the appendix (E.2). Municipalities that did not have an exam site are displayed on the map in white.

Our empirical model exploits the temperature variation described above to assess the effect of temperature on exam scores. Let Y_{imsdt} be the standardized exam score of a student i in a municipality m on a subject s that was taken on a day d in year t. Raw exam scores are standardized within subject-year to have mean 0 and standard deviation 1. Also, let $f(T_{mdt})$ be a transformation of temperature T_{mdt}. As T_{mdt}, we use the dry-bulb and wet-bulb temperatures during exams. The function f can be parametric (e.g., linear function of T_{mdt}) or non-parametric (e.g., 2^oC bins of temperature). Precipitation on the exam days is included in the regressions (X_{mdt}). This variable is intended to account for the possibility that rainfall exam takers experience while traveling to exam sites affects their discomfort level and exam performance. Fixed effects included in the regression are student fixed effects (μ_i), subject fixed effects (η_s), and exam date fixed effects (τ_{dt}). The error term is represented as ϵ_{imsdt}.

Our regression equation is:

$$Y_{imsdt} = f(T_{mdt}) + X'_{mdt}\beta + \mu_i + \eta_s + \tau_{dt} + \varepsilon_{imsdt}. \tag{1}$$

The necessary identification condition is that the temperature variables, T_{mdt}, are uncorrelated with the error term, conditional on the included covariates. One potential concern is that long-run average temperature can be correlated with human capital in municipalities. If, for instance, warmer areas tend to have more low-performance students, then the correlation between heat on the exam date and students' exam performance can be spurious. In Equation (1), we exclude this possibility by including individual fixed effects, which control for unobserved municipality-level and individual-level confounders. Additionally, subject fixed effects control for persistent common differences in performance across different types of exams. Exam-date fixed effects control for average differences in mean performance between the two days and average changes in temperature due to climate cycles/change.

5.1 Results

First, we estimate Equation (1) when $f(T_{mdt})$ is linear, reported in Table 3. Column (1) and (2) contains results using the dry-bulb temperature as the temperature measurement. Column (3) serves as a robustness exercise and provides estimates when using wet-bulb temperature.

The estimates show a negative impact of high temperatures on exam scores. In column (1), a one standard deviation increase in temperature (3.679^oC) reduces exam scores by 3.6 percent[13] of a standard deviation in exam scores.

[13] Interpretation of the results is calculated from $3.679 \times (-0.00972) = 0.0357$ s.d. $\equiv$ 3.6 percent of the exam's standard deviation. The government officially normalizes the ENEM to have mean 500 and standard deviation of 100. Therefore, in the official scale, 3.6 percent of the exam's standard deviation corresponds to 3.6 score points.

Table 3: Regression results: Linear function of temperature, using ENEM Z-score

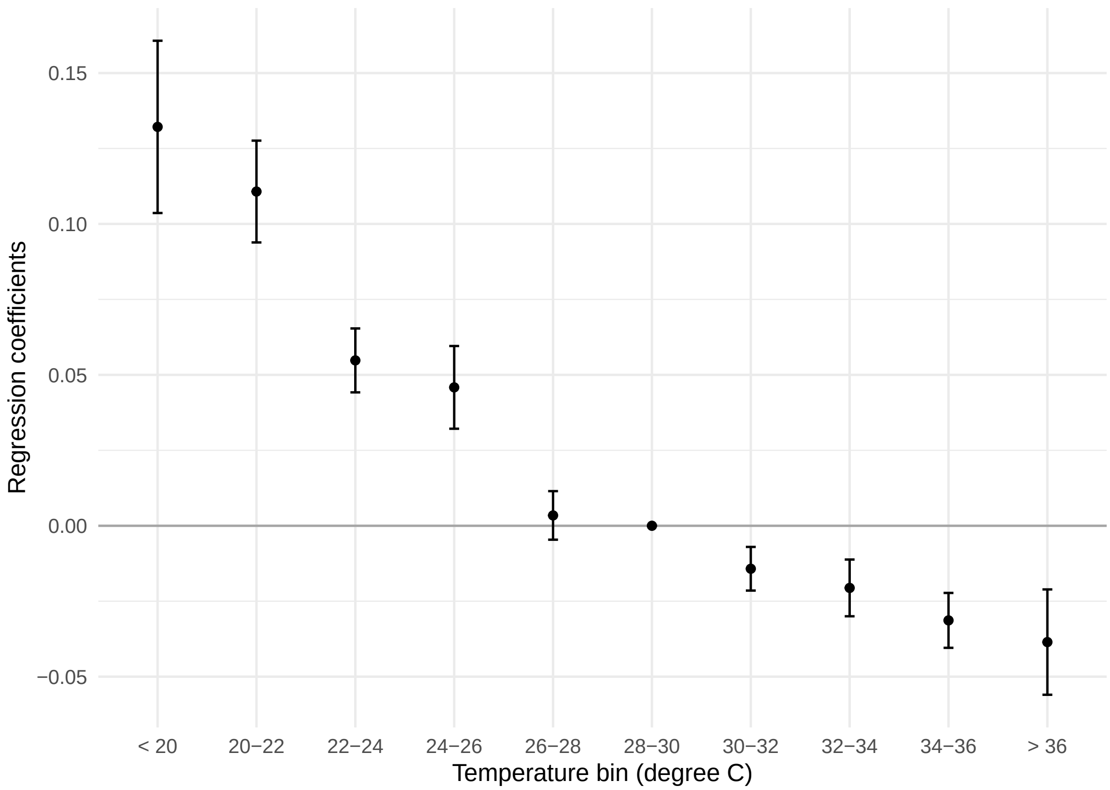
	Dependent variable: ENEM subject-score (z-score)		
	(1)	(2)	(3)
Temperature during exam	-0.00972***	-0.00968***	
	(0.00110)	(0.00100)	
Wet-bulb temperature during exam			-0.0115***
			(0.00123)
Precipitation (m/day) on exam day		0.00658	
		(0.0314)	
Observations	32,392,992	32,392,992	32,392,992
Subject, Individual, Exam date FE	Yes	Yes	Yes
Standard deviation of temperature	3.679	3.679	2.834

Note: This table presents estimates for the linear effects of temperature on exam scores. The unit of observation is subject-student-year. The dependent variable is the Z-scores of exams in each subject and year. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1. Results are robust to standard errors computed based on Conley (1999) with 200km cutoffs (see Table F.2).

Including precipitation on the exam day as an additional variable (column 2) in the regression does not change the point estimate for dry-bulb temperature, indicating that, on average, rainfall has only a negligible impact on exam scores. Results using wet-bulb temperature (column 3) are qualitatively and quantitatively similar to those using the dry-bulb temperature. A one standard deviation increase in wet-bulb temperature (2.834^{o}C) reduces exam scores by 3.3 percent standard deviation in exam scores. For simplicity, we proceed with our discussions based on the results with dry-bulb temperature. Nonetheless, the overall implications are similar if we estimate the effects using wet-bulb temperature.

Figure 7 shows regression results for the non-parametric case, in which flexible temperature effects are allowed using binned temperature. Consistent with the results based on the linear specification, these results show the negative impact of high temperature on exam scores. Our results are aligned with patterns found by Graff Zivin et al. (2020) using a similar non-parametric specification.[14] The estimates also suggest a non-linear effect of temperature on scores. Relative to the reference bin (28-30 ^{o}C), standardized exam scores increase by 0.05 in the 24-26 ^{o}C bin. Meanwhile, they decrease by 0.02 in the 32-34 ^{o}C bin.

Figure 7: Regression results: temperature and exam Z-scores



Note: The figure shows estimates of the effects of temperature on Z-scores using a flexible temperature functional form. The unit of observation is subject-student-year. The dependent variable is the Z-scores of exams in each subject and year. Error bars indicate 95% confidence intervals. Precipitation on the exam days, exam-date fixed effects, subject fixed effects, and individual fixed effects are included in the regression. Standard errors are clustered at the municipality level. Results are robust to standard errors computed based on Conley (1999) with 200km cutoffs (see Table F.2).

The negative effects of temperature on exam scores are consistent with previous findings in the literature, both qualitatively and quantitatively. For example, Park (2020) finds a decrease in 0.13 standard deviations in scores if a student takes an exam under a temperature above 90 ^{o}F (or above 32 ^{o}C) compared to a temperature below 70 ^{o}F (or below 21 ^{o}C). In

[14] Note that Graff Zivin et al. (2020) relies on county variation from the average to identify their effects of interest. In contrast, our paper relies on within-individual variation across two exam days.

our study, the increase in temperature from the 20-22 ^{o}C bin to the 30-32 ^{o}C bin decreases the exam score by 0.13 standard deviations. They also find that the magnitude of the negative impact of high temperature becomes stable above 80 ^{o}F (or above 27 ^{o}C). Potential mechanisms they suggest include (i) extremely high temperatures are rare in their study, which can undermine the power of a statistical test, and (ii) exam scores may be adjusted by graders' compensatory responses. Neither of them explains our non-linear results for the following reasons: First, in our setting, many municipalities experience high temperatures. Second, our outcome variables are based on scores from multiple-choice questions, ruling out compensatory behaviors by graders.

In another study, Li and Patel (2021) find economically and statistically insignificant null impacts of temperature on exam scores, studying the same context as ours. Detailed discussion on the differences between our study and theirs is provided in the Robustness section and Appendix D, where we perform a sensitivity analysis of our results based on their sample restrictions.

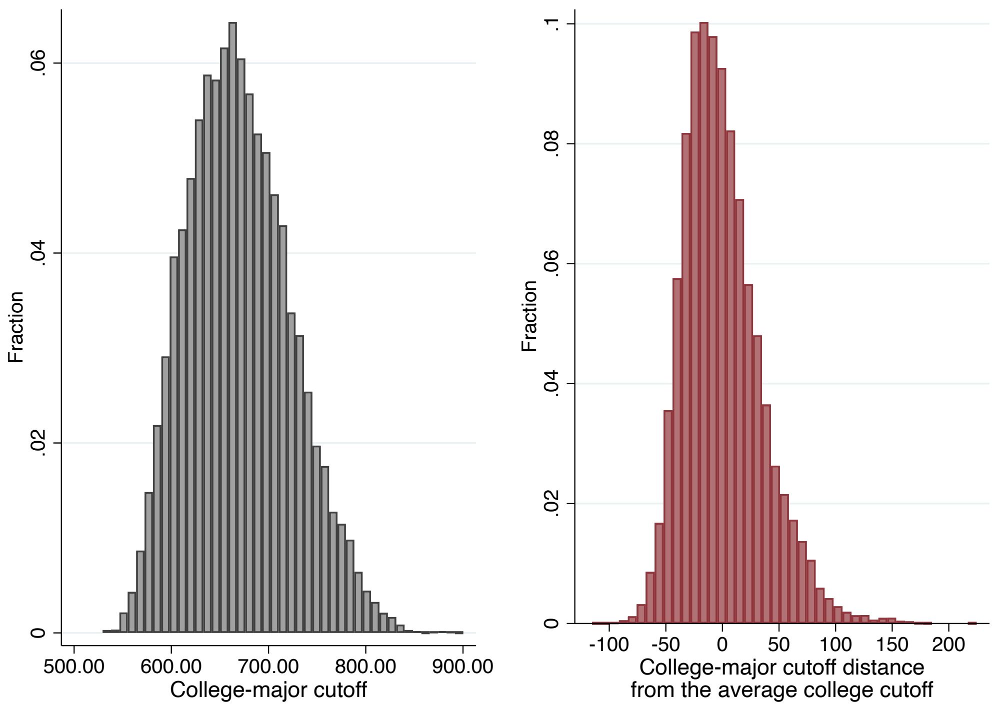
5.2 Are the effects of temperature on achievement economically significant?

