TEMPERATURE AND TEACHER ABSENCE: EVIDENCE FROM 4,085 SCHOOLS IN INDIA

Presented by
Christine Léonard
#300019982
Supervisor: Professor Anthony Heyes

ECO6999
Major Research Paper

Department of Economics
Faculty of Social Sciences
University of Ottawa
Fall 2018

1. Introduction

Theory of human capital suggests that human capital is an asset composed of both education (schooling years) and talent, but also health and many other characteristics. The amount of talent that one is endowed is fixed, but education must be acquired through learning. Teachers should thus be one of the key roles in education, which in turn should determine one's chances of succeeding in life. The economic production is commonly composed of capital and labour. This second component is also known as human capital and tends to be more efficient when the workers have had higher education. The level of high-skilled workers – people who had the opportunity of getting good education – is essential in understanding how human capital is related to the level of economic development, especially since human capital contributes to its own growth through innovation and technical change.

As a result, several authors have attempted to identify different determinants that could affect human capital through teacher absence. For instance, Kremer et al. (2005) study teacher absence in India. They found that 25% of teachers were absent during unannounced visits to schools. In some states, absenteeism rates can even exceed 40%. Kremer et al. (2005) find no evidence that teacher absence is correlated with higher salary. However, they find that some incentives actually work: schools with better infrastructure contribute positively to teacher attendance. In addition, using a randomized experiment in India, Duflo et al. (2012) find that teachers react positively to financial incentives. More specifically, in their experiment, the group exposed to financial incentives was 21% less absent relative to the control group.

The two above-mentioned studies show a certain importance of different determinants in the absenteeism of teachers. It is therefore interesting to look at other potential determinants. An interesting and still very under-studied candidate is temperature. We already know that high temperatures are associated with lower levels of output, especially in poor countries (Dell et al., 2012), and India is not escaping the current global warming trend (IMD, 2015). Therefore, the aim of this paper is to study the *impact of temperature on teacher presence in India*.

We address this issue by conducting an analysis that estimates the effects of temperature on teacher absence in India. We find an inverted U-shaped relationship between temperature and teacher attendance, with a peak around 25°C. Above this threshold, our results show that, if on average 85% of teachers are present at school, one additional degree (°C) will raise absences by 8.1%. In other words, the higher the temperature, the bigger the effect of one additional degree.

Our result is robust to different regression methods (logit, probit, and OLS) and to several specification checks.

This result is crucial as it presents one key factor that affects future economic outcomes through the future workforce's education. With higher teacher absenteeism due to higher temperatures, children are less educated. This leads to a larger proportion of unskilled workers without human capital, which generates lower GDP. Slow growth will in turn limit the capacity to invest in health and education, creating a vicious cycle. In other words, it is likely that high temperatures play a key role in this issue where human capital, economic development, and welfare are closely interconnected. Our analysis also demonstrates that the presence of fans in classrooms and other investments in better infrastructure allow to tackle this problem, by reducing the strong negative impact of heat on attendance.

The rest of the paper is organised as follows. Section 2 summarizes the main results of the previous literature. Section 3 describes the data used for teacher information and weather. Having matched each interview to all the weather information of the location of the interview of the day, Section 4 details the chosen specification. Section 5 presents the benchmark results followed by several robustness checks. Section 6 extends our main results by including different variables that could logically affect the estimated effect. Section 7 concludes the paper.

2. Review of previous works

When thinking about climate change, our mind imagines a polar bear on a melting ice floe that is getting smaller and smaller. It comes only second to mind that higher temperatures can affect our lives in several other ways, such as affecting output production (see for instance Costa et al., 2016), or even causing cognitive impairment. It is quite hard to determine the exact cost generated by increasing temperatures given that they notably affect health, labour supply, and labour productivity (Heal and Park, 2016). The heat balance of the human organism is determined through a combination of temperature and humidity. Heat can cause direct feelings of physical and psychological discomfort, leading to psychomotor performance deterioration, followed by physiological damage. Depending on the activity and the degree of exposure to heat, one can also feel tiredness, or experience troubles concentrating, which in turn diminishes task productivity. According to Barrow and Clark (1998), if the body cannot correctly dissipate the heat, it can lead to dizziness, muscle cramps, and fever. In some extreme cases, it can even lead to mortality, especially for older individuals and children. It is important to note that cold days have impacts on mortality as well, but the effect is greater in response to heat (Fisk et al., 2006; Deschênes and Moretti, 2009).

As shown by Miguel et al. (2004), weather is known to affect agricultural production in the whole world, especially in poor countries, while its impact on non-agricultural labour supply remains underestimated. Also, if we calculate the economic output only by considering capital in the production, but ignoring labour supply, there could be a downward bias caused by climate damages (Hsiang, 2010). Even in the US, hot weather has negative impact on economic performance, even though individuals have more opportunities to make hot days more sufferable (e.g., by getting an A/C or a fan). The solutions are however harder to reach in poorer countries with lower levels of electrification and less infrastructure (Barreca et al., 2012; Deryugina and Hsiang, 2014). In 2016, only 84.5% of the Indian population had access to electricity (World Bank). It is however interesting to note that, according to Chen and Yang (2017), output in high-temperature regions suffers less from hot temperature compared to low-temperature regions, suggesting that individuals are able to adapt to heat. These authors argue that the long-term and, to a lesser extent, the short-term adaptations to heat make it more difficult to determine the precise future costs of climate change. Using US data, Graff Zivin and Neidell (2014) find evidence that, in order to calculate welfare costs, behavioural responses by firms and individuals need to be taken into account. They also show that on days with maximum temperature above 30°C, individuals working in high heat exposure will reduce time allocated to labour by one hour. They find an inverted U-shaped relationship between maximum temperature and time spent outdoors. Somanathan et al. (2015) find that sustained heat leads to a decrease in worker attendance in Indian manufacturing. On the other hand, it appears that the presence of a cooling system on the workplace dissipates the impact of heat on productivity, but makes no difference on absenteeism.

