Too Hot to Get Tired: How Fatigue Mitigates Temperature Effects in Long Cognitive Tasks

Tales Rands*

Abstract

Recent literature has documented that economic agents under-perform when exposed to heat plausibly due to a difference in effort. Yet, it has also have been recently documented that mental fatigue is an important detractor of performance, which raises one natural question: could a difference in mental fatigue induced by heat mitigate the negative effect of heat itself over hours-long cognitive tasks? This paper studies this possibility both empirically, leveraging a national exam in Brazil used as selection criteria for most universities in the country, and theoretically, by building on recent models of temperature effects over performance by adding to it the dimension of time and the element of mental fatigue. Together these analyses suggest that a temperature shocks may induce a difference in initial effort that causes a difference in mental fatigue later that mitigates the direct effect of the temperature shock.

JEL Classification Codes: I21, Q54, J24
Keywords: temperature, effort, mental fatigue, achievement

*EPGE Brazilian School of Economics and Finance. E-mail: talesrands@gmail.com

1 Introduction

Thermal discomfort have negative consequences to humans and their cognitive capabilities. Recent literature have shown these effects both on the long-run, jeopardizing learning (Park et al. (2020), Behrer (2021), Roach and Whitney (2022)), and on the short-run, declining agents' abilities to perform cognitive tasks (Melo (2023), Park (2022), Graff Zivin, Hsiang, and Neidell (2018), Vu (2019)). Furthermore, in Melo (2023) this effect is shown to decline in higher stakes situations, reinforcing the importance of effort as a mediator of the temperature effect. This difference in effort caused by differences in temperature can generate further consequences, and the present work takes a deeper look into how these other consequences affect cognitive performance throughout time on hours-long tasks.

Introducing the dimension of time to this problem demands the introduction of fatigue, understood here as a direct consequence of recent exerted effort and as an element that induces a decline in performance. The importance of fatigue to cognitive performance is also documented in recent literature (Balart, Oosterveen, and Webbink (2018), Reyes (2023), Brown et al. (2022)), and in the present work it is accounted as the driving force of the indirect consequences of temperature shocks. In the theoretical model, a temperature shock causes a difference in initial effort, inducing a difference in fatigue that later affects performance in the opposite direction, mitigating the effect of temperature in time-extensive tasks. This prediction is then tested using a large-scale national high-stakes exam in Brazil, and results suggest that at the last question of the exam the total effect of a temperature shock could be completely nullified and even overturn by this indirect mechanism.

The quasi-experiment leveraged in this work is Brazil's ENEM, a national exam that plays some role in selecting students to most major public universities in the country - generally being the only criteria. Conveniently, this set-up was previously used to study effects from both temperature (Melo (2023)) and fatigue (Reyes (2023)), therefore both effects have been documented in this very context. Moreover, two features of this quasi-experiment make it specially adequate to study both phenomena: (i) student's are induced to answer questions in a given order, and this order is randomized between students; and (ii) the exam is divided in two days, therefore it is possible to observe the same student at two different temperatures. These are necessary to isolate students' and questions' fixed effects, as well as the effect of mental fatigue through positions' fixed effects; then one can finally estimate the temperature effect and its evolution throughout the exam days.

The main contribution of this work, therefore, is to highlight how the well-documented effect of thermal discomfort over cognitive performance evolves in time. This is a contribution to recent literature investigating external determinants of performance, more specifically to its understanding of the aforementioned fatigue and temperature effects. While it seems no other work has looked specifically at the evolution of the temperature effect over time, there are some studies regarding the effect of temperature on mental fatigue - although outside the economic literature. In order to understand how our findings fit within the existing literature on the temperature effect on performance, it is important to first understand what has been established about this effect and its interaction with time.

Graff Zivin, Hsiang, and Neidell (2018) provide the first estimate of short-term performance decline caused by heat, followed by Vu (2019) and Park (2022). The later goes further and models the impact of thermal discomfort acting through two different channels: a direct impact on performance, and a indirect impact on the disutility of exerting effort. This is specially important because Melo (2023) then shows that result stakes are an important mediator of the temperature effect, being able to mitigate as much as 80% of the total effect - highlighting the importance of effort in this mechanism, which is essential for our hypothesis. None of these works, however, adds the dimension of time to this problem.

As for the effect of mental fatigue on performance, it appears as a measurement of non-cognitive skills in Borghans and Schils (2018), Reyes (2023) and Brown et al. (2022) - always focusing on individuals resistances to fatigue and its consequences rather than on the causes of the fatigue itself. Outside the realm of economic literature, within health literature mental fatigue is *a psychobiological state caused by prolonged periods of demanding cognitive activity and is characterized by various subjective, physiological, and behavioral alterations* (Van Cutsem (2017)), which is a broader definition than necessary for this work. In these broader terms Qian et al. (2015) and Solivio (2021) present evidence of heat increasing mental fatigue. The first finds significant effects on biological measurements and on a reflex-based performance test, but insignificant effects on self-reported mental fatigue; and the second finds effects of temperature on self-reported ”irritability”, ”unproductivity” and ”concentration difficulty”, which the authors associate to mental fatigue, yet there is no significant effect on self-reported ”mental fatigue”.

Yet, the more elucidating results for how our work dialogues with this literature might lie a little further from our question. Moore (2012) study the effects of exercise-induced fatigue on cognitive performance, and find that different cognitive tasks are affected differently; additionally, Smith (2019) highlights the different characteristics of the mental fatigue induced by different cognitive tasks. These results showcase that the mechanisms of mental fatigue and its effect over cognitive performance are nuanced and often ambiguous. In the present work mental fatigue is accumulated through a repetitive process[1] and measured as the decline in performance on the very same task generating the mental fatigue. The interaction of a variable defined in that manner with temperature have not been previously studied, thus the novelty of our results.