We find statistically significant and non-linear effects of temperature on test scores, ranging from -0.05 to 0.12 s.d. in the non-parametric approach, averaging to 0.036 s.d. in the linear specification. Are these effects economically significant? Do they translate into changes in applicants' probability of college admission?

We provide evidence of the economic relevance of our results by calculating the change in the quantity of available college major options resulting from a temperature shock. We use yearly national data on college-major cut-offs for universities within the centralized admissions system. We calculate the difference between each major's cut-off to the average cut-off per university-campus yearly.[15] As shown in Table F.3, on average, the number of majors across college-campuses is 16, with 9 majors below the mean cut-off.

Figure 8 confirms the common perception that public college admissions are highly competitive, with a high density of majors' cut-off lying within 20 points from the average cut-off. These descriptive statistics indicate that temperatures shocks can potentially translate into economically sizable changes in the applicants' set of major options, particularly for those applicants scoring at the college's cut-off average.

Figure 8: Distribution of majors' cut-off and share of majors per college by distance between the major's cutoff and the college's average major cutoff

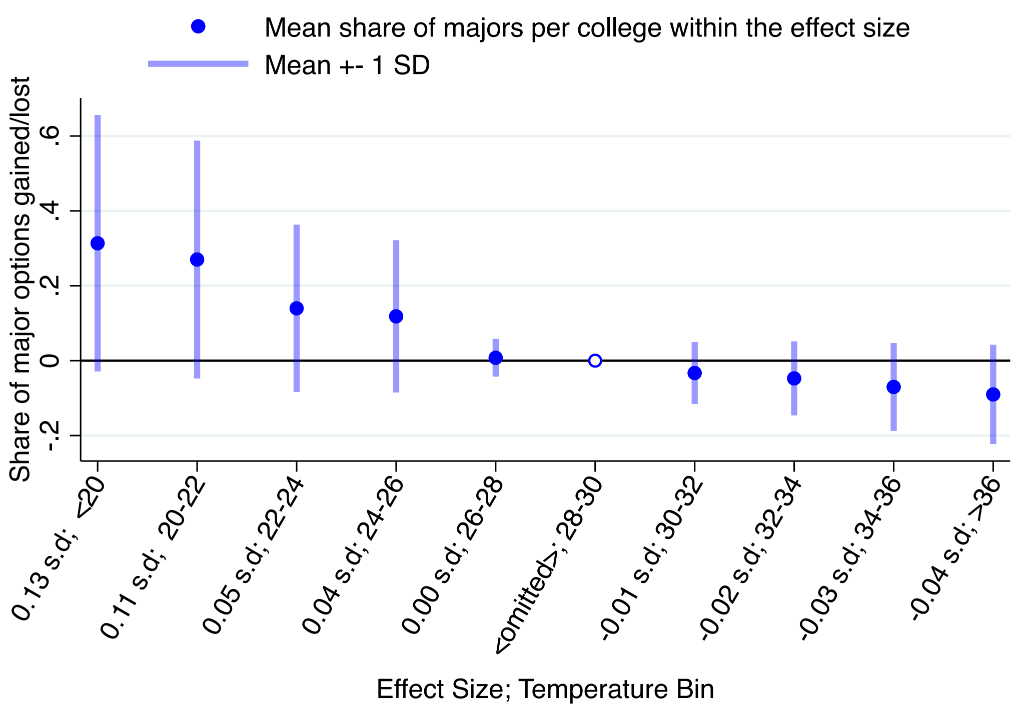


Note: we use data on college-major cut-offs of the non-quota group for all majors offering seats in the centralized system platform from 2011 to 2017 (corresponding to ENEM 2010-2016, the period of our study). ENEM scores are standardized by the government to have mean 500 and standard deviation 100. We exclude campuses with less than five majors - results including these campuses are robust and larger. Estimates include 83 state and federal universities, 1,196 campuses, and 6,272 majors.

[15]With the adoption of affirmative action policies, each college-major option had several different cut-offs, one for each non-quota and quota groups. We use the non-quota group cut-off as the reference in our calculations.

Quantifying the main effects in terms of applicants' set of attainable majors, we conclude the results are economically sizable. We find that, a one standard deviation increase in temperature reduces the set of options available to the average applicant by 8.2 percent. Figure 9 shows results based on the effect in each temperature bin. For applicants with scores at the average college-major cut-off, the range of temperature shocks we estimated in our paper translate into a variation in the number of attainable majors from roughly minus 10 percent to plus 30 percent, on average.

Figure 9: Average change in students' choice set of majors



Note: This figure shows calculations for the average share of majors per college for each effect size relative the average major cut-off per college. We use data on college-major cut-offs of the non-quota group for all majors offering seats in the centralized system platform from 2011 to 2017 (corresponding to ENEM 2010-2016, the period of our study). ENEM scores are standardized by government to have mean 500 and standard deviation 100. We exclude campuses with less than five majors - results including these campuses are robust and larger. Estimates include 83 state and federal universities, 1,196 campuses, and 6,272 majors.

6 Heterogeneous effects: how does the temperature effect interact with the exam stakes?

Our study also investigates exam stakes as a mechanism behind the relationship between temperature and exam scores. For this purpose, we include an interaction term between temperature and a unique measure of exam stakes in our main estimation equation and analyze how the temperature effect changes as exam stakes vary.

More specifically, we estimate the following regression:

$$Y_{imsdt} = f(T_{mdt}) + \theta\left(f(T_{mdt}) \times H_{mt}\right) + \mu_i + \delta_s + \tau_d + \epsilon_{ismdt}, \tag{2}$$

where H_{mt} is a proxy for exam stakes.

We measure exam stakes by exploiting variation in the number of universities adopting a centralized admissions system. When a university joins this system, the national exam becomes a necessary and often the sole criterion for college admissions, raising the exam stakes for students applying for college.

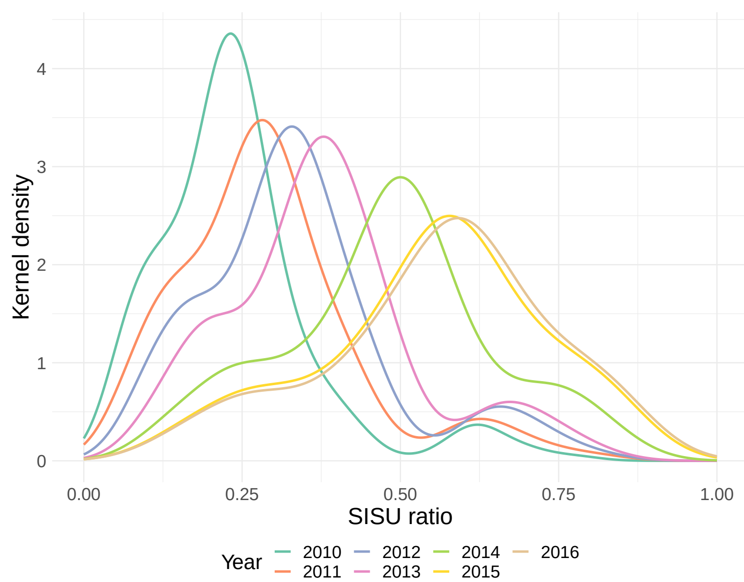
A few contextual facts underlie our identification strategy. First, while all state and federal universities were allowed to participate in the system, adoption was not mandatory, with the timing of adoption varying across universities. Second, there are high migration costs for college purposes in Brazil, with substantial economic barriers for students interested in attending university outside their hometown in Brazil. For instance, existing student loan programs in Brazil do not allow students to cover living expenses, being exclusive for tuition payments. Housing provided by universities is rare and often allocated to extremely low-income applicants. Therefore, moving to another state to attend a college is expensive, which might explain why most college students attend universities in their home state ($\approx$ 90 percent of college students). Given the high mobility costs, we expect applicants residing closer to a university to be more affected by the SISU adoption than a student living further from that university.

To capture both the timing and geographic variation induced by universities adopting the new system, we create a municipality-level variable based on the proportion of universities adopting SISU nationwide, with larger weights for municipalities closer to the municipality where a student takes the ENEM. The variable H_{mt} is then calculated using the following equation, corresponding to the proportion of universities adopting SISU at a municipality, weighted by the geographic distance between municipalities.[16]

$$H_{mt} = \frac{\sum_{n \in \text{all municipalities in Brazil}} w_{nm} (\#\text{ universities adopting SISU})_{nt}}{\sum_{n \in \text{all municipalities in Brazil}} w_{nm} (\#\text{ universities})_{nt}} \tag{3}$$

where w_{nm} are weights defined as $w_{nm} = \frac{1}{1+(\text{Distance between } n \text{ and } m \text{ (km)})}$. Figure 10 shows the H_{mt} distribution by year and illustrates the increase in SISU adoption over time.

Figure 10: Distribution of H_{mt} in each year



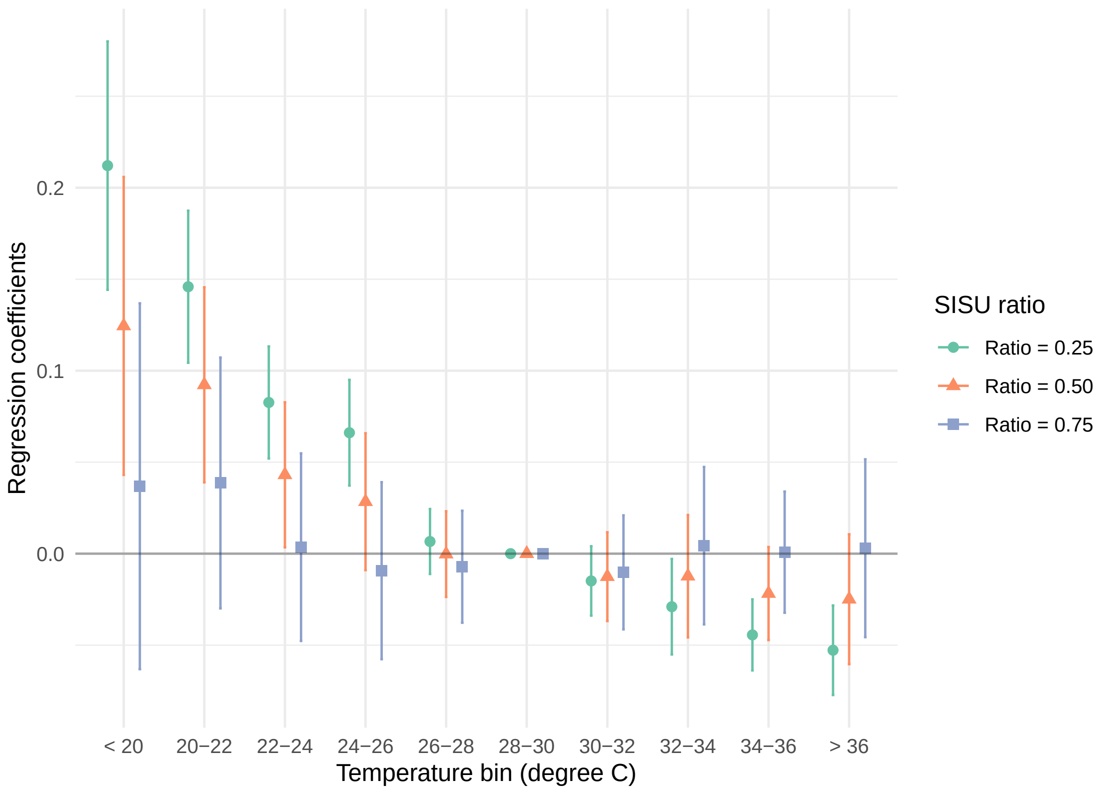
Note: The figure shows the distribution of H_{mt}, the SISU adoption ratio, by year. The unit of observation is municipality-year. The construction of the variable is described in detail in the main text.