In India, high temperatures are associated with poorer mathematics and reading performance among students (Garg et al., 2018). Other evidence shows that children in Texas tend to be more absent due to carbon monoxide (Currie et al., 2009).

With more than 200 million children across India, education represents a major concern. The fiscal cost caused by teacher absence in India is estimated to $1.5 billion per year, but we hardly know the real costs of lower performance from students, who represent the future human capital, especially in a country that is still catching up in terms of development. Huge investments in primary schools, such as better infrastructure, teacher quality, or student-teacher ratios, have been made by the government in the past decade. It seems however that increases in inspections are the only real way to positively affect teacher attendance, and offset the consequences of weak governance (Das et al., 2017). Kremer et al. (2005) find results in line with the positive relationship between more inspections and attendance, while suggesting a negative relationship between upgrades in infrastructure and absenteeism. They also find higher absence rates in poorer states, but no effect from higher pays.

We here want to study the consequences caused by heat on labour supply in the context of Indian schools. Our main study question is to determine heat impacts on the probability of a teacher missing school in India. It is common sense to think that teachers are the main school-based determinant for students' academic accomplishments. Teacher absences are worth worrying as empirical results show that they are associated with lower student achievement in the US (Clotfelter et al., 2007; Miller et al., 2008). The absence rate is quite alarming both in developed and underdeveloped countries. The numbers go up to 36% of absent teachers in the US, and 42% in some states of India (Kremer et al., 2005; Miller, 2012). Some studies on teacher absence in India have already been done, but this work is the first one linking it to temperatures.

Our research shows an inverted U-shaped relationship between maximum temperature of the day and teacher presence at school. The peak of this inverted U shape is at 20°C to 25°C, which is probably considered as the most comfortable temperature to work. When temperatures are higher than this threshold, we find that any 1°C increase will raise teacher absence by 8% to 9.4%, depending on the regression method. This result is robust to several changes. Our findings also establish an impact due to maximum temperature of the day before, and to the amount of precipitations encountered two days before. We find no significant difference due to teacher gender, but good infrastructure, such as the presence of fan, indoors classrooms, or access to electricity, plays a strong positive role on teacher attendance. The extensions of our main result strongly suggest that investment in infrastructure could be a big step towards better levels of teacher attendance in India.

3. Data

School data

The school level data used in this study come from the Indian Human Development Survey[1] (IHDS) jointly organised by the University of Maryland and the National Council of Applied Economic Research. This study uses the second round of interviews, covering 24,459 teachers working in 4,085 primary schools in 355 districts of India. Following a specific agenda, the interviewers go once in each school in order to complete a questionnaire regarding school facilities, and school staff. Questions about school characteristics are very broad, going from the language of the school to the presence of toilets for students, allowing us to take into consideration some of those features to extend our main results.

We only keep observations with no missing values regarding teacher presence on the day of the interview, and drop observations about employees who are neither teachers or headmasters. This results in 24,459 observations going from January 2011 to May 2013.

Weather data

In order to measure daily maximum temperature, minimum temperature, precipitation, and relative humidity, we use weather data provided by the National Climatic Data Center at the National Oceanic and Atmospheric Administration. The file contains 4,324 weather observation points throughout India. Each point contains weather data for every day for 14 years. The whole data set is observed by satellite. We match schools to all the weather stations that are within a 100km radius using their respective longitude and latitude. Then, we take the inverse-distance weighted average information regarding the day of each interview (temperatures, precipitations, humidity index) given by all the weather stations within the 100km radius. This way, each school finds its matched weather information for the day of interview.

4. Empirical approach

Having matched each interview to all the weather information of the day, we summarize in Table 1 all these weather variables, and our dependent variable (teacher presence).

[1] For more information on the survey and the states visited, please see https://ihds.umd.edu

Table 1: Summary statistics

	Mean	Std. Dev.	Min	Max	Obs
Presence	0.85	0.35	0	1	24,459
Maximum temperature (°C)	31.52	6.86	-0.35	48.56	24,459
Minimum temperature (°C)	18.91	7.08	-10.57	32.67	24,459
Precipitation (mm)	4.63	11.60	0	163.78	24,459
Relative humidity (%)	55.11	0.23	9.50	98.09	24,459

We estimate the effects of temperature on the presence of teachers from school *s* in district *d* during month m on day *t* using the following regression specification:

$$y_{sdmt} = \alpha + \beta\, Temp_{sd} + M_{dm} + W_t + \varepsilon_{sdmt} \quad (1)$$

where $y = 1$ when the teacher is present on the day of the interview, and $y = 0$ if he is absent. $Temp_{sd}$ is the maximum temperature in school *s* area in district *d* on the day of interview, M_{dm} is a vector of district-by-month controls, W_t controls for potential effects due to the day of the week, and ε_{sdmt} is the error term. We cluster standard errors on school-by-season level in order to address potential spatial correlation in the error term, which leads to 3298 clusters.