In section 3 we present a 2-periods model that accounts for effort choices, fatigue accumulation and temperature effects, showcasing how the process of fatigue accumulation can create a mechanism to mitigate the total effect of temperature shocks over performance. Then in section 2 we explain the empirical strategy to test the intuitions from the theoretical model and the context of the quasi-experiment leveraged to these tests. Section 4.2 presents the main results and tests for other possible mechanisms. Section 5 concludes and discusses what these results add to the literature.

2 Theoretical Framework

The theoretical trigger for the intuitions in this work is Melo (2023)'s result suggesting that effort is the main channel through of thermal discomfort's effects on performance. Her result is presented under the light of Park (2022)'s model, with agents benefiting from performing better ($y(a)$) due to higher effort (a) but balancing this to the physical discomfort of exerting effort. This cost of exerting effort is higher under higher temperature, causing agents choose to exert less effort. Our goal is to expand this context by adding to it the dimension of time, allowing for this resulting difference in effort to indirectly affect performance through a second mechanism: mental fatigue.

In our version of the model, agents perform two consecutive tasks and derive utility from the sum of their performances $Y(a_1, a_2) = y(a_1) + y(a_2)$. Exerting effort causes a disutility that depends on current environmental conditions (γ) and mental fatigue (f). Therefore the agents' utility maximization problem is:

$$\max_{a_1,a_2} \quad U = u(y(a_{i1}) + y(a_{i2})) - d(a_{i1}, \gamma, f(0)) - d(a_{i2}, \gamma, f(a_{i1})) \tag{1}$$

For the sake of simplicity we assume utility to be linear in performances and physical discomfort to be

[1] i.e., answering questions from the ENEM exam

linear in present effort. The positive impact of effort in performance is decreasing in itself and mediated by individual ability α, and the physical discomfort caused by present effort is aggravated by thermal discomfort and mental fatigue. We model thermal discomfort as e^{γ}, because (1) exerting effort still generates discomfort in ideal environmental conditions and (2) a 20ºC temperature offset should be clearly more than twice as harmful as a 10ºC temperature offset. As for mental fatigue, we assume it to be an increasing function of a_{t-1}, $f(a_{t-1}) = a_{t-1}^{\rho}$, mediated by the fatigue-accumulation factor ρ. Additionally, effort should have a greater impact on present physical discomfort than later physical discomfort, therefore $0 < \rho < 1$. Finally, the utility maximization problem is:

$$\max_{a_1,a_2} \quad U = \alpha(\ln a_1 + \ln a_2) - a_1(e^{\gamma} + 0^{\rho}) - a_2(e^{\gamma} + a_1^{\rho}) \tag{2}$$

It is worthy noting that effort in this model is the only channel for temperature or fatigue to affect performance, which is a simplification that helps the interpretation of the necessary intuitions without polluting the results with too many mechanisms. This simplification is supported by Melo (2023), that finds that up to 80% of the temperature effect can be mitigated with differences in incentives - and are therefore mediated by effort. From this point, our goal is to understand how a temperature shock affects effort and performance in $t = 1$ and $t = 2$, i.e., to calculate $\frac{da_1}{d\gamma}$ and $\frac{da_2}{d\gamma}$.

The model's First Order Conditions can be rearranged as:

$$\frac{\alpha}{a_1} - \rho a_2 a_1^{(\rho-1)} = e^{\gamma} \tag{3}$$

$$\frac{\alpha}{a_2} - a_1^{\rho} = e^{\gamma} \tag{4}$$

Equation (4) evidentiates that in this model higher levels of environmental distress (γ), on the right-hand side of both equations, induce some combination of lower effort levels (a_1 and a_2) on the left-hand side. From that alone it is not yet possible to state that either a_1 or a_2 necessarily decreases if γ increases, however it implies that at least one of a_1 and a_2 have to decrease.

In order to understand how the chosen levels of effort a_t (and thus the performance) reacts to variations in thermal-discomfort γ we want to calculate $\frac{da_1}{d\gamma}$ and $\frac{da_2}{d\gamma}$. The values of a_1 and a_2 are co-determined implicitly in (3) and (4) for a given value of γ. Therefore the derivatives of a_1 and a_2 with respect to γ can be calculated using the Implicit Function Theorem, as is shown in Appendix A, and it results in:

$$\frac{da_1}{d\gamma} = \Delta(-\rho a_1^{(\rho-1)} - \frac{\alpha}{a_2^2}) \tag{5}$$

$$\frac{da_2}{da} = \Delta(-\rho a_1^{(\rho-1)} - \frac{\alpha}{a_1^2} - \rho(\rho-1)a_2 a_1^{\rho-2}). \tag{6}$$

Where Δ is a common positive multiplier.[2] Equations (5) and (6) were rearranged to highlight the relevant contrast between $\frac{da_1}{d\gamma}$ and $\frac{da_2}{d\gamma}$: the effect is stronger (more negative) in the first period than in the later period, represented by the presence of an additional positive term.

3 Data and Background Information

This section intends to provide all necessary information to understand and interpret our empirical approach and its results. First sub-section 3.1. presents a general description of the exam leveraged in our quasi-experiment, its recent history, what is at stakes for exam-takers and details of its application. Then sub-section 3.2 explains how we measure thermal discomfort on exam-takers, the necessary datasets and the available information. Sub-section 3.3 follows by describing the micro-data from the exam and the selected universe of interest. Finally, sub-section 3.4 explains how we drew samples from the original data to use in our empirical analysis.