[16]The calculation of H_{mt} is illustrated in Appendix C with a toy example.

6.1 Results

Estimates indicate both economically and statistically significant impacts of exam stakes on temperature effects. Figure 11 shows the regression results using binned temperature, illustrating the attenuating effects of the increased stakes induced by the centralized system. As the ratio increases (more universities adopt the system), the temperature effects across bins become smaller.

Figure 11: Regression coefficients of temperature $\times$ H_{mt}



Notes: Effects are calculated based on point estimates from regressions of Z-scores on temperature and its interaction with H_{mt}, the SISU adoption ratio. Error bars indicate 95% confidence intervals. Municipalities whose distance from the municipality where a student took ENEM is less than 60km are used, and the inverse of the distance between municipalities are used as weights. Precipitation is controlled for in all regressions. Standard errors are clustered at the municipality level. Results are robust to standard errors computed based on Conley (1999) with 200km cutoffs (see Table E.3).

Based on the theoretical model, these results suggest that an increase in exam stakes induces students to exert more effort, attenuating the adverse effects of high temperatures on performance. To put this result in perspective with the main ones, the changes due to the increased stakes correspond, on average, to a 58 percent reduction in the effect of temperature on exam scores. In the lowest bin, the temperature effect decreases by over 80 percent if the SISU adoption ratio increases from 0.25 to 0.75.

These results shed light on an important mechanism behind the effect of temperature on cognitive performance: temperature affects the level of effort, which changes the outputs of cognitive tasks. Exploiting the staggered adoption of SISU, we provide the first empirical evidence that the increase in exam stakes mitigates the temperature effect.

Motivated by previous findings in the literature,[17] we extend our analysis to test whether we find gender differences in (i) the direct effect of temperature on performance; (ii) the interacted effects of temperature changes and exam stakes. The results for the non-parametric (Figure E.4) transformations of T_{mdt} in Equation 1, estimated by male and female separately, show that girls and boys similarly under-perform if the temperatures during the exam are high.

When considering the varying stakes, we find suggestive evidence of differences in effort responses across gender. Figure E.5 shows the estimates of Equation 2 by male and female separately. We find that females' test scores are less responsive to a variation in stakes from relatively low to high than males', specifically for lower temperature bins. These findings align with the literature on gender differences in performance across different exam stakes, which suggests that males respond to the increased pressure more successfully than females (Azmat et al., 2016; Schlosser et al., 2019). However, since these differences are not statistically significant at conventional levels, we interpret these results as suggestive yet inconclusive.

[17] Differential responses in performance to temperature documented in different contexts shows that women outperform men at higher temperatures (Chang and Kajackaite, 2019; Lee et al., 2021). Additionally, there is a well-documented difference in how males and females respond to high vs. low stakes exams (Azmat et al., 2016; Schlosser et al., 2019)

7 Robustness checks

7.1 Confounders related to the adoption of the centralized admissions system

One potential concern is that our measure of exam stakes - the SISU ratio - may capture other factors affecting the relationship between temperature and exam scores. We consider three possibilities: (i) selection bias; (ii) endogenous SISU adoption; (iii) time-varying factors.

First, SISU adoption can change the composition of exam takers. As described before, SISU adoption increases the exam stakes as it becomes a necessary condition for college admissions. This change in stakes can induce more students to register and take the exam. For example, if students induced to take ENEM are less affected by temperature, the observed results could be partially attributed to this compositional effect. In fact, as seen in Table F.1, there has been a significant increase in the number of ENEM takers following the introduction of the centralized system, increasing from 4.6 in 2010 to 8.6 million in 2016, peaking in 2014 at 8.7 million applicants. In our restricted sub-population of high school seniors, the number of exam takers is relatively more stable, increasing from about 900 thousand to 1.3 million between 2010 and 2016. We add to the regression equation interaction terms of temperature and individual-level variables to deal with the possibility that temperature effects might differ based on students' characteristics. We include an indicator of high-income households (above median income) and the type of high school (federal, state, municipal, and private).

Second, SISU adoption by universities can be endogenous. SISU may be adopted in regions with better education systems and high-quality educational infrastructure. Exam takers in these regions can be less affected by temperature. If this is the case, the positive coefficients of the interaction between temperature and SISU adoption ratio could be caused by the municipality-level factors correlated with SISU adoption. We include interaction terms of temperature and municipality-level variables to investigate this concern. The municipality-level variables included are Gini coefficients, poverty rate, and education indexes.

Finally, there may be omitted time-varying factors. For instance, more universities may adopt SISU while exam sites may install air conditioners, which can also reduce the temperature effect on exam scores. We include interactions between temperature and year dummies to account for this possibility and other time-varying factors. Equation (4) includes all interaction terms described above. Our coefficient of interest is θ, which captures how SISU adoption affects individual responses to temperature.

$$
\begin{aligned}
Y_{imsdt} \\
&= f(T_{mdt}) + \theta \left(f(T_{mdt}) \times H_{mt} \right) \\
&\quad + \gamma_1 \left(f(T_{mdt}) \times X_{imt} \right) + \gamma_2 \left(f(T_{mdt}) \times Z_{mt} \right) + \gamma_3 \left(f(T_{mdt}) \times \sum_{t=2010}^{t=2016} Year_t \right) \\
&\quad + \mu_i + \delta_s + \tau_d + \epsilon_{ismdt},
\end{aligned} \tag{4}
$$

Panel A of Table 4, columns (2)-(5), shows that the interaction terms have limited effects on the estimated coefficient of the main interaction term between temperature and the SISU adoption ratio.[18] Including interactions between temperature and individual-level variables (column (2)) and municipality-level variables (column (3)) barely changes the estimates of interest.

However, including the interactions with year dummies reduces the coefficient of interest (column (4)) more substantially. These smaller estimates could be due, for instance, to an increase in air conditioner installation or other time-varying factors correlated with SISU adoption. Nonetheless, the mitigating effect of the SISU ratio remains statistically and

[18]Table F.4 provides estimates of all interaction terms used in the regressions.

economically significant. Finally, when considering all factors, increasing the SISU ratio by one standard deviation reduces the temperature effect by 42 percent. This suggests that the results in Table 4 may capture effects of non-SISU factors to some extent but not entirely.

Table 4: Regression results: Interaction effects of temperature and ENEM stakes on exam Z-score

	Dependent variable: ENEM subject-score (z-score)				
		Robustness			
	Main	Selection Bias Bias	Endogenous SISU	Time-varying factors	All
	(1)	(2)	(3)	(4)	(5)
Panel A: Main SISU measure (Extensive Margin)					
Temp. × SISU ratio (extensive margin)	0.0254*** (0.00474)	0.0246*** (0.00477)	0.0252*** (0.00514)	0.0186*** (0.00480)	0.0180*** (0.00478)
Observations	32,392,992	32,392,960	32,392,992	32,392,992	32,392,960
Standard dev. SISU ratio	0.2219	0.2219	0.2219	0.2219	0.2219
Average temperature effect	-0.00968	-0.00968	-0.00968	-0.00968	-0.00968
Panel B: Alternative SISU measure (Intensive Margin), years: 2010-2014					
Temp. × SISU ratio (intensive margin)	0.0244*** (0.0056)	0.0234*** (0.0057)	0.0234*** (0.0049)	0.0243*** (0.0061)	0.0214*** (0.0050)
Observations	22,090,472	22,090,472	22,090,472	22,090,472	22090472
Standard dev. SISU ratio	0.2221	0.2221	0.2221	0.2221	0.2221
Average temperature effect	-0.0131	-0.0131	-0.0131	-0.0131	-0.0131
Subj., Ind., Exam data FE	Yes	Yes	Yes	Yes	Yes
Ind. Var. Interactions	No	Yes	No	No	Yes
Mun. Var. Interactions	No	No	Yes	No	Yes
Year Interactions	No	No	No	Yes	Yes

Notes: Panel A shows results for the robustness checks for the heterogenous effects induced by SISU adoption. Extensive margin refers to using a binary measure to classify whether a university allocated any seats to SISU. Panel B shows robustness results using an alternative measure that captures the intensive margin of SISU adoption: the proportion of seats allocated to SISU. Data is from Mello (2022), which is restricted to years 2010-2014 (SISU 2011-2015).
Average temperature effect is the estimated coefficient $\widehat{\alpha}$ from a regression equation, $Y_{imsdt} = \alpha T_{mdt} + X'_{mdt}\beta + \mu_i + \eta_s + \tau_{dt} + \varepsilon_{imsdt}$. Precipitation on exam days is controlled for in all regressions. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1. Results are robust to standard errors computed based on Conley (1999) with 200km cutoffs (see Table F.5).

Next, we will discuss three main potential factors that could explain the time-varying counfounders: (i) changes in SISU adoption at the intensive margin; (ii) other correlated policies that affect exam stakes, particularly affirmative action policies; and (iii) air-conditioning coverage.

7.1.1 Intensive margin of the centralized system adoption

Our main empirical strategy exploits variation in the adoption of the centralized system using a binary adoption definition. However, although many universities allocated 100 percent of their seats to SISU once they joined, there was still substantial yearly variation (Figure E.7). One possibility is that the time-varying effects we observe in Column (4), Panel A, Table 4 capture more granular changes in the adoption of the centralized system.

To investigate for this possibility, we use an alternative measure for H_{mt} that exploits intensive changes in stakes, the number of seats allocated to SISU. We use data from Mello (2022), which covers a sub-set of our study's years: 2010-2014. Results in Panel B of Table 4 show that the mitigating effects of SISU are reasonably robust and comparable to the intensive margin. More importantly, when adding temperature-year interactions, we see virtually no change in the coefficient of interest. The findings suggest that the time-varying effects captured at the extensive margin are driven by yearly variation in the intensity of exam stakes.

7.1.2 Alternative policy affecting exam stakes: affirmative action

The centralized admissions system is the policy change that most directly affects exam stakes for all applicants. However, another important time-varying factor during the period of our study is the adoption of affirmative action (AA) policies, which were first enacted in Brazil in the early 2000s.[19] Until 2012, universities individually adopted affirmative action

[19]Since 2010, another policy was also announced that affects admissions to college and the college entrance exam take-up: the expansion of the federal student loan system (FIES - *Fundo de Financiamento Estudantil*).

either by internal deliberation or due to state laws. In 2012, a federal affirmative action law was enacted, mandating that all federal institutes of higher education reserve 50 percent of their seats to students from public high schools, with additional income and race sub-criteria. The 2012 Quotas Law primarily affected the number of universities and proportion of seats allocated to affirmative action (see Figures E.6 and E.7), changing the relative probabilities of admissions between beneficiaries and non-beneficiaries in unprecedented ways.