Having controlled for both location and temporal fixed effects, temperature realisation on day of visit is as good as randomly assigned. This assumption allows us to draw the results presented in the next section.

Considering 25°C as the most comfortable temperature to work, we omit all observations where the maximum temperature on the day of interview is beneath this threshold. The idea here is to analyse the impact of temperature getting hotter as we expect to see an inverted U-shaped relationship between temperature and teacher presence. This way we avoid the effects of colder temperatures on our coefficient of interest β. All the following regressions are done using observations with maximum temperature higher or equal to 25°C, unless specified otherwise.

5. Results

We first interpret how the temperature affects the presence of teachers in India using OLS, probit, and logit. Then, we demonstrate the robustness of this result, using for instance alternative fixed effects, different cut-offs, different independent variables.

Main results

Table 2 shows that temperature has a statistically significant negative impact on teacher presence in India. In column (1), we include no control, in column (2) we include weather controls and both days of the week and district-by-month fixed effects, and in columns (3a) to (3c) we only include both fixed effects. The coefficient obtained from the OLS regression implies that, if on average 85% of the teachers are present, one additional degree (°C) raises absences by 8.1%.

The logit interpretation works using probabilities. The probability of teacher *i* from school *s* during month *m* on day *t* of being present on the day of interview can be calculated using:

$$P(presence_i = 1) = \frac{1}{1 + e^{-z}}$$

where $z = \alpha + \beta\ Temp_s + M_{dm} + W_t$

The probability of teacher *i* being present can also be interpreted as the presence rate in the school. For instance, the presence rate in a school located in district Ernakulam in June (M_{dm}=1.46) on a Friday (W_t = - 0.44), where the maximum temperature is 30°C will be:

$$P(presence = 1|Temp = 30) = \frac{1}{1 + e^{-(3.95-0.1045*30+1.46-0.44)}} = 86.24\%$$

If the maximum temperature rises from 30°C to 31°C, everything else held constant, then

$$P(presence = 1|Temp = 35) = \frac{1}{1 + e^{-(3.95-0.1045*31+1.46-0.44)}} = 84.95\%$$

A rise of 1°C when maximum temperature is 30°C is associated with a 9.4% increase in teacher absences. The impact of hotter temperature will be different depending on the district, the month, the day of the week, but also on the level of the temperature used in the comparison (i.e. a rise of 1°C will have a larger impact on absence if it is 40°C compared to the case with a maximum temperature of 25°C).

We need to use a cumulative standard normal distribution to interpret the coefficient generated by the probit regression. We can compute the probability of a teacher being present using

$$P(presence = 1) = \phi(\alpha + \beta\ Temp_s + M_{dm} + W_t)$$

where ϕ is the cumulative standard normal distribution.

The easiest interpretation is to compute the predicted probability of presence for alternative values of temperature. Let's take the same example used for the logit interpretation, but with probit we get M_{dm} = 0.76 and W_t = - 0.22:

$$P(presence = 1|Temp = 30) = \phi(2.17 - 0.0556 * 30 + 0.76 - 0.22) = \phi(1.042)$$
$$= 85.08\%$$

and

$$P(presence = 1|Temp = 30) = \phi(2.17 - 0.0556 * 31 + 0.76 - 0.22) = \phi(0.9864)$$
$$= 83.89\%$$

When the maximum temperature is 30°C, the coefficient can be interpreted as a 8% increase in absences when there is a 1°C rise in temperature, everything else held constant. The impacts of temperature also depend on other variables, as with logit.

The result is robust when adding weather controls to the specification. We do not include school controls because there would be too many variables compared to the number of observations, which saturates the model.

Table 2: Main results

	(1)	(2)	Preferred		
	No controls	Weather inclusion	(3a) OLS	(3b) Logit	(3c) Probit
Temperature	-0.0132***	-0.0121***	-0.0121***	-0.1045***	-0.0556***
	[0.0011]	[0.0017]	[0.0017]	[0.0142]	[0.0074]
Observations	20,044	20,044	20,044	19,064	19,064
Weather controls	N	Y	N	N	N
Day of the week FEs	N	Y	Y	Y	Y
District-month FEs	N	Y	Y	Y	Y

Notes: Standard errors are clustered on school-season in brackets. Column (1) has no controls and is regressed using OLS. Column (2) includes weather controls to the preferred specification and is regressed using OLS. Columns (3a), (3b), and (3c) are the preferred specification, with day of the week and district-by-month fixed effects, and are regressed respectively using OLS, logit, and probit.
* significant at 10% ** significant at 5% *** significant at 1%.

Alternative fixed effects

We compute in Table 3 other structures of our model in order to examine whether our coefficient of interest is robust to modelling choices. We choose district-by-month and day of the week fixed effects as the preferred regression, as they seem to be the most natural ones. Table 3 shows that other combinations of fixed effects do not affect the sign of the coefficient, nor the size of the effect. In column (1), we start with a specification that only includes year and district fixed effects; the structure of the response remains unchanged here, but the coefficient becomes smaller, while still statistically significant. In column (2), we replace district fixed effects with school fixed effects. We use season and school fixed effects in column (3). To allow spatial heterogeneity, we use district-by-month fixed effects in columns (4) and (6), and district-by-season fixed effects in column (5). We add day of the week fixed effects in column (6), making it our preferred specification.