3.1 Background Context

The quasi-experiment leveraged in this work is the set of main applications of the Brazilian ENEM (*Exame Nacional do Ensino Médio*) from 2013 to 2016. The ENEM was created in 1998 as tool for evaluating public schools, but in 2004 it received a role in distributing public scholarships to private universities. Finally in 2009 the Brazilian Ministry of Education began to invest in ENEM to be the national unified selection process to public universities in the country, providing financial incentives for public universities that voluntarily adopted ENEM as their selection process. In spite of that, ENEM adoption was not instantaneous

[2] From (4) we know that at least one of $\frac{da_1}{d\gamma}$, $\frac{da_2}{d\gamma}$ must be negative. The term in parenthesis in (5) is necessarily negative, thus Δ being negative can only possibly be negative if the term in parenthesis in (6) is positive, so that the derivatives have opposite signals and there is still one of the two variables reacting negatively to γ. However, this term can only be positive if a_2 is considerably higher than a_1 and α is very low, which is not a reasonable equilibrium for any values of ρ and α.

and it encountered some resistance. Indeed, it is only four years later, in 2013, that the official website of ENEM celebrates the fact that *"for the first time, almost all federal institutions adopted ENEM as selection criteria"*.

The exam consists of 180 questions, divided in 4 major subjects (Math, Portuguese, Human Sciences and Natural Sciences) and a writing exam. The exam is applied in two consecutive days (Saturday and Sunday)[3]. The first day is dedicated to natural and human sciences (90 questions) and students have 4.5 hours to complete it. The second day is dedicated to Math, Portuguese (90 questions) and a 20-to-30 lines writing exam, and students have 5.5 hours. As the ENEM gained importance nation-wide a lot of emphasis was put into the fact that it is too long even considering students have 4.5 hours. This also reflects the style of the test, with long questions that require a lot of reading. Appendix C presents one page from each subject and each year of our quasi-experiment to illustrate this point.

The exam application begins at 12:30pm in over 1000 municipalities, and students are randomly sorted to different facilities in their selected municipalities. Most of these facilities are public schools and their infrastructures might vary deeply, from the type of chairs to the presence of air-conditioning - most public schools in Brazil lack air-conditioning. In fact, Brazil's 2022 School Census documented that nearly 70% of all public classrooms in the country lack air-conditioning. This is specially relevant because the exam takes place between the end of October and the beginning of November, when temperatures average 27,76ºC across all municipalities that hosts ENEM exams during the time of the tests. In fact, more than 35% of the test application sites register over 32ºC during a daily application.

One convenient procedure of the ENEM exam is that the gates to every test site closes half an hour before the beginning of the exam. This time allow for students to be at thermal equilibrium with the temperature conditions of the room in which they will take the test from the beginning of the exam.

Another convenient feature of the ENEM application is its anti-cheating procedure. The same 45-questions composing each subject are distributed in 4 different booklets with questions appearing in different orders, and these booklets are randomly distributed to test-takers student's questions are not ordered the same as the questions for other students sitting near them. Therefore every question is answered by every student, however in 4 different positions, randomly sorted among students. The order of subjects is fixed to all students and through the whole time period of interest (2013-2016). Each booklet has all the questions for the

[3]In 2018 it changed to two consecutive Sundays, but this is not in our sample

day, therefore this randomization applies to the order in which questions present themselves to the students, but ultimately they are free to choose the order in which they want to answer the questions or subjects of the day.

3.2 Temperature

For weather data we use the Princeton Global Meteorological Forcing Dataset for land surface modeling.[4] This dataset estimates 3-hourly temperature information, along with other information such as humidity and daily rainfall, on a regular grid over the whole globe. There are measurements for 12h, 15h and 18h[5] (local time), and the main temperature measure used in most regressions (referred as *mean temperature*) is the average of the temperature measured at these three times. The openly available measurements currently ends in 2016, which is the reason why the analysed period ends in 2016.

To measure the temperature at a given point x, we take the average of the four points closest to x and weight it inversely proportional to its squared distance to x. Because we only know the municipality in which students took the exam, and not the exact site, we estimate the temperatures at the municipality's centroid - obtained in a openly available dataset from IBGE, the *Instituto Brasileiro de Geografia e Estatística* or the Brazilian Institute of Geography and Statistics.

It is worthy noting that higher temperatures are associated with more thermal discomfort in the context of this work because the exam occurs in Brazil's spring. On average over half of the exam sites registers temperatures above 30ºC during the exam, while less than 10% registers temperatures under 20ºC. The effect measured in the empirical section 4 of this work is the effect of a 1ºC increase in temperature in a context in which this is most certainly a discomfort.

One important part of our argument is that the temperature variation during the exam-days are as good as randomly assigned to municipalities, since the estimation of the temperature effects com from observing the same students vary performance under different temperatures in the two (consecutive) exam days. Figures 1 and 2 at Figures section 6 show how temperature has varied between exam days for each year in the sample. The main threat to our identification that could come from patterns of temperature variations would be if they were connect with some sort of regional subject-specific bias, but the images do not suggest any kind

[4]Details of the dataset are provided in Sheffield et al. (2006)

[5]and so on every three hours. These are the measurements of interests for encompassing the whole exam

of pattern for the temperature variations between exam days.

3.3 Exam Data

The openly available data from ENEM provides information on every answer provided by every students, allowing also for the identification of the order in which questions were presented. It also informs the municipality in which exams' were taken. These are the minimum data necessary to operate the identification strategy described in sub-section 4.1.