Unlike the centralized system, AA policies do not affect how students apply to college or the exam they take. Yet, it likely affects the perceived importance of the college entrance exam, changing, like the centralized system, both the probability of exam taking and the returns to effort. These two policies are often and increasingly adopted together (Figure E.6), but at different intensities (Figure E.7). AA policies are more widely adopted, whereas the centralized system started a decade after AA and is adopted by fewer universities. When the two policies are adopted together, AA seats are often offered through SISU. Previous studies have discussed the effects of affirmative action on exam take-up and college enrollment as well as the combined impact of these two admissions policies (Mello, 2022; Mello and Melo, 2023; Otero et al., 2021).

To test for the extent AA policies are raising the exam stakes, we estimate Equation 2 with an alternative measure of exam stakes as a function of the adoption of affirmative action by universities, also using the 2010-2014 data from Mello (2022). We also provide evidence for the simultaneous effects of the two policies by jointly estimating their effects at the intensive margin.

In Columns (1) and (2) of Table F.6, we show that SISU and AA have significant and

This federal program provides low-interest loans to low-income applicants to attend college at a private institution. The criteria for admissions are mostly income-based. However, there is a minimum ENEM score of 450 that students need in order to qualify, and not used to rank students. In the 2010 ENEM, 77 percent of applicants scored above the minimum threshold to qualify for federal student loans. Therefore, regarding raising the stakes and incentives to perform in the exam, we believe the federal loan program has relatively limited influence relative to the centralized system or affirmative action.

comparable effects. A one standard deviation increase in the intensive SISU ratio reduces the average temperature effect by 41 percent. In comparison, a one standard deviation increase in the intensive AA ratio reduces the average temperature effect by 52 percent. When considered together (Column (3)), we see that both policies jointly mitigate the impact of temperature on achievement.

Overall, these consistent results corroborate our main hypothesis of the interacting effects of exam stakes and individual responses to temperature shocks. The results remain virtually the same after accounting for potential selection bias, endogenous SISU adoption, and other time-varying factors (Column (4)). Notably, the robustness of these alternative results to the inclusion of time-varying effects is also relevant supportive evidence that estimates are not capturing other potential time-varying confounders.

7.1.3 Air-conditioning

Air conditioning (AC) usage can potentially downward bias our results since room temperature control mitigates the effects of higher temperatures. Unfortunately, we do not have data on air conditioning usage during the exam to test this for this possibility. Alternatively, we provide contextual details on temperature, ventilation, and AC coverage across schools and households, arguing that air conditioning coverage in exam locations is likely low.

First, we provide information regarding classroom conditions in high schools. To the best of our knowledge, high school classrooms are the most common exam sites for ENEM. A 2019 national survey administered to school teachers asked them to report classroom conditions related to natural ventilation and temperature.[20] According to teachers' reports, about 50 percent of schools have inadequate ventilation. As for temperature, 57 percent of

[20]Teacher module, SAEB 2019 - Sistema de Avaliação da Educacão Básica (Saeb). Questions relative to ventilation and temperature are based on a scale of 1 (Inadequate) to 4 (Adequate). We interpret answers less than 4 (adequate) as less than adequate.

classrooms have less than adequate temperature. Although the temperature question in the survey does not specify air conditioning usage, the high share of classrooms with inadequate ventilation and uncomfortable temperatures suggests that most high schools do not use an air conditioning system. Second, we assessed information on AC coverage in households nationwide as another proxy for general AC coverage in the country. For instance, in 2019, 16 percent of households in Brazil reported having air conditioners. Moreover, virtually no household has central air conditioning: 99.5 percent report having a window, portable, or a split unit.[21]

Even though we cannot rule out that AC adoption increased over the years of our study, this descriptive data supports the idea that the diffusion of air conditioners is limited in Brazil. Moreover, our previous empirical exercises showed that most of the time-varying effects are more likely to be driven by yearly variations in the adoption of the centralized system and the interacting effects of SISU and AA, with little room for variations in AC adoption to be driving our results.

7.2 Other SISU ratios

Another robustness check we conduct investigates if our results are driven by how we construct the SISU ratio. For this, we attempt two changes. First, we change the municipalities used to calculate the SISU ratio. In our main estimates, all municipalities in Brazil are included in the calculations of the SISU ratio. That assumes that all exam takers demanding higher education, that is, taking the ENEM, can be potentially affected by any university adopting SISU. However, it is more likely that only universities in neighboring municipalities may affect exam takers.

Alternatively, we use the proportion of universities adopting SISU in neighboring mu-

[21] Source: ELETROBRAS. Relatório de resultados do Procel 2020: ano base 2019. Rio de Janeiro: PROCEL, 2020.

nicipalities to calculate the ratio. The set $\mathcal{M}_m$ represents the set of municipalities included in this calculation, for which we use two definitions: (i) municipalities within 60km from a municipality m, and (ii) in the same microregion (defined mainly by a commuting zone) as m. We provide estimates with and without distance weighting to calculate the ratio. Specifically, we use the following formula:

$$H_{mt} = \begin{cases} 0 & \text{if there is no university in any municipalities in } \mathcal{M}_m \\ \frac{\sum_{n\in\mathcal{M}_m} w_{nm}(\#\text{ universities adopting SISU})}{\sum_{n\in\mathcal{M}_m} w_{nm}(\#\text{ universities})} & \text{otherwise.} \end{cases}$$

We expect the different measures to change the estimates' magnitudes since a fraction of exam takers that were previously considered treated to some extent are now treated by a SISU ratio equal to zero. Our robustness check relies on the qualitative interpretation of the coefficients of interest. Results in Table F.7 show qualitatively similar results that SISU mitigates the effects of temperature, which supports the robustness of our empirical results.

Lastly, we also provide estimates for different units of distance used to calculate the weights from kilometers to meters and miles (Table F.8).

7.3 Spatial correlation

To account for the spatial correlation beyond municipalities, we also use the standard errors proposed by Conley (1999) for inference.[22] They take into account the dependence due to geographical proximity. First, temperature variations can be similar across neighboring municipalities. Second, exam takers in these close municipalities experience similar changes in exam stakes. Tables F.2 and F.5, and Figure E.3 show that our results are robust to the use of this alternative method to calculate standard errors does not affect the statistical significance.

[22] We use 200km as a cutoff to account for the spatial autocorrelation.

7.4 Other robustness checks

Finally, we perform a series of checks to contextualize our results on the direct effects of temperature with those in another related study (Li and Patel, 2021). Regression outputs and a detailed discussion of how results from these alternative research choices differ from our primary analysis are discussed in Appendix D.

In summary, the different temperature data and frequencies explain the distinct average effects of temperature on performance between ours and their studies. Our paper uses high-frequency temperature variation, allowing us to isolate the temperature *during* the exam, whereas theirs use daily average temperature. Second, their paper selects exam takers 14 to 22 years old, including high-school seniors and individuals who decided to take (or retake) the exam one or more years after graduating high school. Instead, we restrict our analysis to high school seniors. Since ENEM is used for college admissions, high school seniors are most likely taking the exam for the first time and are a more homogeneous group. Third, their estimates mix outcomes for multiple-choice and essay questions, whereas we restrict to multiple-choice exams. We choose not to use essay scores in the primary analyses for the following reasons: (i) Since the essay is not a multiple-choice exam, we expect the nature of the exam to be different from other subjects; (ii) Since humans grade the essays, the temperature can affect grading (Park, 2020), which prevents us from isolating the temperature effect on exam takers' performance.

8 Conclusion

This paper provides new evidence on the effects of temperature on achievement and evaluates an important channel through which temperature affects exam scores: effort. Our theoretical framework suggests that temperature can affect performance through cognitive and effort channels. We derive two hypotheses. One is that temperature negatively affects exam scores. The other is that these adverse effects are lower as the exam stakes increase, mitigated by compensatory changes in effort. Our identification strategy exploits within-individual variation in temperature across two consecutive exam days. Using data on millions of exam takers in a national standardized exam in Brazil, we estimate the differential effects of temperature by exam stakes. A unique feature of the Brazilian context provides variation in stakes, where a national exam's stakes increase from relatively low to high.

Our paper provides the first evidence on how exam stakes mitigate the effects of temperature on exam scores. Our baseline results show that temperature negatively impacts exam scores. These effects are comparable to other studies in China and the US. Moreover, we provide evidence that these effects are economically sizable, with substantial implications for applicants' college major options When exploiting the variation in exam stakes, we find that the higher the stakes, the lower the effects of temperature, suggesting that effort mitigates the effects of temperature on performance.

The understanding that low-stakes exams are affected by motivation and effort during the exam is largely discussed in the literature (see Finn (2015) for a review). Our paper shows that the harmful effects of temperature on performance are less of a concern if stakes are sufficiently high, such as when admissions to selective universities are exclusively based on one exam's outcome. Yet, at selective environments, even small changes in scores induced by inadequate temperature control can lead to large consequences in college applicants' admissions probabilities. At lower stakes contexts, students have lower incentives to compensate for the negative effects of temperature when exams are not directly linked to their outcomes. These findings are particularly relevant since low stakes test scores are widely used to allocate financial resources to schools, college seats, teacher's bonuses, and rank countries. Negative effects of temperature can result in inaccurate rankings and translate into unequal redistribution of resources or biased cross-country evaluations.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2020). The light and the heat: Productivity co-benefits of energy-saving technology. *The Review of Economics and Statistics 102*(4), 779–792.

Azmat, G., C. Calsamiglia, and N. Iriberri (2016). Gender differences in response to big stakes. *Journal of the European Economic Association 14*(6), 1372–1400.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences 113*(47), 13354–13359.

Budd, G. M. (2008). Wet-bulb globe temperature (WBGT) - Its history and its limitations. *Journal of Science and Medicine in Sport 11*(1), 20–32.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In L. N. Christofides, K. E. Grant, and R. Swidinsky (Eds.), *Aspects of labour market behavior: Essays in Honour of JOhn Vanderkamp*, pp. 201–222. University of Toronto Press.

Chang, T. Y. and A. Kajackaite (2019, 05). Battle for the thermostat: Gender and the effect of temperature on cognitive performance. *PLOS ONE 14*(5), 1–10.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

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

Ebenstein, A., V. Lavy, and S. Roth (2016, oct). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics 8*(4), 36–65.

Finn, B. (2015). Measuring motivation in low-stakes assessments. Research Report Series RR-15-19, Educational Testing Service.

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

Geruso, M. and D. Spears (2018, July). Heat, humidity, and infant mortality in the developing world. Working Paper 24870, National Bureau of Economic Research.

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

Graff Zivin, J., Y. Song, Q. Tang, and P. Zhang (2020). Temperature and high-stakes cognitive performance: Evidence from the national college entrance examination in China. *Journal of Environmental Economics and Management 104*, 102365.

Hyman, J. (2017, 07). ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy 12*(3), 281–311.

Lee, Y., B. Haile, G. Seymour, and C. Azzarri (2021, June). The heat never bothered me anyway: Gender-specific response of agricultural labor to climatic shocks in Tanzania. *Applied Economic Perspectives and Policy 43*(2), 732–749.

Li, X. and P. C. Patel (2021). Weather and high-stakes exam performance: Evidence from student-level administrative data in Brazil. *Economics Letters 199*, 109698.