Table 3: Alternative fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	-0.0086***	-0.0141***	-0.0120***	-0.0120***	-0.0120***	-0.0121***
	[0.0017]	[0.0012]	[0.0015]	[0.0017]	[0.0015]	[0.0017]
Year FE	Y	Y	N	N	N	N
District FE	Y	N	N	N	N	N
Season FE	N	N	Y	N	N	N
Day of week FE	N	N	N	N	N	Y
District-season FE	N	N	N	N	Y	N
District-month FE	N	N	N	Y	N	Y
School FE	N	Y	Y	N	N	N
Observations	20,044	20,044	20,044	20,044	20,044	20,044

Notes: We regress using OLS. Standard errors are clustered on school-season in brackets. Column (1) includes year and district fixed effects. Column (2) includes year and school fixed effects. Column (3) includes season and school fixed effects. Column (4) includes district-by-month fixed effects. Column (5) includes district-by-season fixed effects. Column (6) includes district-by-month and day of the week fixed effects, and is the preferred combination.

Robustness to different cut-offs

In Table 4, we check whether the size of the coefficient could be caused by the cut-off we chose to put at 25°C. We predict an inverted U-shaped relationship between temperature and attendance, with a peak around 25°C, so that higher temperatures have stronger negative effects on presence, while lower temperatures have a weaker effect that could even be positive (it can be more comfortable to go to work by 15°C than by 3°C). Column (2) in Table 4 shows that the inclusion of observations with a lower maximum temperature cut-off does weaken the size of our coefficient, as predicted, but its sign and amplitude still remain unchanged. Column (3) in Table 4 confirms that the size of the coefficient will grow bigger due to the inclusion of higher temperature observations, but the difference is negligible.

Table 4: Different cut-offs

	(1) Preferred	(2) Cut-off at 22.5°C	(3) Cut-off at 27.5°C
Temperature	-0.0121***	-0.0110***	-0.0125***
	[0.0017]	[0.0015]	[0.0019]
Observations	20,044	22,304	17,624

Notes: We regress using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification, using all the observations with maximum temperature superior or equal to 25°C. Column (2) uses a 22.5°C cut-off. Column (3) puts the threshold at 27.5°C.
* significant at 10% ** significant at 5% *** significant at 1%.

Robustness to outliers

In Table 5, we test whether the appearance of particularly high maximum temperatures in our data can distort the regression and the size of the effect. One could argue that the size of our result is mainly driven by these outliers. In order to address this potential issue, we can trim or winsorize the data set. Table 5 shows our coefficient of interest, first after trimming the database, consecutively dropping observations with maximum temperature values higher than the 99%, the 95%, and the 90% range, and then after winsorising for the same ranges. These thresholds are respectively 45.42°C, 42.68°C, and 41.17°C. It appears that these changes have no real impact on our coefficient of interest as it remains negative, statistically significant, and the change in size is negligible. The small shrinkage of our coefficient of interest when trimming the database is predictable as the highest temperatures have the most important impact on attendance due to the inverted U-shaped relationship.

Table 5: Robustness to trimming and winsorising

	(1) No change	(2) 99% range	(3) 95% range	(4) 90% range
Temperature (Trimming)	-0.0121***	-0.0104***	-0.0091***	-0.0095***
	[0.0017]	[0.0016]	[0.0017]	[0.0019]
Observations	20,044	19,846	19,032	18,041
Temperature (Winsorising)	-0.0121***	-0.0119***	-0.0114***	-0.0116***
	[0.0017]	[0.0017]	[0.0017]	[0.0017]
Observations	20,044	20,044	20,044	20,044

Notes: We regress using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is our preferred specification. Columns (2), (3), and (4) are respectively trimmed and winsorised for values of maximum temperatures higher than the 99%, 95%, and 90% range.
* significant at 10% ** significant at 5% *** significant at 1%.

Robustness to other choices of independent variable

In Table 6, we use maximum temperature as the main independent variable of our preferred specification, as minimum temperature is mainly felt at night, so that it probably has a smaller impact on the decision of going or not to work. Besides, it can be 18°C for several days at night, while the maximum temperature during the day can vary from 30°C to 43°C for these days, giving different levels of heat. Column (2) in Table 6 shows that minimum temperature still has an impact on presence, with the same sign as the maximum temperature coefficient, but its size effect is smaller. We think that the mean temperature is not the best estimation tool as there can actually be a difference of 30°C between maximum temperature and minimum temperature on the same day. The mean temperature does not best represent the real feeling of temperature during a day. However, it appears from column (3) that the sign of the coefficient is unaffected, while somewhat bigger.

Table 6: Other independent variable

	(1) Max Temp	(2) Min Temp	(3) Mean Temp
Temperature	-0.0121***	-0.0098***	-0.0158***
	[0.0017]	[0.0018]	[0.0021]
Observations	20,044	20,044	20,044

Notes: We regress using OLS with district-by-month and day of the week fixed effects, using observations where maximum temperature is above 25°C. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification with maximum temperature of the day as main variable. Column (2) uses minimum temperature. Column (3) uses the mean between maximum and minimum temperature.