We also use ENEM's micro-level data to restrict the population of interest, with the goal to level stakes for all subjects of our quasi-experiment. Therefore we defined our universe of interest to students aged 17 to 19 finishing high-school who were present for all four test (Math, Portuguese, Human Sciences and Natural Sciences) at the years of 2013, 2014, 2015 and 2016. The decision to being our data in 2013 also intents to stabilize the stakes of the exam, considering it is the year that ENEM's official website marks as the one in which "for the first time, almost all federal institutions adopted ENEM as selection criteria".

This ENEM micro-level data is also used in Table 1 to summarise who the profile of the universe of interest as defined here. As one can see, there is a majority of females (58.4%) aged 17 from public schools (76.9%) that can barely outperform completely random answers[6] in both Math (25.8%) and Natural Sciences (26.3 %).

3.4 Sampling

The size of the total universe of interest, even after filtering the exam-takers as described in the last sub-section, and the nature of our empirical strategy force us to select a sample from the total universe of almost 5 million students within our criteria. Due to computational reasons, we aim at working with a sample of around 110.000 students. Since the goal of these estimations is to assess how the probability of a question being correctly answered varies according to its position on test and to municipality's mean temperature during the exam, the process of selecting a sub-sample should (1) optimize the estimation of students' fixed effects, to better isolate the desired effects, and (2) maximize the number of municipalities, which is the root of variations in temperature.

[6]Every question is multiple choice and has 5 options

In order to respect these priorities, the construction of the working sample focus on (1) keeping all answers given by every student in the sub-sample and (2) keeping all municipalities present in the original sample. In practice, the municipalities are divided into three groups according to their sizes on the original sample: big municipalities, the top 1% in number of students during the period of interest; medium-sized municipalities, the ones in percentiles 2 to 5; and the small municipalities defined as the bottom 95%. From each municipality we sort 56, 86, 569 students[7] (from small, medium or big municipalities, respectively), when available, from any year of the period of interest (2013-2016) without distinction. From each student (near 109.741) we take all 180 answers as observation units. Observations from a student from municipality m are then weighted by the ratio of total students from m in the universe to total students from m in the sample.

Table 1 summarises this sample, both weighted and unweighted, and compares its composition to the one of the complete universe of available observations.

4 Empirical Results

This section intents to analyse results from the quasi-experiment described on sub-section 3.1 and discuss how they dialogue with the theoretical model presented in section 2. Sub-section 4.1 explains the modeling process and the assumptions necessary to our interpretation of the quasi-experiment, which leads to the statistical tests presented in sub-section 4.2 along with their results. Finally, sub-section 4.3 addresses the main identification threat to the results.

4.1 Empirical Approach

The first thing to understand about our identification strategy is that the unit of observation is one question being answered by one student at a given day and a given municipality, and therefore at a given temperature. Since the full exam has 180 questions, there are 180 observations for each student taking the exam. More precisely, our dependent variable is a dummy for whether the answer is correct or not. Here we are trying to model the probability of a student correctly answering a given question presented at given position - using

[7] This allocation was chosen so that small municipalities had 90% of the average number of students per municipality, and big municipalities had 10x more students than small municipalities - medium municipalities got what was left. Than we took the integer part of the resulting number of students for each municipality.

Table 1: Data and Sample

	Universe	**Unweighted Sample**	**Weighted Sample**
Females	58.4%	59.5%	58.7%
Federal Schools	2.3%	2.2%	2.1%
Other Public Schools	74.6%	84.7%	78.1%
Private Schools	23.1%	13.0%	18.6%
White	44.4%	38.2%	45.2%
Aged 17	57.8%	56.7%	59.3%
Aged 18	32.9%	32.3%	31.8%
Aged 19	9.3%	11.0%	8.9%
Mean Temperature	26.7ºC	27.8ºC	26.8ºC
E(Y | subject = Reading)	0.384	0.357	0.383
E(Y | subject = Math)	0.258	0.243	0.258
E(Y | subject = N. Science)	0.263	0.250	0.263
E(Y | subject = H. Science)	0.392	0.371	0.391

Note: The unit of observation is student. The universe of interest is defined as students aged 17 to 19 finishing high-school who were present for all tests, comprising a set of 4.708.419 observations. We rank municipalities by the number of students who took the test there in our universe, and then take 56, 86 and 569 students from big, medium and small municipalities - defined here as the top 1%, percentiles 2-5% and the bottom 95% - for the sample, which is presented is the second and third columns. While the second column shows the raw proportions of the units in the sample, the third column weights the units as they are used for the empirical results

the position as a proxy for mental fatigue. Since each student can only appear on the sample once, it implies one municipality and one year, which combined with the subject of the questions implies the day and finally the temperature $a(i, q)$. Therefore the probability of a student i correctly answering a question q at position p is:

$$E(y_{i,q,p,a(i,q)}) = \alpha + \alpha_q + \alpha_p + \beta_{a,p} a(i, q) \tag{7}$$

In this specification, the student fixed effect is being identified through the variation among students, while position and question fixed effects are separately identified from the randomization of question orders. This position fixed effect allow for flexible formats for the effect of mental fatigue throughout the exam. Additionally, the mental fatigue effect that is independent of temperature is not a direct object of this study.