Liljegren, J. C., R. A. Carhart, P. Lawday, S. Tschopp, and R. Sharp (2008). Modeling the wet bulb globe temperature using standard meteorological measurements. *Journal of Occupational and Environmental Hygiene 5*(10), 645–655.

Lim, C. L., C. Byrne, and J. K. Lee (2008). Human thermoregulation and measurement of body temperature in exercise and clinical settings. *Annals Academy of Medicine Singapore 37*(4), 347.

Machado, C. and C. Szerman (2021). Centralized college admissions and student composition. *Economics of Education Review 85*, 102184.

Mello, U. (2022, August). Centralized admissions, affirmative action, and access of low-income students to higher education. *American Economic Journal: Economic Policy 14*(3), 166–97.

Mello, U. and A. P. Melo (2023). Affirmative action and demand for schooling: evidence from nation-wide policies. *Unpublished Manuscript*.

Otero, S., N. Barahona, and C. Dobbin (2021). Affirmative action in centralized college admission systems: Evidence from brazil. *Unpublished Manuscript*.

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

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

Reardon, S. F., D. Kalogrides, E. M. Fahle, A. Podolsky, and R. C. Zárate (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades. *Educational Researcher 47*(5), 284–294.

Schlosser, A., Z. Neeman, and Y. Attali (2019, 05). Differential performance in high versus low stakes tests: Evidence from the GRE Test. *The Economic Journal 129*(623), 2916–2948.

Sheffield, J., G. Goteti, E. F. Wood, J. Sheffield, G. Goteti, and E. F. Wood (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate 19*(13), 3088–3111.

A Proof of $\frac{dy}{da} < 0$

Remember that the first order condition (FOC) is

$$\frac{\partial u_2}{\partial w}\frac{\partial w}{\partial y}\frac{\partial y}{\partial e^*} + \frac{\partial u_1}{\partial e^*} = 0,$$

and we make the following assumptions:

(i) higher future wages increase utility $\left(\frac{\partial u_2}{\partial w} > 0\right)$,

(ii) exerting effort is costly $\left(\frac{\partial u_1}{\partial e} < 0\right)$,

(iii) a higher temperature gives discomfort and decreases utility $\left(\frac{\partial u_1}{\partial a} < 0\right)$,

(iv) marginal returns to future wages diminish $\left(\frac{\partial^2 u_2}{\partial w^2} < 0\right)$,

(v) the cost of effort to utility is convex $\left(\frac{\partial^2 u_1}{\partial e^2} < 0\right)$,

(vi) an effort cost is higher under a hotter environment $\left(\frac{\partial^2 u_1}{\partial e \partial a} < 0\right)$,

(vii) a positive relationship between exam score and future wages $\left(\frac{\partial w}{\partial y} > 0\right)$,

(viii) effort increases scores $\left(\frac{\partial y}{\partial e} > 0\right)$,

(ix) the effort effect diminishes $\left(\frac{\partial^2 y}{\partial e^2} < 0\right)$,

(x) effort is less effective in improve the test score when temperature is higher due to cognitive impairment $\left(\frac{\partial^2 y}{\partial e \partial a} < 0\right)$, and

(xi) a higher temperature has an adverse impact on cognitive performance and hence on exam scores $\left(\frac{\partial y}{\partial a} < 0\right)$.

Using the implicit function theorem on the FOC, we get

$$\frac{\partial e^*}{\partial a} = -\frac{\frac{\partial y}{\partial a}\frac{\partial y}{\partial e^*}(\frac{\partial w}{\partial y})^2\frac{\partial^2 u_2}{\partial w^2} + \frac{\partial w}{\partial y}\frac{\partial^2 y}{\partial e^* \partial a}\frac{\partial u_2}{\partial w} + \frac{\partial y}{\partial a}\frac{\partial y}{\partial e^*}\frac{\partial^2 w}{\partial y^2}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^* \partial a}}{(\frac{\partial y}{\partial e^*})^2\left((\frac{\partial w}{\partial y})^2\frac{\partial^2 u_2}{\partial w^2} + \frac{\partial^2 w}{\partial y^2}\frac{\partial u_2}{\partial w}\right) + \frac{\partial^2 y}{\partial e^{*2}}\frac{\partial w}{\partial y}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^{*2}}}.$$

Substituting this into $\frac{dy}{da} = \frac{\partial y}{\partial e^*}\frac{\partial e^*}{\partial a} + \frac{\partial y}{\partial a}$, we obtain

$$\begin{aligned}
\frac{dy}{da} &= -\frac{\partial y}{\partial e^*}\frac{\frac{\partial y}{\partial a}\frac{\partial y}{\partial e^*}(\frac{\partial w}{\partial y})^2\frac{\partial^2 u_2}{\partial w^2} + \frac{\partial w}{\partial y}\frac{\partial^2 y}{\partial e^* \partial a}\frac{\partial u_2}{\partial w} + \frac{\partial y}{\partial a}\frac{\partial y}{\partial e^*}\frac{\partial^2 w}{\partial y^2}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^* \partial a}}{(\frac{\partial y}{\partial e^*})^2\left((\frac{\partial w}{\partial y})^2\frac{\partial^2 u_2}{\partial w^2} + \frac{\partial^2 w}{\partial y^2}\frac{\partial u_2}{\partial w}\right) + \frac{\partial^2 y}{\partial e^{*2}}\frac{\partial w}{\partial y}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^{*2}}} + \frac{\partial y}{\partial a} \\
&= \frac{-\frac{\partial y}{\partial e^*}\left(\frac{\partial w}{\partial y}\frac{\partial^2 y}{\partial e^* \partial a}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^* \partial a}\right) + \frac{\partial y}{\partial a}\left(\frac{\partial^2 y}{\partial e^{*2}}\frac{\partial w}{\partial y}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^{*2}}\right)}{(\frac{\partial y}{\partial e^*})^2\left((\frac{\partial w}{\partial y})^2\frac{\partial^2 u_2}{\partial w^2} + \frac{\partial^2 w}{\partial y^2}\frac{\partial u_2}{\partial w}\right) + \frac{\partial^2 y}{\partial e^{*2}}\frac{\partial w}{\partial y}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^{*2}}} \\
&= \frac{-\frac{\partial y}{\partial e^*}\left(\frac{\partial w}{\partial y}\frac{\partial^2 y}{\partial e^* \partial a}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^* \partial a}\right) + \frac{\partial y}{\partial a}\left(\frac{\partial^2 y}{\partial e^{*2}}\frac{\partial w}{\partial y}\frac{\partial u_2}{\partial w} + \frac{\partial^2 u_1}{\partial e^{*2}}\right)}{(\frac{\partial y}{\partial e^*})^2(\frac{\partial w}{\partial y})^2\frac{\partial^2 u_2}{\partial w^2} + \left(\frac{\partial u_2}{\partial w}\frac{\partial^2 w}{\partial y^2}\left(\frac{\partial y}{\partial e^*}\right)^2 + \frac{\partial u_2}{\partial w}\frac{\partial w}{\partial y}\frac{\partial^2 y}{\partial e^{*2}} + \frac{\partial^2 u_1}{\partial e^{*2}}\right)} < 0.
\end{aligned}$$

For the last inequality, we use the assumption that the utility function is globally concave in the effort level: $\frac{\partial u_2}{\partial w}\frac{\partial^2 w}{\partial y^2}\left(\frac{\partial y}{\partial e^*}\right)^2 + \frac{\partial u_2}{\partial w}\frac{\partial w}{\partial y}\frac{\partial^2 y}{\partial e^{*2}} + \frac{\partial^2 u_1}{\partial e^{*2}} < 0$.

B Construction of weather-related variables

We use the Princeton Meteorological Forcing Dataset to obtain weather information. This reanalysis dataset combines the climate model information and observational data from various sources, such as weather stations and satellite images. This allows us to use weather information in remote places where observational data tends to be scarce. The Princeton Meteorological Forcing Dataset is a 3-hourly dataset: weather information each day is recorded at 0 am, 3 am, . . . , and 9 pm in the Greenwich time zone. Also, the weather information is recorded on a 0.25-degree global grid. For details on the dataset, see Sheffield et al. (2006).

We use dry-bulb temperature, specific humidity, air pressure, and rainfall information in the dataset. To obtain each of these variables at each municipality where an exam is held, we use measures at four grid points surrounding municipality centroids and take their weighted average, with the inverse distance between the centroids and each of the four grid points as a weight. For weather variables other than precipitation during exams, we calculate the average across temperature measures from the latest time before the start time of exams and the earliest time after the end time of exams. For example, in Brasilia in 2016, the exam on the first day started at 1:30 pm and ended at 5:30 pm at the local time. In this case, we take the average across the temperatures at 12 pm, 3 pm, and 6 pm at the local time and use this average as a temperature measurement on a particular day. For precipitation, we use the precipitation on the "exam day" in the weather dataset. This measure is the precipitation from 9 pm on the previous day to 9 pm on the exam day, provided by the dataset.

In the analyses, we use two different measures for temperature: dry-bulb temperature and wet-bulb temperature. Dry-bulb temperature is directly obtained from the Princeton Meteorological Forcing Dataset, and the wet-bulb temperature is calculated based on dry-bulb temperature, specific humidity, and air pressure, using the following formula (Geruso and Spears, 2018):

$$\begin{aligned} T_{wb} = T_{db} * & \left[atan(0.151977 * (R + 8.313658)^{1/2}\right] + atan(T_{db} + R) \\ & - atan(R - 1.676331) + 0.00391838 R^{3/2} * atan(0.023101 R) - 4.686035 \\ R = & \ 0.263 * p * s * \left[\exp\left(\frac{17.67 T_{db}}{T_{db} + 243.5}\right)\right]^{-1}, \end{aligned}$$

where T_{wb} is wet bulb temperature (°C), T_{db} is dry-bulb temperature (°C), R is relative humidity (%), p is air pressure (Pa), and s is specific humidity.

C Illustration of how to construct the exam stakes variable

Remember that the SISU ratio is calculated based on the following formula:

$$H_{mt} = \frac{\sum_{n \in \text{all municipalities in Brazil}} w_{nm} (\#\text{ universities adopting SISU})_{nt}}{\sum_{n \in \text{all municipalities in Brazil}} w_{nm} (\#\text{ universities})_{nt}}$$

where w_{nm} are weights defined as $w_{nm} = \frac{1}{1+(\text{Distance between } n \text{ and } m \text{ (km)})}$. Here we provide a toy example to illustrate the calculation of this variable.