Robustness to other specifications

Heavy rainfall can have effects on diarrhea through contamination of drinking water (Carlton et al., 2014). A teacher could miss work because he is sick due to water contamination, and not because of the temperature. We can also imagine that mud roads become impassable due to heavy rain on previous days. Table 7 presents results showing that precipitation has really little effect on teacher's attendance at school. It seems however that precipitation two days before has a negative significant effect on attendance.

We added a squared term of maximum temperature to our preferred specification in column (5). Our main coefficient becomes positive, but our squared term coefficient is negative, suggesting an inverted U-shaped relationship between temperature and attendance, as predicted. When taking the first derivative of this equation, we find that the turning point is at 22.5°C. This result shows that the relationship between temperature and teacher presence is positive up to 22.5°C, but that it becomes negative after this point. This result is in line with our predictions.

Table 7: Robustness to other specifications

	(1) Preferred	(2) Rain	(3) Rain lags	(4) Interaction	(5) Squared term
Temperature$_t$	-0.0121***	-0.0123***	-0.0124***	-0.0122***	0.0185***
	[0.0017]	[0.0017]	[0.0017]	[0.0018]	[0.0040]
Rain$_t$	-	0.0002	-	0.0027	-
		[0.0005]		[0.0050]	
Rain$_{t-1}$	-	-	0.0008*	-	-
			[0.0004]		
Rain$_{t-2}$	-	-	-0.0013***	-	-
			[0.0004]		
Rain$_t$*Temperature$_t$	-	-	-	-0.0001	-
				[0.0002]	
(Temperature$_t$)^2	-	-	-	-	-0.0004***
					[0.0001]
Observations	20,044	20,044	20,044	20,044	24,459

Notes: We regress using OLS with district-by-month and day of the week fixed effects for columns (1) to (4), and no fixed effects for column (5). Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Column (2) adds a rain variable. Column (3) adds lagged values for rain. Column (4) adds a rain and an interaction terms. Column (5) adds a squared term of maximum temperature and is regressed on the whole data set.

Robustness to lagged values of maximum temperature

It is reasonable to think that the human metabolism might not only react to the temperature of the day, but also to the weather it went through during the previous days. We run in Table 8 the preferred specification, adding up to two lags to our variable of interest. In both columns (2) and (3) the size of our coefficient decreases, sharing temperature effects on attendance with lags coefficients. Nonetheless, the lagged coefficient is only statistically significant at 5% in column (2) while none the lagged coefficients are statistically significant in column (3).

Table 8: Robustness to lagged values of maximum temperature

	(1) Preferred	(2) 1 lag	(3) 2 lags
Temperature$_t$	-0.0121***	-0.0079***	-0.0076***
	[0.0017]	[0.0024]	[0.0024]
Temperature$_{t-1}$	-	-0.0048**	-0.0022
		[0.0023]	[0.0028]
Temperature$_{t-2}$	-	-	-0.0034
			[0.0024]
Observations	20,044	20,044	20,044

Notes: The regression is done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Columns (2) and (3) respectively adds one and two lags of maximum temperature.

Other robustness checks

In Table 9, we run several other specifications and controls. Column (2) is run just as our preferred specification but excluding schools that have more than one shift. Column (3) is run on a subsample of schools that have piped water as it is an element of good infrastructure. It appears that our coefficient of interest reduces in size, but the difference with our preferred specification is quite small. We also check if the presence of a written evaluation process for teachers plays a role in their attendance. Column (4) runs the preferred specification but adding an interaction term for the existence or not of an evaluation process in the school. It appears that the existence of an evaluation process has no impact on our main coefficient. One explanation could be that even a bad evaluation leads to no sanction, so that it does not affect teachers' decision of coming to work. Column (5) runs the preferred specification but only on schools for which the most important criterion on which teachers are evaluated is attendance. In this case, our coefficient gets very small, but insignificant due to the insufficient number of observations. Column (6) excludes observations regarding headmasters. Our coefficient grows in size, suggesting that teachers tend to be more absent than headmasters when temperatures are high. Column (7) runs the preferred specification but only on schools visited in the morning. It appears from Table 9 that our coefficient of interest is not much disturbed by these changes, except in column (5) where its size shrinks, but gets insignificant.

Table 9: Other robustness checks

	(1) Preferred	(2) Shifts	(3) Piped water	(4) Evaluation process	(5) Evaluation criteria	(6) Excludes headmaster	(7) Visits in the morning
Temperature	-0.0121***	-0.0127***	-0.0112***	-0.0120***	-0.0010	-0.0135***	-0.0153***
	[0.0017]	[0.0017]	[0.0025]	[0.0017]	[0.0057]	[0.0019]	[0.0022]
Temperature* Evaluation	-	-	-	0.0002	-	-	-
				[0.0003]			
Observations	20,044	17,646	10,063	19,597	2,535	16,768	11,523