When using the question's position as proxy for mental fatigue, we are assuming that students answer these questions in the order presented to them - or at least that a questions' position influence when it is going to be answered. While this seems reasonable, there is one important threat to the relevance of the standard ordering: the 90 questions answered in each day are clearly divided in two subjects. Students receive a single booklet of questions each day, and every booklet have all 90 questions for the day, yet there is a clear separation between the two daily subjects.[8] Because of this, it is much more likely for students not to follow the subject order than the in-subject order. In our main estimations positions are deemed 1 to 90, assuming students follow the standard order, yet we also present results modeling positions as four sets of questions 1 to 45.

The essence of this work lies in the set $\{\beta_{a,p}\}_p$, which estimates the effect of thermal discomfort a, measured by the temperature, on the probability of correctly answering a question at position p. To study the evolution of the total effect of temperature on cognitive performance is to study the relation between $\beta_{a,p}$ for different values of p. The main intuition from the theoretical model in Section 2 is that because of a difference in initial effort the total effect of temperature is mitigated throughout the duration of the exam, therefore that $\beta_{a,p}$ is increasing in p.

The source of variation for this effect comes from the fact that ENEM is a 2-days exam, and thus we observe the same students answering questions at the same position under two different temperatures. In order to have a cleaner estimate of the effect of 1ºC on the margin of the realized temperatures, the temperature variable used in the regressions was the mean temperature minus the minimum mean temperature faced by each student, i.e., if a student faces one day of 30ºC and one day of 32ºC this variable is 0 on the colder day and 2 on the hotter day. We further simplify this estimation by adding structure to $\beta_{a,p}$ in assuming that it is linear in p. The variable p in this approach is the question's position normalized so that the first question of the day is $p = 0$ and the last at $p = 1$, and the total effect of a degree Celsius $\beta_{a,p} = \beta_a + p * \beta_{ap}$ has a component that is fixed throughout the exam and one that is directly proportional to the question's position p. Therefore:

$$E(y_{i,q,p,a(i,q)}) = \alpha + \alpha_q + \alpha_p + \beta_a a(i,q) + p\beta_{ap} a(i,q) \tag{8}$$

[8] Questions are officially numbered 1 to 180

4.2 Results

The main result comes from the estimation of (8) and suggests that a 1ºC increase in temperature causes a reduction of 0.1835% on the probability of answering any question correctly. However, it also causes another effect that is 0 on the first question and increases linearly throughout the test resulting in a positive impact of 0.1760% on the probability of answering the last question right, summing to a net effect of only -0.0075% on this specific question. In terms of (8), $\beta_a = -0.1835$ and $\beta_{ap} = 0.1760$. This result is presented on column (1) of Table 2 and it suggests that in this case there are indeed secondary consequences to the effect of temperature on cognitive performance, and these secondary consequences have a component that increases in time and have effect contrary to the direct one.

Table 2: Results Table

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Temperature	-0.33%***	-0.165%***	-0.14%***	-0.33%***	-0.23%**
(sd)	*(0.019%)*	*(0.026)*	*(0.017%)*	*(0.019%)*	*(0.096%)*
Temperature * Position	0.49%***	0.1642%***	0.12%***	0.51%***	0.375%***
(sd)	*(0.019%)*	*(0.041%)*	*(0.012)*	*(0.019%)*	*(0.061%)*
Temp.Drop * Position				-0.13%***	
(sd)				*(0.026%)*	

Note: ”Temperature” is measured in Celsius degrees as a daily average between the temperatures measured at 12h, 15h and 18h (the exam goes from 13h30 from 18h/19h on day 1/day 2). ”Position” measures the position in which questions are presented, and it is normalized so the variable goes linearly goes from 0 to 1 during the exam day (each exam day contains 90 questions divided in two subjects of 45 questions). ”Temp.Drop” measures how much the temperature drops from 12h to 18h in Celsius degrees (if temperature is 30ºC at 12pm and 27ºC at 18h then ”Temp.Drop” = 4). Column (1) presents the main result, and in column (2) the temperature measurement is normalized within municipality-year so that it is 0 at the colder day. In column (3) the positions are redefined so positions within subjects goes from 0 to 1 (in 45 questions). Column (4) adds a new variable and column (5) uses a sub-sample of municipality-year pairs that do not register temperature drops higher than 3ºC

Columns (2) and (3) test for slightly different specifications. In column (2) the temperature specification changes, normalizing every municipality-year pair so that temperature is 0 on the colder day and on the hotter day it measures the difference between days.[9] In column (3) the questions positions are normalized as 4 sets of 1-45, which stops the assumption that students follow the standard order of subjects (Human

[9] Thus if a municipality averaged 32ºC and 29ºC on the two exam days for a given year, the new values would be 3 and 0

Sciences before Natural Sciences; Reading before Math) - this also means that now a full day of exam goes from position 0 to position 1 twice. In this specification the difference in accumulated fatigue from being the second test should be absolved to questions' fixed effects.

The change in temperature specification gives a more accurate measure of the importance of a Celsius degree on the margin of the realized temperatures, while changing the position definitions is a robustness test that seems natural given the quasi-experiment.

In columns (4) and (5) of Table 2 we test for the main threat to identification: the fact that the temperature goes down during the test, which could be polluting the results. Column (4) adds a new variable that measures the temperature drop from the first to the last temperature measurement used in the daily averages. Column (5) limits the sample to municipalities that did not experience relevant temperature drops during the exam days. These results are discussed in more depth in the next sub-section.

All results points towards the existence of a mechanism as proposed in the theoretical model from section 2, and results are statistically solid. Evidence is enough to establish a causal relation temperature and cognitive performance through time in this experiment, in that an increase of 1ºC on the margin is causing students to have an increasing bonus in performance through the exam, although the total impact is negative.[10] The next sub-section discusses how to interpret the external valid of this paper given the biggest threat faced by these results.