Suppose that there are three municipalities (A, B, and C) and four universities, one in A and C and two in B (Figure C.1). We consider a situation where one university in B adopts SISU (panel (a)). In this case, the SISU ratio for the municipality A, H_{At}, is calculated as

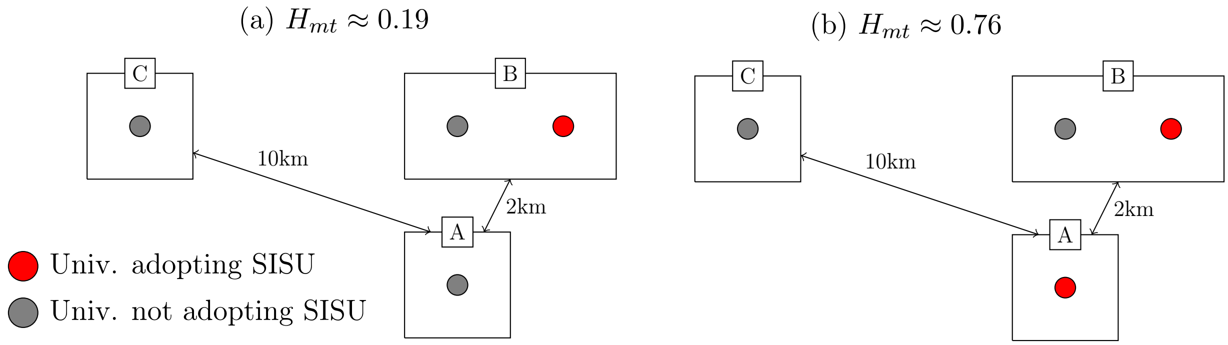
$$\begin{aligned} H_{At} &= \frac{w_{AA} \cdot 0 + w_{AB} \cdot 1 + w_{AC} \cdot 0}{w_{AA} \cdot 1 + w_{AB} \cdot 2 + w_{AC} \cdot 1} \\ &= \frac{\frac{1}{1+0} \cdot 0 + \frac{1}{1+2} \cdot 1 + \frac{1}{1+10} \cdot 0}{\frac{1}{1+0} \cdot 1 + \frac{1}{1+2} \cdot 2 + \frac{1}{1+10} \cdot 1} \\ &\approx 0.19. \end{aligned}$$

Now suppose that the university in municipality A also adopts SISU (panel B). We expect an increase in the stakes of ENEM for students in A since ENEM becomes more important for admission to the university in A. Therefore, we expect that H_{At} increases due to the SISU adoption. If we calculate H_{At} in this new situation, we obtain

$$\begin{aligned} H_{At} &= \frac{\frac{1}{1+0} \cdot 1 + \frac{1}{1+2} \cdot 1 + \frac{1}{1+10} \cdot 0}{\frac{1}{1+0} \cdot 1 + \frac{1}{1+2} \cdot 2 + \frac{1}{1+10} \cdot 1} \\ &\approx 0.76. \end{aligned}$$

This is higher than the SISU ratio in the previous situation, consistent with our expectations.

Figure C.1: Illustration of ENEM stakes variable H_{mt}



D Robustness check based on differences to Li and Patel (2021)

The first part of our paper - the direct effect of temperature on exam scores - directly relates to Li and Patel (2021) (henceforth LP). They study the direct effects of temperature on the ENEM scores. However, they find positive and statistically significant but economically negligible impacts of temperature on exam scores. This section compares our results to theirs and performs additional robustness checks by adopting some of their design decisions.

First, we follow their research design by restricting the analysis to 2012-2016, using daily average temperature, and including essay scores as one of the outcomes of interest. With this exercise, we can demonstrate where the main differences in results between our studies come from. Second, as a robustness check to our results, we will use our preferred time frame, 2010 to 2016, and discuss how adopting their research design affects our results (Table D.2).

Results for the first exercise are shown in Table D.1. Column (1) replicates our research design but restricts the data from 2012 to 2016. This estimate is the closest to our paper's main specification, only slightly smaller, showing our results are not driven by the first two of years of SISU adoption. We include the essay as a subject-score in column (2), and the estimate drops to almost a third compared to column (1). As argued in the main text, we exclude essay scores mainly because scores are not comparable due to the different nature of the essay score relative to the multiple-choice ones.

In columns (3) and (4), we keep only high school seniors in our main analysis, but we use the average daily temperature. Both columns show that using average daily temperature substantially reduces the estimates. We see statistically and economically insignificant effects using both the essay score and average daily temperature (column 4) as in D.1.

In columns (5) to (8), we restrict the sub-population of interest to ages 14 to 22, as in LP. In all columns, the estimates are considerably smaller when compared to the analysis restricted to high school seniors. One possibility for the lower average effects is that the stakes are higher for older students taking the exam for the second or more time, lowering the average temperature effects. Column (8) replicates all specifications used by LP and shows a negligible and statistically insignificant temperature effect. It is also important to note that we cannot precisely replicate their regression results. One possible difference is that we use weather information from a reanalysis dataset, which integrates data from various sources, such as weather stations and satellite observations. Instead, they use data collected from weather stations.

Now, in table D.2, we use the same time frame as in our main specification (2010-2016), but adopt the same sampling criteria as in LP and it substantially reduces our estimates. Including essay scores seem to be the most important factor in reducing the estimates, comparing columns (1) and (2). Using average daily temperature is the second most important factor (Columns (1) vs. (3)), followed by including high-school graduates (columns (1) vs. (5)). In column (8), combining all the above, estimates are negligible and statistically insignificant. For reasons explained above, we consider our main specifications the preferred ones.

Table D.1: Regression results: Comparison with results in Li and Patel (2021) (data between 2012 and 2016)

	Dependent variable: ENEM subject-score (z-score)							
Sample:	HS seniors				Ages 14-22			
Subjects:	Multiple-choice (1)	Including essay (2)	Multiple-choice (3)	Including essay (4)	Multiple-choice (5)	Including essay (6)	Multiple-choice (7)	Including essay (8)
Temp. during exam	-0.0085*** (0.0010)	-0.0031*** (0.0007)			-0.0055*** (0.0008)	-0.0005 (0.0006)		
Avg. daily temp. on exam day			-0.0036*** (0.0008)	0.0003 (0.0009)			-0.0028*** (0.0007)	0.0011 (0.0009)
Observations	24,459,856	30,574,820	24,459,856	30,574,820	73,807,204	92,259,005	73,807,204	92,259,005
R-squared	0.738	0.679	0.738	0.679	0.725	0.665	0.725	0.665
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exam date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SD of temperature var.	3.917	3.917	3.356	3.356	3.875	3.875	3.315	3.315

Notes: We use the sample between 2012 and 2016. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1

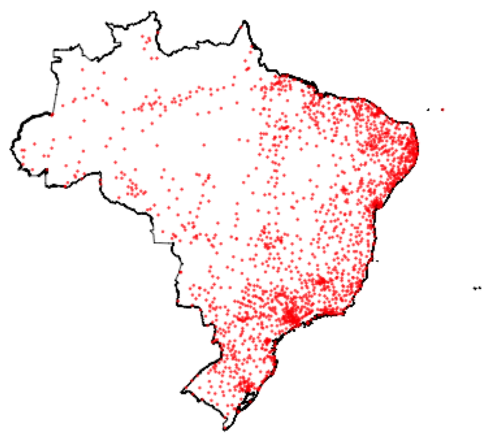
Table D.2: Regression results: Comparison with results in Li and Patel (2021) (data between 2010 and 2016)

	Dependent variable: ENEM subject-score (z-score)							
Sample:	HS seniors				Ages 14-22			
Subjects:	Multiple-choice (1)	Including essay (2)	Multiple-choice (3)	Including essay (4)	Multiple-choice (5)	Including essay (6)	Multiple-choice (7)	Including essay (8)
Temp. during exam	-0.0097*** (0.0011)	-0.0032*** (0.0007)			-0.0069*** (0.0009)	-0.0003 (0.0006)		
Avg. daily temp. on exam day			-0.0052*** (0.0009)	-0.0016* (0.0009)			-0.0039*** (0.0008)	0.0004 (0.0008)
Observations	32,392,992	40,491,240	32,392,992	40,491,240	92,976,512	116,220,640	92,976,512	116,220,640
R-squared	0.750	0.677	0.750	0.677	0.736	0.665	0.736	0.665
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exam date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SD of temperature var.	3.716	3.716	3.210	3.210	3.714	3.714	3.204	3.204

Notes: We use the sample between 2010 and 2016. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1

E Appendix figures

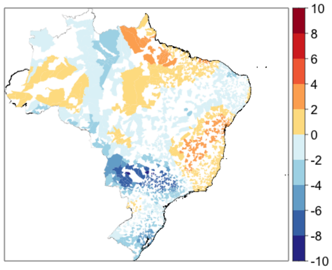
Figure E.1: Exam locations - municipality centroid



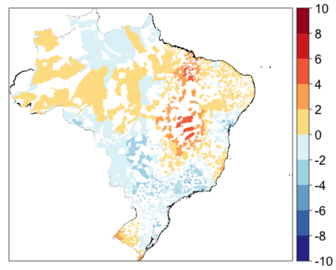
Note: The figure shows the municipalities with exam locations. The dots represent the municipality centroid. The one red dot off the continent refers to Fernando de Noronha, a district administered by the state of Pernambuco, Brazil. A municipality can have more than one exam site, but we do not have information on the exact exam location.

Figure E.2: Variation in temperature during exam from day one to day two, from 2010 and 2016