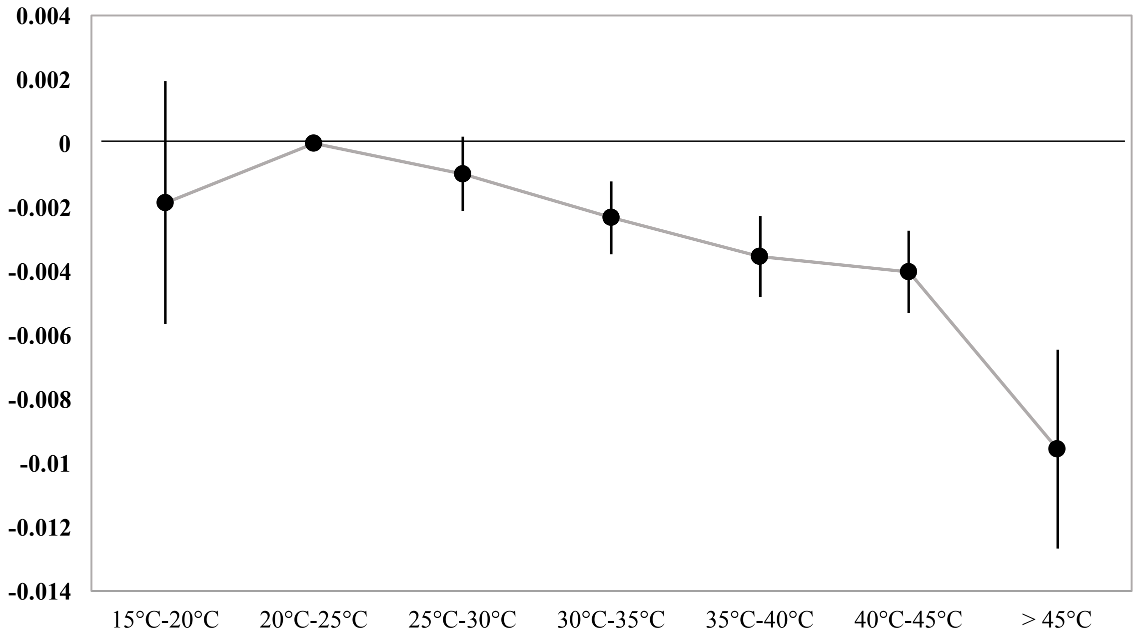
Notes: The regression is done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Column (2) excludes observations with more than one shift. Column (3) only includes observations with piped water. Column (4) adds an temperature times existence of an evaluation process interaction term. Column (5) only includes observations where attendance is the first evaluation criterion. Column (6) excludes observations with headmasters. Column (7) only includes observations with visits to school done in the morning.
* significant at 10% ** significant at 5% *** significant at 1%.

Robustness to bins

In Figure 1, we check if our results are robust to a specification with bins. We construct nine bins, each corresponding to a category of maximum temperatures. The first bin corresponds to any day during which the maximum temperature is equal or lower to 10°C. The second corresponds to days where the maximum temperature is between 10°C and 15°C (15°C included). It goes on for a total of seven bins, each bin including a range of 5 degrees, and the maximum bin being any maximum temperature higher than 45°C. This way, if the maximum temperature of the day for an observation is 26°C, the variable corresponding to the fifth bin will display 1 for this observation, while all the other ones will be 0. We then construct new bins where all the ones are replaced by their corresponding maximum temperature, while the zeros remain zeros. We regress presence on all these bins, omitting the fourth bin (20°C – 25°C) as it appears to be the most comfortable temperature category. We drop bin 1 and bin 2 as they respectively contain 140 and 11 observations different from 0, which we do not find sufficient in order to get significant effects.

As predicted, Figure 1 shows an inverted U-shaped relationship between presence and maximum temperature. The range of maximum temperatures between 15°C and 20°C and between 25°C and 30°C has a coefficient with no statistically significant impact in comparison with the category of reference. This could suggest that the comfort zone is set between 15°C and 30°C. Higher than that, temperatures have strong negative effect on attendance. The higher the temperature, the more negative is the impact of a 1°C increase on presence.

Figure 1: Impact of temperature on teacher presence



Notes: The regression is done using OLS with day of the week and district-by-month fixed effects. The standard errors are clustered on school-season. X-axis corresponds to temperatures. Y-axis corresponds to our coefficient of interest. Bin 20°C – 25°C is the omitted category.

6. Extension of the main results

We extend our main results to take into consideration several school's and teacher's characteristics. We focus on teacher gender, and different elements of school infrastructure in order to make some potential recommendations.

Male vs. Female

In Table 10, we investigate whether males or females are more impacted by higher temperatures than the other gender. It appears that women feel more often uncomfortably hot or uncomfortably cold than men (Karjalainen, 2007). This may suggest that females might try to avoid working under especially hot temperatures, while males could be less sensitive to heat. Table 10 shows that the gender of the teacher has no statistically significant incidence on their attendance. We regress using

using a gender control variable in column (2). The gender coefficient is statistically insignificant at the 10% level, suggesting that the gender of the individual plays no role in the decision of presence. Column (3) is regressed adding both control and interaction terms. The coefficient for the interaction term is negative, suggesting that being a female lowers the probability of attendance when the temperature gets higher. When regressing separately for males and females, we find in columns (4) and (5) that the size of the coefficient is slightly bigger for females compared to males, but the difference is not substantial. This table suggests that there is no clear effect due to gender on attendance. The only potential difference would be that females suffer slightly more from heat than males, but the effect is compensated by their overall higher presence rate.