4.3 Identification Threat

The main threat to our identification strategy using the ENEM, or the main threat to the external validity of its results, is the fact that the exam goes from 13h30pm to 18:00pm or 19:00pm - which naturally implies that the temperature goes down during the exam. There are two possible mechanisms for this within-day temperature variation to be driving our main results, a direct and an indirect one.

The direct mechanism is for students to improve their results throughout the exam because the temperature is decreasing and its point-wise negative effect on current performance is fading away. One aspect worthy noting of this possibility is that our estimations account for position fixed-effects, therefore if the bonus of a temperature drop does not vary with the temperature itself it should be captured by positions

[10] The total effect of a 1ºC is certainly negative during most of the exam and more importantly in the total impact; this negative effect, however, is already established in the literature.

fixed-effects. However, data indicates that higher temperatures at the beginning of the exam are associated with sharper temperature decreases during the exam.[11] Therefore it should be possible for the increasing positive impact captured in our regressions to be the result of this temperature drop.

Nonetheless, although temperature drops are indeed associated with higher temperatures at the beginning of the exam, they are not a necessary consequence of high temperatures and its value can vary a lot even for a given initial temperature. We leverage this variations to assess the possibility of it being the driving force behind our results. If this complication were in fact driving the results, than two things were to be expected: first the effect should be moderated not only by the mean temperature faced on that day but also by how much the temperature dropped between 12pm and 18pm. Second, the existence of the effect should depend on the existence temperature drop during the exam. None of these two predictions are true, as shown in columns (4) and (5) of Table 2.

Column (4) presents the results for the main sample under a slightly different specification, adding a new term $p\beta_{dp}d(i,q)$ to equation (8) that interacts the question position p to the total temperature drop $d(i,q)$ experienced by student i on that day between 12:00pm and 18:00pm.[12] This allows the temperature drop to cause an increasing effect throughout the exam, which should better capture the effect of this temperature drop if it were the driving force of our main result.

As can be seen in Table 2, the addition of this interactive term directly accounting for the temperature drop does not decreases the estimates of the effect of interest - and seem to actually jeopardize agent's performance, i.e., higher temperature drops are causing agents' performance to decrease performance throughout the exam. However this is not a good test for the effect of the temperature drop and estimating this effect is not the concern of this work, but rather its possible interaction with the fatigue effect we are estimating. And on that matter the fact that it barely affected the original estimation is enough to argue that these are separate mechanisms.

Column (5) presents the results for the main specification for a different sample. Instead of drawing a sample representative of the total universe, as explained in subsection 3.4, for this regression the sample was limited to the 992 pairs municipality-year that experienced temperature drops of at most 3ºC during either exam days [13]. This sample contained just over 84.000 students and the results were similar to the general

[11] this relationship is shown in Figure 2

[12] Therefore if the temperature is 32ºC at 12:00pm and 27 at 18:00pm, d(i,q) = 5

[13] For a minimum decrease of 2ºC there were 228 municipality-year pairs, thus we chose 3ºC

sample, indicating that the estimated effect is not being mediated by the daily decrease in temperature during the exam.

The indirect mechanism for this daily temperature drop to be driving our results is for students to consciously choose to self-preserve on earlier periods when temperature is higher and so is the cost of effort, projecting to exert more effort under colder temperatures later. This possible mechanism is not too different from the one we are purposing, in that both argue that higher temperatures are causing students to withhold effort on earlier periods and it is allowing them to exert more effort on later periods. Thus, they agree with withholding initial effort as the mediator of this effect.

In spite of these two mechanisms being similar in nature, it still is an important threat to the external valid of our result. If the only mechanism in play is this possible indirect effect of temperatures dropping, as the effect relies on the expectation of the existence of a later period under milder temperature conditions for it to be worth withholding effort, then our result is only valid for hours-long cognitive tasks under very specific conditions (when temperatures are expected to drop). What generalizes this effect in the mechanism we are proposing is that the temperature drop is not the relevant difference between the costs of exerting effort in different periods - that role is rather played by mental fatigue.

This mechanisms could be captured by $Temperature * Position$ because higher temperatures could induce the expectation of higher temperature drops, causing students to withhold more initial effort. However, this effect should be better captured by $TemperatureDrop * Position$ and thus the addition of this term to the formula in column (4) of Table 2 should capture at least a significant share of this mechanism that would otherwise be causing our results on $Temperature * Position$. Therefore the results in columns (4) and (5) constitute strong evidence that the temperature drop is not the driving force behind the results in column (1).

5 Conclusion

This work presents solid evidence that in the specific scenario studied a one degree increase in temperature causes a negative effect that fades away throughout the duration of the exam. The theoretical model suggests that the mechanism responsible for mitigating this effect in time is innate to the nature of the temperature effect, and therefore should be present in all hours-long cognitive task. The empirical section 4 also presents evidence to support this hypothesis, supporting the existence of a mechanism in this direction and discarding other possible explanations for this results.

This contribution uncovers nuances on how people's performing capabilities reacts to both thermal distress and mental fatigue. Mainly by highlighting the evolution of the effect of heat on cognitive performance through time, exploring other consequences from its documented impact on effort. This also have practical implications regarding dealing with this effect. One could argue that the fact that the exam leveraged as quasi-experiment here is so long is actually reducing the unfair impact of temperature on average relative performance.