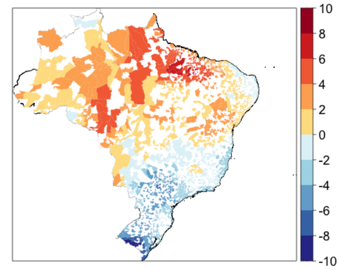
(a) 2010



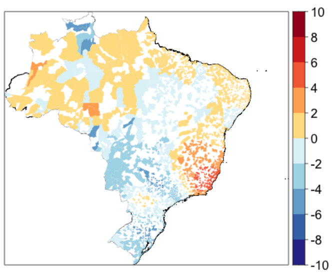
(b) 2011



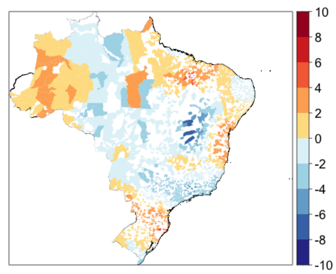
(c) 2012



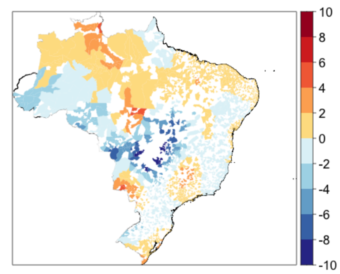
(d) 2013



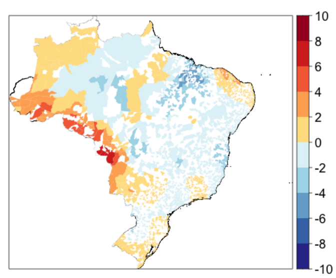
(e) 2014



(f) 2015

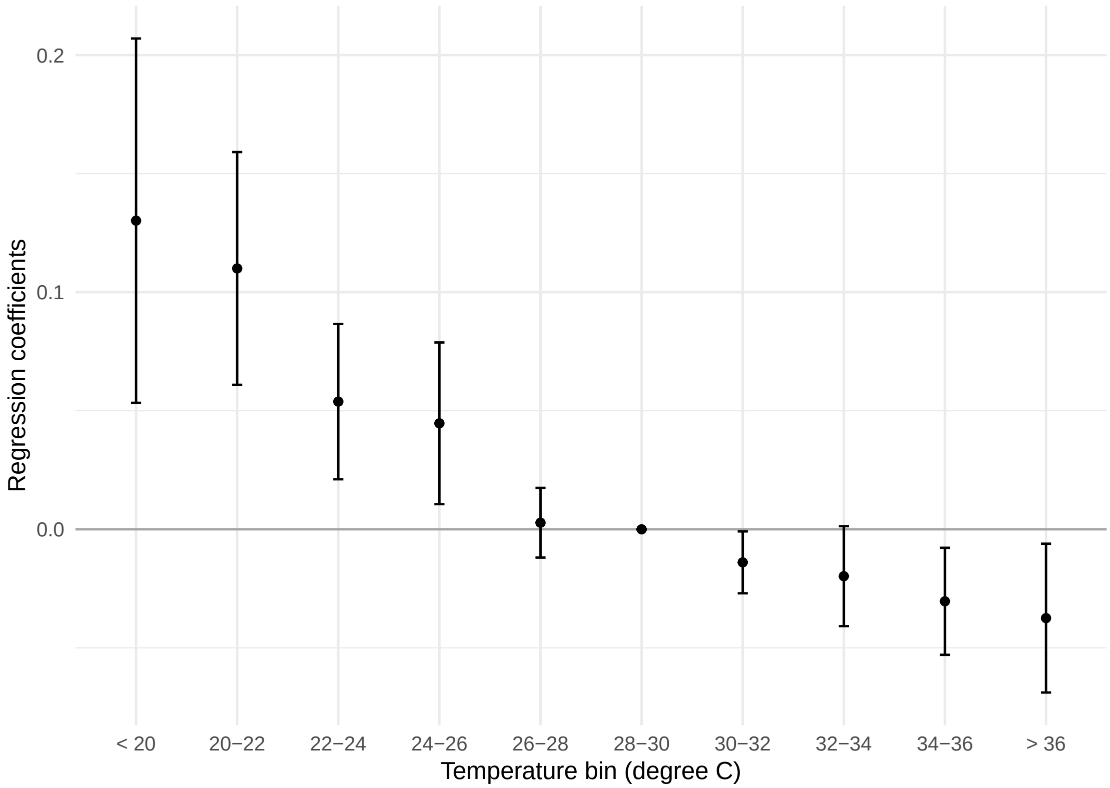


(g) 2016



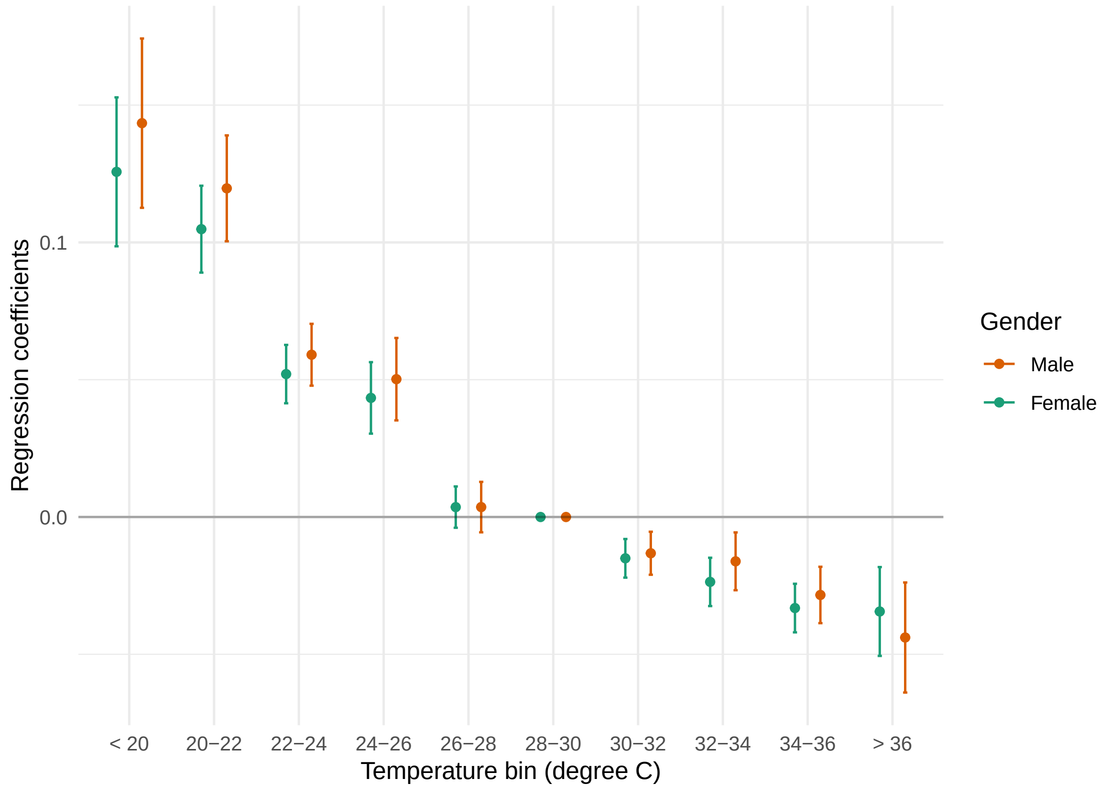
Notes: The figure shows the municipality level variation in temperature from day 1 to day 2 (difference = day 2 − day 1) from 2010 to 2016. In the municipalities with white color, nobody in our sample took ENEM in the year.

Figure E.3: Regression results: temperature and exam Z-scores (with Conley standard errors)



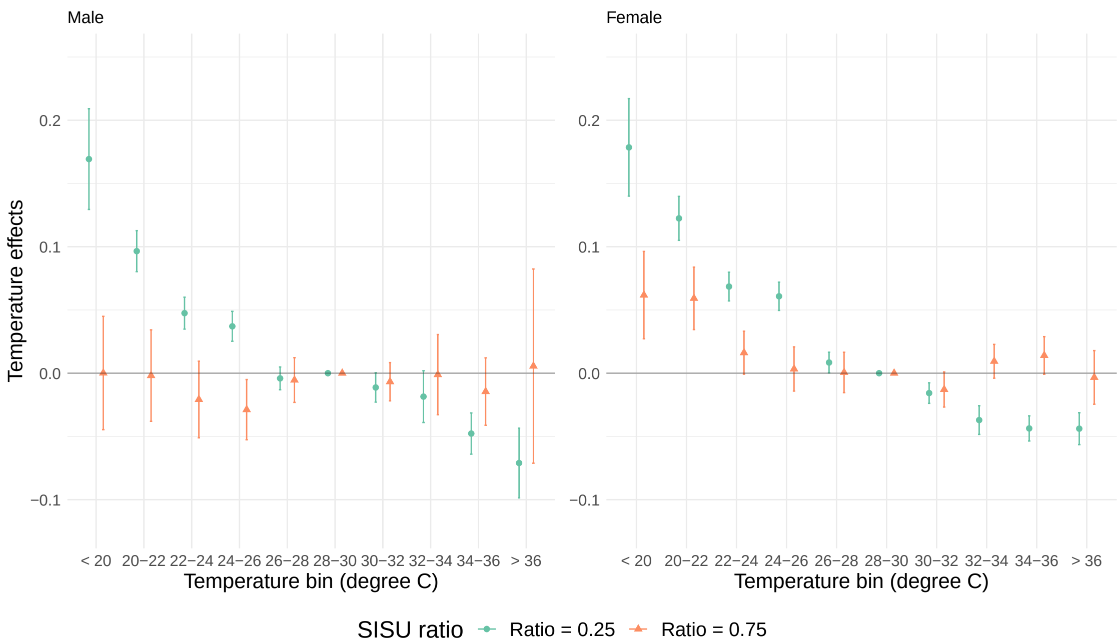
Note: The figure shows estimates of the effects of temperature on Z-scores using a flexible temperature functional form. The unit of observation is subject-student-year. The dependent variable is the Z-scores of exams in each subject-year. Error bars indicate 95% confidence intervals. Precipitation on the exam days, exam-date fixed effects, subject fixed effects, and individual fixed effects are included in the regression. Standard errors are computed based on Conley (1999) with 200km cutoffs.

Figure E.4: Regression results: temperature and exam Z-scores, by male and female



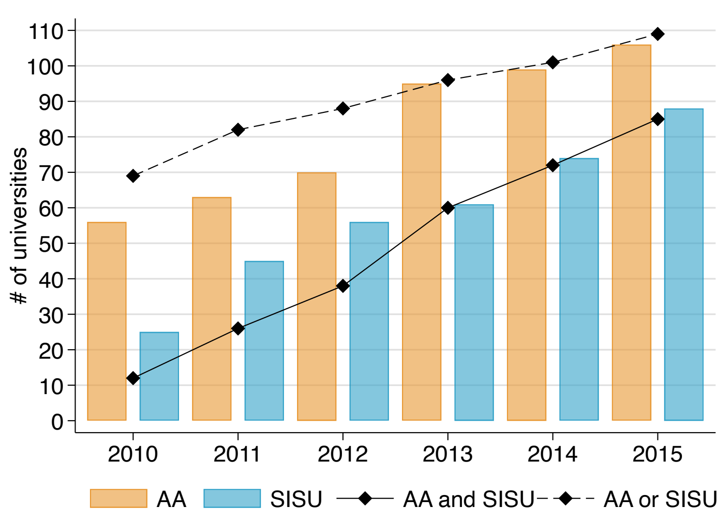
Note: The figure shows estimates of the effects of temperature on Z-scores using a flexible temperature functional form, estimated separately by gender. The unit of observation is subject-student-year. The dependent variable is the Z-scores of exams in each subject-year. Error bars indicate 95% confidence intervals. Precipitation on the exam days, exam-date fixed effects, subject fixed effects, and individual fixed effects are included in the regression. Standard errors are clustered at the municipality level.

Figure E.5: Heterogeneity results: regression coefficients of temperature × SISU adoption ratio, by gender



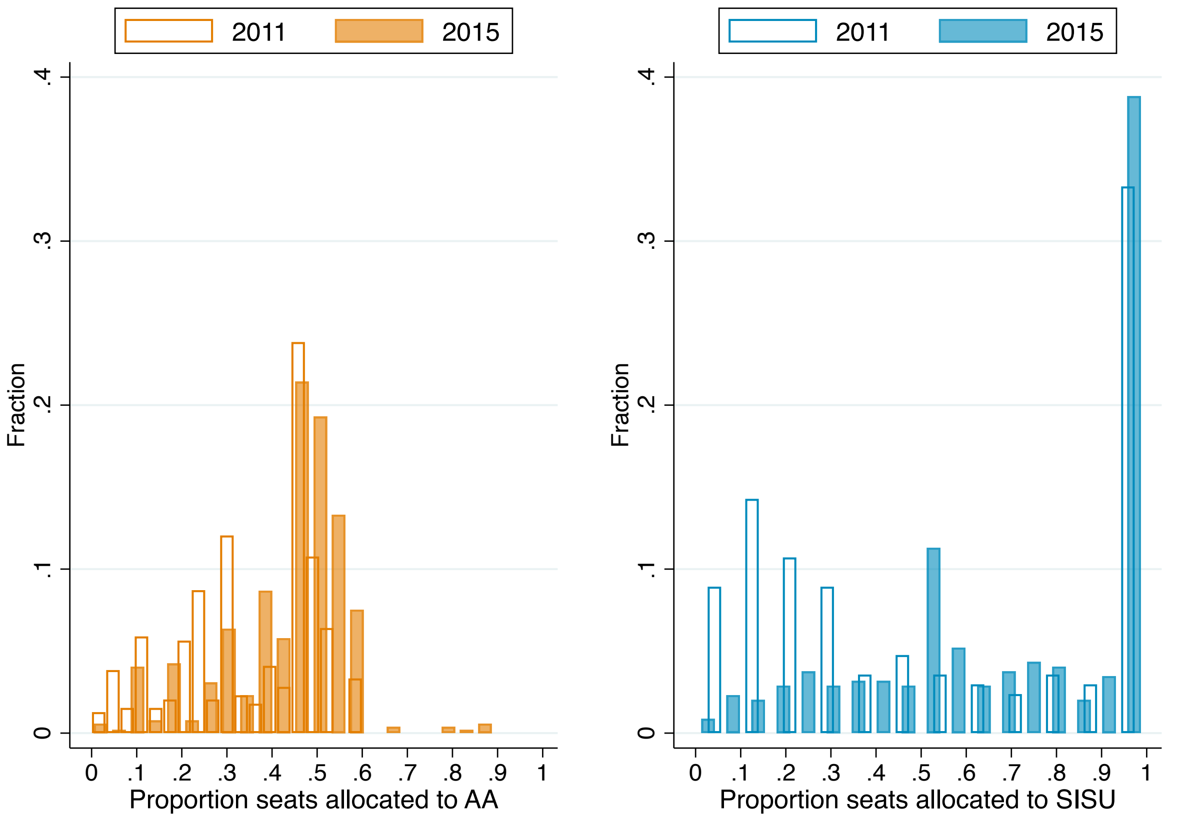
Notes: Effects are calculated based on point estimates from regressions of Z-scores on temperature and its interaction with SISU adoption ratio. Error bars indicate 95% confidence intervals. Municipalities whose distance from the municipality where a student took ENEM is less than 60km are used, and the inverse of the distance between municipalities is used as weights. Precipitation is controlled for in all regressions. Standard errors are clustered at the municipality level.