Table 10: Gender

	(1) Preferred	(2) Control	(3) Control & interaction	(4) Males	(5) Females
Temperature	-0.0121***	-0.0122***	-0.0106***	-0.0107***	-0.0118***
	[0.0017]	[0.0017]	[0.0018]	[0.0018]	[0.0018]
Female*Temperature	-	-	-0.0026**	-	-
			[0.0012]		
Female	-	-0.0046	0.0841**	-	-
		[0.0060]	[0.0397]		
Observation	20,044	20,044	20,044	8,723	13,581

Notes: Regressions are done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. In column (2), we re-estimate the preferred specification, adding a female dummy variable. Column (3) adds both a control term and an interaction term to our preferred specification. Column (4) is regressed using only male teachers observations while column (5) is regressed using only female teachers observations.
* significant at 10% ** significant at 5% *** significant at 1%.

Distance to home

The distance the teacher has to travel from home to school could be an important cause of absenteeism; as we can imagine, it can be easier to walk 2km under 40°C than 15km. Table 11 shows that distance does indeed have a statistically significant negative impact on attendance. We put a threshold at 3km. Column (2) shows that living more than 3km away from school decreases the probability of being present at school. We get from columns (4) and (5) that a 1°C increase in temperature impacts potentially more the attendance of teachers living at more than 3km from school. This result could suggest that schools should potentially hire teachers living in a small perimeter around school in order to tackle absenteeism.

Table 11: Distance

	(1) Preferred	(2) Control	(3) Control & interaction	(4) ≤ 3km	(5) > 3km
Temperature	-0.0121***	-0.0122***	-0.0113***	-0.0102***	-0.0144***
	[0.0017]	[0.0016]	[0.0018]	[0.0020]	[0.0027]
Temp* Distance > 3km	-	-	-0.0024	-	-
			[0.0015]		
Distance > 3km	-	-0.0285***	0.0522	-	-
		[0.0069]	[0.0494]		
Observations	20,044	20,044	20,044	12,420	7,624

Notes: Regressions are done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Column (2) adds an distance between home and school superior to 3km control variable. Column (3) adds both an interaction term and a control variable. Column (4) regresses the preferred specification on a subsample of observations with a distance to home inferior or equal to 3km, while column (5) regresses the preferred specification on a subsample of observations with a distance to home superior to 3km.
* significant at 10% ** significant at 5% *** significant at 1%.

Presence of fan

There is evidence that the presence of a cooling system in the workplace reduces the negative effects of heat on output (Somanathan et al., 2015). Table 12 (except for column (4)) suggests that schools with fans have higher attendance, but heat affects similarly schools with and without fans. Column (3) shows that having a fan almost completely offsets the negative impact on teacher presence caused by a 4°C rise. In column (4), neither of the control and interaction terms are statistically significant. When regressing separately observations with and without a fan in the classroom, we find in columns (5) and (6) that the impact of heat on attendance if 1.5 times stronger if the teacher has to give class in a school without fan, compared to schools with fans.

Table 12: Presence of fan

	(1) Preferred	(2) Control	(3) Control & interaction	(4) Fan	(5) No fan
Temperature	-0.0121***	-0.0124***	-0.0130***	-0.0098***	-0.0158***
	[0.0017]	[0.0017]	[0.0021]	[0.0021]	[0.0028]
Temperature*Fan	-	-	0.0009	-	-
			[0.0020]		
Fan	-	0.0457***	0.0131	-	-
		[0.0105]	[0.0653]		
Observations	20,044	20,017	20,017	13,870	6,147

Notes: Regressions are done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Column (2) re-estimates the preferred specification adding a dummy variable indicating the presence of fan in school. Column (3) runs the preferred specification both with control and interaction terms. Columns (4) and (5) run the preferred specification respectively on subsamples of schools with fans and schools with no fan.
* significant at 10% ** significant at 5% *** significant at 1%.

Private vs. Public school

Kremer et al. (2005) find that attendance is slightly higher in private schools compared to government run schools. A potential explanation to their result could be that private school teachers have a larger risk of being dismissed for repeated absence. Column (2) from Table 13 reveals that private schools have a strong positive impact on teachers attendance; these schools might have a lower tolerance for absenteeism, or have more money to invest in infrastructure in order to offset the negative effects caused by heat. It appears from columns (2) and (3) that private schools have higher attendance, but heat affects more teachers in private schools. Columns (4) and (5) tend to confirm this increased sensitivity to heat in private schools, even though the difference is minor. This general result agrees with the findings of Kremer et al. (2005).

Table 13: Private schools vs. public schools

	(1) Preferred	(2) Control	(3) Control & interaction	(4) Private	(5) Public
Temperature	-0.0121***	-0.0121***	-0.0098***	-0.0124***	-0.0109***
	[0.0017]	[0.0017]	[0.0019]	[0.0023]	[0.0023]
Temperature*Private	-	-	-0.0039**	-	-
			[0.0018]		
Private	-	0.0424***	0.1743***	-	-
		[0.0083]	[0.0589]		
Observations	20,044	20,044	20,044	11,934	8,110

Notes: Regressions are done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Column (2) re-estimates the preferred specification adding a private school interaction variable. Column (3) regresses the preferred specification both with control and interaction terms. Columns (4) and (5) run the preferred specification respectively on subsamples of private schools and public schools.
* significant at 10% ** significant at 5% *** significant at 1%.