Another information ill-documented in the literature that is essential to the narrative presented connecting the results in this work is the intuitive relationship assume between effort and mental fatigue. It is worthy reminding that mental fatigue is defined here as the decrease in performance throughout a long cognitive task, measured on the same activity inducing the fatigue. Although its definition by itself involve the fact that this decline come from previously performing a cognitive task, what is not yet clearly documented in the literature is that choosing to exert more effort during the performance of a task increases the fatigue accumulated through this process.

References

Balart, Pau, Matthijs Oosterveen, and Dinand Webbink (2018). "Test scores, noncognitive skills and economic growth". *Economics of Education Review* 63, pp. 134–153. ISSN: 0272-7757. DOI: https://doi.org/10.1016/j.econedurev.2017.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0272775717306854.

Behrer A. Patrick, Park R. Jisung Goodman Joshua (2021). "Learning is inhibited by heat exposure, both internationally and within the United States". *Nat Hum Behav* 5, 19–27. DOI: 10.1038/s41562-020-00959-9.

Borghans, Lex and Trudie Schils (Dec. 2018). "Decomposing Achievement Test Scores Into Measures of Cognitive and Noncognitive Skills". *Available at SSRN*. DOI: http://dx.doi.org/10.2139/ssrn.3414156.

Brown, Christina L, Supreet Kaur, Geeta Kingdon, and Heather Schofield (2022). *Cognitive Endurance as Human Capital*. Working Paper 30133. National Bureau of Economic Research. DOI: 10.3386/w30133. URL: http://www.nber.org/papers/w30133.

Graff Zivin, Joshua, Solomon M. Hsiang, and Matthew Neidell (2018). "Temperature and Human Capital in the Short and Long Run". *Journal of the Association of Environmental and Resource Economists* 5.1, pp. 77–105. DOI: 10.1086/694177. eprint: https://doi.org/10.1086/694177. URL: https://doi.org/10.1086/694177.

Melo A.P., Suzuki M. (2023). "Temperature, Effort, and Achievement". *JMP*, pp. 1–61. URL: https://www.apmelo.com/files/manuscript_MM2.pdf.

Moore R. D., Romine M. O'connor P. Tomporowski P. (2012). *Journal of Sports Sciences* 30.9, pp. 841–850. DOI: 10.1080/02640414.2012.675083. URL: https://pubmed.ncbi.nlm.nih.gov/22494399/.

Park, R. Jisung (2022). "Hot Temperature and High-Stakes Performance". *Journal of Human Resources* vol. 57, issue 2, pp. 400–434.

Park, R. Jisung, Joshua Goodman, Michael Hurwitz, and Jonathan Smith (2020). "Heat and Learning". *American Economic Journal: Economic Policy* 12.2, pp. 306–39. DOI: 10.1257/pol.20180612. URL: https://www.aeaweb.org/articles?id=10.1257/pol.20180612.

Qian, Shaowen, Min Li, Guoying Li, Kai Liu, Bo Li, Qingjun Jiang, Li Li, Zhen Yang, and Gang Sun (2015). "Environmental heat stress enhances mental fatigue during sustained attention task performing: Evidence from an ASL perfusion study". *Behavioural Brain Research* 280, pp. 6–15. ISSN: 0166-4328. DOI: https://doi.org/10.1016/j.bbr.2014.11.036. URL: https://www.sciencedirect.com/science/article/pii/S0166432814007724.

Reyes, Germán (2023). "Cognitive Endurance, Talent Selection, and the Labor Market Returns to Human Capital". *JMP*.

Roach, Travis and Jacob Whitney (2022). "Heat and learning in elementary and middle school". *Education Economics* 30.1, pp. 29–46. DOI: 10.1080/09645292.2021.1931815. eprint: https://doi.org/10.1080/09645292.2021.1931815. URL: https://doi.org/10.1080/09645292.2021.1931815.

Smith M.R., Chai R. Nguyen H.T. Marcora S.M. Coutts A.J. (2019). "Comparing the effects of three cognitive tasks on indicators of mental fatigue". *The Journal of Psychology*. DOI: 10.1080/00223980.2019.1611530. URL: https://pubmed.ncbi.nlm.nih.gov/31188721/.

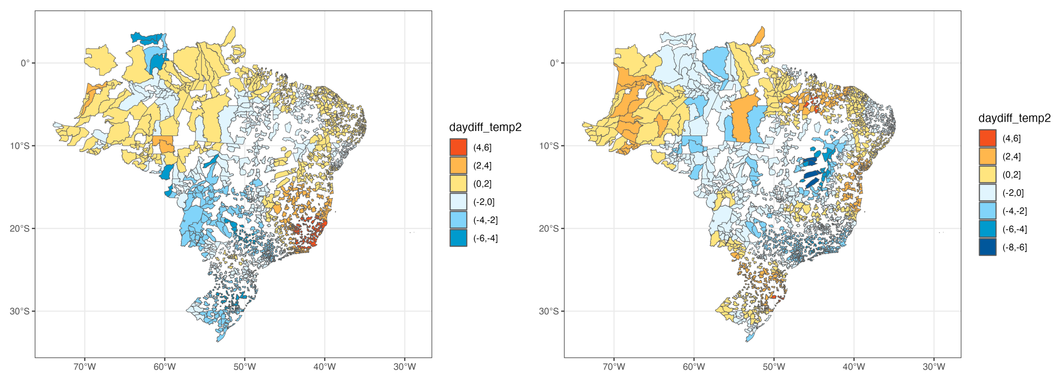
Solivio M., Haboy D. Robas K. Gumasing M. (2021). "A Case Study of Perceived Effect of Thermal Comfort to the Mental Fatigue Level oF Students during Online Class".