Figure E.6: Number of federal and state universities adopting AA or SISU



Notes: This figures shows the number of universities adoption AA, SISU, either or both policies.

Figure E.7: Distribution of the share of total seats allocated to AA or SISU by municipality



Notes: This figures shows the distribution of percent of seats allocated to AA or SISU, conditional on the municipality having at least one university adopting the policy.

F Appendix tables

Table F.1: Total number of exam takers in the targeted sampled by year

	# sub-population	Total applicants
2010	914,725	4,626,094
2011	1,068,559	5,380,857
2012	1,096,451	5,791,332
2013	1,178,558	7,173,574
2014	1,264,333	8,722,290
2015	1,247,157	7,792,025
2016	1,328,465	8,627,371

Notes: targeted population is composed by high-school seniors, who are 16-20 years old, that were present in all exams and not disqualified due to, for example, cheating.

Table F.2: Regression results: Linear function of temperature, using ENEM Z-score (with Conley standard errors)

	Dependent variable: ENEM subject-score (z-score)		
	(1)	(2)	(3)
Temperature during exam	-0.0097***	-0.0097***	
	(0.0032)	(0.0028)	
Wet-bulb temperature during exam			-0.0115***
			(0.0036)
Precipitation (m/day) on exam day		0.0066	
		(0.0935)	
Exam date FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes
Observations	32,392,992	32,392,992	32,392,992

Note: This table presents estimates for the linear effects of temperature on exam scores. The unit of observation is subject-student-year. The dependent variable is the Z-scores of exams in each subject-year. WB stands for wet-bulb temperature. Standard errors are computed based on Conley (1999) with 200km cutoffs.

Table F.3: Summary statistics on majors and cut-offs, by university, campus, and year

	Mean	Median	Min.	Max.
# Majors	16.04	9	5	91
# Majors below the mean cut-off	8.95	5	1	55
Mean cut-off	662.15	659.66	541.64	825.88
# of college-campus		1,196		
# of majors		6,272		

Table F.4: Regression results: Interaction effects of temperature and ENEM stakes and other potentially confounding factors on exam z-score

	Dependent variable: ENEM subject-score (z-score)				
	(1)	(2)	(3)	(4)	(5)
Temp. × SISU ratio	0.0254***	0.0246***	0.0252***	0.0186***	0.0180***
	(0.00474)	(0.00477)	(0.00514)	(0.00480)	(0.00478)
Temp. × High Inc		-0.00605***			-0.00578***
		(0.000684)			(0.000532)
Temp. × State HS		-0.0106***			-0.0105***
		(0.00149)			(0.00150)
Temp. × Mun HS		-0.0195***			-0.0180***
		(0.00340)			(0.00314)
Temp. × Pri HS		-0.00262*			-0.000852
		(0.00145)			(0.00133)
Temp. × Gini			-0.00111		-0.00190
			(0.00183)		(0.00183)
Temp. × Percent Poor			-0.00118		-8.13e-05
			(0.00177)		(0.00172)
Temp. × Education Index			-0.00339**		-0.00279**
			(0.00144)		(0.00135)
Temp. × 2011				-0.00172	-0.00203
				(0.00199)	(0.00201)
Temp. × 2012				-0.00516***	-0.00553***
				(0.00187)	(0.00197)
Temp. × 2013				0.00329***	0.00370***
				(0.00125)	(0.00114)
Temp. × 2014				-0.00277	-0.00266
				(0.00210)	(0.00212)
Temp. × 2015				0.0135***	0.0141***
				(0.00243)	(0.00274)
Temp. × 2016				0.00710*	0.00634
				(0.00390)	(0.00407)
Observations	32,392,992	32,392,960	32,392,992	32,392,992	32,392,960
Subject FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Exam date FE	Yes	Yes	Yes	Yes	Yes
SD of SISU ratio	0.220	0.220	0.220	0.220	0.220
Average temperature effect	-0.00968	-0.00968	-0.00968	-0.00968	-0.00968

Notes: Individual-level variables interacted with temperature are an indicator of above-median-income household and the type of students' high school (federal, private, state, or municipal). Municipality-level variables interacted with temperature are the Gini coefficient, the poverty rate, and the education index, which are standardized (mean 0 and sd 1). Precipitation on exam days is controlled for in all regressions. Average temperature effect is the estimated coefficient $\widehat{\alpha}$ from a regression equation, $Y_{imsdt} = \alpha T_{mdt} + X'_{mdt}\beta + \mu_i + \eta_s + \tau_{dt} + \varepsilon_{imsdt}$. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1

Table F.5: Regression results: Interaction effects of temperature and ENEM stakes on exam Z-score (with Conley standard errors)

	Dependent variable: ENEM subject-score (z-score)				
	(1)	(2)	(3)	(4)	(5)
Temp. × SISU ratio	0.0254***	0.0246***	0.0251***	0.0186**	0.0180***
	(0.0083)	(0.0080)	(0.0074)	(0.0082)	(0.0064)
Observations	32,392,992	32,392,960	32,392,992	32,392,992	32,392,960
Ind.; Var.; Interactions	No	Yes	No	No	Yes
Mun.; Var.; Interactions	No	No	Yes	No	Yes
Year Interactions	No	No	No	Yes	Yes
Exam date FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes

Notes: Average temperature effect is the estimated coefficient $\widehat{\alpha}$ from a regression equation, $Y_{imsdt} = \alpha T_{mdt} + X'_{mdt}\beta + \mu_i + \eta_s + \tau_{dt} + \varepsilon_{imsdt}$. Precipitation on exam days is controlled for in all regressions. Standard errors are computed based on Conley (1999) with 200km cutoffs. *** p<0.01, ** p<0.05, * p<0.1.

Table F.6: Robustness check: the joint effects of SISU and AA (2010-2014)

	Dependent variable: ENEM subject-score (z-score)			
	SISU Effect	AA Effect	Both	Both w/ controls
	(1)	(2)	(3)	(4)
Temp. × SISU ratio (intensive)	0.0244***		0.0128***	0.0125***
	(0.0056)		(0.0045)	(0.0040)
Temp. × AA ratio (intensive)		0.0541***	0.0433***	0.0485***
		(0.0103)	(0.0099)	(0.0079)
Observations	22,090,472	22,090,472	22,090,472	22,090,472
Subject, Individual, Exam data FE	Yes	Yes	Yes	Yes
Ind.; Mun.; Year Interactions	No	No	No	Yes
SD of SISU ratio	0.2220	0.1270	.	.
Average temperature effect	-0.0131	-0.0131	-0.0131	-0.0131

Notes: This table show results using an alternative measure of SISU and AA adoption: proportion of seats allocated to each policy. Data on AA and SISU adoption is from Mello (2022) and is restricted to years 2010-2014 (SISU 2011-2015). Average temperature effect is the estimated coefficient $\widehat{\alpha}$ from a regression equation, $Y_{imsdt} = \alpha T_{mdt} + X'_{mdt}\beta + \mu_i + \eta_s + \tau_{dt} + \varepsilon_{imsdt}$. Precipitation on exam days is controlled for in all regressions. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1.

Table F.7: Robustness checks by varying the municipalities is included to calculate SISU ratios

	Dependent variable: ENEM subject-score (z-score)				
	(1)	(2)	(3)	(4)	(5)
Temp. × SISU ratio (weighted, all)	0.0254***				
	(0.00474)				
Temp. × SISU ratio (unweighted, 60km)		0.0109***			
		(0.00163)			
Temp. × SISU ratio (weighted, 60km)			0.00884***		
			(0.00180)		
Temp. × SISU ratio (unweighted, CZ)				0.00747***	
				(0.00127)	
Temp. × SISU ratio (weighted, CZ)					0.00761***
					(0.00164)
Observations	32,392,992	32,392,992	32,392,992	32,392,992	32,392,992
Subject FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Exam date FE	Yes	Yes	Yes	Yes	Yes
SD of SISU ratio	0.220	0.349	0.381	0.378	0.399
Average temperature effect	-0.00968	-0.00968	-0.00968	-0.00968	-0.00968

Notes: The variable "SISU ratio" is the proportion of universities adopting SISU, with different municipalities used for calculations. For column (1), all municipalities in Brazil are used. For columns (2) and (3), municipalities whose distance from the municipality where a student took ENEM is less than 60km are used. For columns (4) and (5), municipalities that belong to the same commuting zone as the municipality where a student took ENEM are used. For columns (2) and (4), the weights were not used, and for columns (1), (3), and (5), the inverse of the distance (km) between municipalities is used as weights. Precipitation on exam days is controlled for in all regressions. Average temperature effect is the estimated coefficient $\widehat{\alpha}$ from a regression equation, $Y_{imsdt} = \alpha T_{mdt} + X'_{mdt}\beta + \mu_i + \eta_s + \tau_{dt} + \varepsilon_{imsdt}$. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1

Table F.8: Robustness checks by different distance units used to calculate SISU ratios

Dependent variable: ENEM subject-score (z-score)	(1)	(2)	(3)
Temp. × SISU ratio (all, km)	0.0254***		
	(0.00474)		
Temp. × SISU ratio (all, mile)		0.0316***	
		(0.00558)	
Temp. × SISU ratio (all, meter)			0.0117***
			(0.00253)
Observations	32392992	32392992	32392992
R-squared	0.750	0.750	0.750
Subject FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Exam date FE	Yes	Yes	Yes
SD of SISU ratio	0.220	0.196	0.337
Average temperature effect	-0.00968	-0.00968	-0.00968

Notes: The variable "SISU ratio" is the proportion of universities adopting SISU, with different weights used for calculations. For column (1), kilometer distances are used as weights. For column (2), mile distances are used as weights. For column (3), meter distances are used as weights. Precipitation on exam days is controlled for in all regressions. Average temperature effect is the estimated coefficient $\widehat{\alpha}$ from a regression equation, $Y_{imsdt} = \alpha T_{mdt} + X'_{mdt}\beta + \mu_i + \eta_s + \tau_{dt} + \varepsilon_{imsdt}$. Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.1