Van Cutsem, Jeroen et al. (2017). "Effects of Mental Fatigue on Endurance Performance in the Heat." *Medicine and science in sports and exercise* 49.8, pp. 1677–1687. DOI: 10.1249/MSS.0000000000001263.

Vu, Tien Manh (Jan. 2019). *Effects of weather on human capital in Vietnam*. AGI Working Paper Series 2019-01. Asian Growth Research Institute. URL: https://ideas.repec.org/p/agi/wpaper/00000150.html.

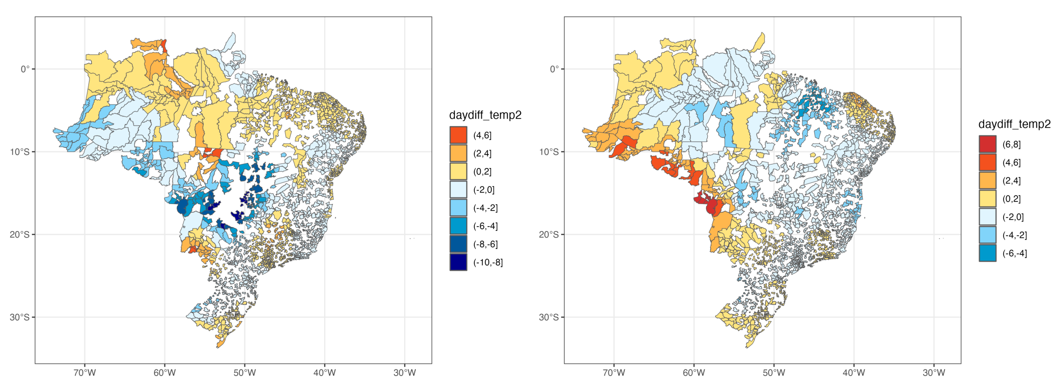
6 Figures

Figure 1: Figure 1: Exam-Days Temperature Variation for 2013 and 2014



Note: The maps in figures 1A (left) and 1B (right) display the difference in temperature between the first and the second day of ENEM's exams for each municipality hosting exam sites, for the years 2013 and 2014 respectively. The same maps for 2015 and 2016 are available on Figure 2 in section 6

Figures 1C and 1D: Temperature Variation between Exam Days for 2015 and 2016



Note: The maps in figures 1A (left) and 1B (right) display the difference in temperature between the first and the second day of ENEM's exams for each municipality hosting exam sites, for the years 2014 and 2015 respectively.

7 Appendix A:

From the Implicit Function Theorem:

$$\begin{bmatrix} \frac{da_1}{d\gamma} \\ \frac{da_2}{d\gamma} \end{bmatrix} = - \begin{bmatrix} \frac{d(2)}{da_1} & \frac{d(2)}{da_2} \\ \frac{d(3)}{da_1} & \frac{d(3)}{da_2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \frac{d(2)}{d\gamma} \\ \frac{d(3)}{d\gamma} \end{bmatrix} \tag{9}$$

Yet

$$\begin{bmatrix} \frac{d(2)}{da_1} & \frac{d(2)}{da_2} \\ \frac{d(3)}{da_1} & \frac{d(3)}{da_2} \end{bmatrix}^{-1} = \frac{1}{\det \begin{bmatrix} \frac{d(2)}{da_1} & \frac{d(2)}{da_2} \\ \frac{d(3)}{da_1} & \frac{d(3)}{da_2} \end{bmatrix}} \begin{bmatrix} \frac{d(3)}{da_2} & -\frac{d(2)}{da_2} \\ -\frac{d(3)}{da_1} & \frac{d(2)}{da_1} \end{bmatrix}. \tag{10}$$

Then consider:

$$\frac{d(2)}{da_1} = -\frac{\alpha}{a_1^2} - \rho(\rho-1)a_2 a_1^{(\rho-2)}, \tag{11}$$

$$\frac{d(2)}{da_2} = \frac{d(3)}{da_1} = -\rho a_1^{(\rho-1)}, \tag{12}$$

$$\frac{d(3)}{da_2} = -\frac{\alpha}{a_2^2}, \tag{13}$$

$$\frac{d(2)}{d\gamma} = -e^{\gamma}, \tag{14}$$

$$\frac{d(3)}{d\gamma} = -e^{\gamma}. \tag{15}$$

Substituting (11)-(15) in the fraction in the right-hand side of (10):

$$\begin{bmatrix} \frac{da_1}{d\gamma} \\ \frac{da_2}{d\gamma} \end{bmatrix} = -\Delta \begin{bmatrix} -\frac{\alpha}{a_2^2} & \rho a_1^{(\rho-1)} \\ \rho a_1^{(\rho-1)} & -\frac{\alpha}{a_1^2} - \rho(\rho-1)a_2 a_1^{(\rho-2)} \end{bmatrix} \cdot \begin{bmatrix} -e^{\gamma} \\ -e^{\gamma} \end{bmatrix} \tag{16}$$

where

$$\Delta^{-1} =$$

Now going back to (16), the desired derivatives can be calculated as:

$$\frac{da_1}{d\gamma} = \frac{2\gamma}{\Delta}(-\frac{\alpha}{a_2^2} - \sqrt{a_1} - \frac{1}{2\sqrt{a_1}}) \tag{17}$$

$$\frac{da_2}{da} = \frac{2\gamma}{\Delta}(-\frac{\alpha}{a_2^2} - \frac{1}{2\sqrt{a_1}} + \frac{a_2}{4a_1\sqrt{a_1}}). \tag{18}$$