Indoor vs. Outdoor classrooms

If the teacher has to give class outside in the heat, he might not have the same incentive to come compared to a situation in which he could teach inside where the air might possibly be fresher. On the one hand, column (2) and (3) from table 14 show no statistically significant impact of indoor teaching. On the other hand, columns (4) and (5) show that the size of our main coefficient is drastically different when running the preferred specification separately on indoor and outdoor teaching subsamples. The size of heat impact on attendance is twice as large when class is given outdoors, compared to schools where class is given indoors. This result could potentially be explained by the inexistence of any outdoor cooling system, which negatively affects attendance (see above). It is also reasonable to suggest that schools with no indoor classrooms could have less infrastructure in general, and less money to deal with heat issues.

Table 14: Indoor classrooms

	(1) Preferred	(2) Control	(3) Control & interaction	(4) Indoors	(5) Outdoors
Temperature	-0.0121***	-0.0101***	-0.0133***	-0.0102***	-0.0201***
	[0.0017]	[0.0016]	[0.0034]	[0.0017]	[0.0079]
Temperature*Indoors	-	-	0.0035	-	-
			[0.0034]		
Indoors	-	0.0218	-0.0918	-	-
		[0.0160]	[0.1032]		
Observations	20,044	19,824	19,824	18,284	1,540

Notes: Regressions are done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Column (2) re-estimates the preferred specification adding a indoor classrooms interaction variable. Column (3) adds both control and interaction terms to the preferred specification. Column (4) runs the preferred specification on a subsample of schools where teaching is done indoors, while column (5) is run on a subsample of schools where teaching is done outdoors.
* significant at 10% ** significant at 5% *** significant at 1%.

Access to electricity

The school needs access to electricity for a potential cooling system to be able to run. It appears from column (2) from Table 15 that access to electricity has a statistically positive effect on attendance. On the other hand, regressing with both control and interaction terms in column (3) shows no statistically significant impact of access to electricity. We regress the preferred specification on subsamples in columns (4) to (6). The maximum hours of access to electricity in our sample is 8 hours, probably because they only care for electricity during class time. It seems that the subsample of schools having one to four hours of access to electricity has the weakest relationship between heat and attendance, followed by schools with no access to electricity at all. It is worth noting these two coefficients are respectively only statistically significant at 10% and 5%.

Table 15: Access to electricity

	(1) Preferred	(2) Interaction	(3) Control & interaction	(4) No electricity	(5) 1 – 4 hours	(6) 5 – 8 hours
Temperature	-0.0121***	-0.0135***	-0.0115***	-0.0081**	-0.0057*	-0.0114***
	[0.0017]	[0.0017]	[0.0027]	[0.0039]	[0.0031]	[0.0023]
Electricity	-	-	0.0793	-	-	-
			[0.0838]			
Temperature* Electricity	-	0.0019***	-0.0004	-	-	-
		[0.0004]	[0.0026]			
Observations	20,044	20,044	20,044	2,477	6,174	11,393

Notes: Regressions are done using OLS with district-by-month and day of the week fixed effects. Standard errors are clustered on school-season in brackets. Column (1) is the preferred specification. Column (2) re-estimates the preferred estimation, adding a maximum temperature times access to electricity interaction term. Column (3) runs the preferred specification both with a control and an interaction term. Column (4) runs the preferred specification on a subsample of schools with no access to electricity. Column (5) runs the preferred specification on a subsample of schools with 1 to 4 hours of access to electricity while column (6) runs the regression on a subsample of schools with 5 to 8 hours of access to electricity.
* significant at 10% ** significant at 5% *** significant at 1%.

7. Concluding remarks

High temperatures in India are currently affecting the country's economic development, and in a global warming time, the situation is going to be even harder to improve. Several studies already show the negative effects provoked by heat on the labour force. We study the impact of high temperatures on teacher absenteeism in India. This question is central as teachers are the first source of transmission of any human capital. Teacher absence in India remains quite high, despite all the investments made by the government during the past decade. We contribute to the literature by determining the impact of temperature and other potential weather factors on this major issue affecting the human capital. We use interviews developed by the IHDS covering 24,459 teachers in 4,085 Indian primary schools in 34 states. Each interview reports important school features, staff characteristics, and whether teachers are present at the time of the interview. We match each school to weather variables using the date of the interview and latitude and longitude of both schools and weather stations.

We find an inverted U-shaped relationship between maximum temperature and teacher attendance, with a peak between 20°C and 25°C, which is probably considered as the most comfortable temperature for work. When temperatures are higher than 25°C, our results show that a 1°C increase in maximum temperature leads to a rise of 8% to 9.4% in teacher absence, depending on the regressing method used. Because the relationship is nonlinear, the higher the temperature, the bigger the impact of a 1°C increase. This main result is robust to several other specifications. It also appears that moving averages of temperatures up to a month preceding the day of the interview have an impact on attendance, and it seems that teachers' metabolism hardly adapts to heat. Lagged values of precipitations also have an effect on teacher absence. Rain is an important factor to take into account as it can transport diseases such as diarrhea, and make mud roads harder to walk, especially since we also find that teacher living at more than 3km from school tend to be more absent. This last result should be an incentive for India to invest in better roads and in water sanity.

The extension of our main results shows strong positive correlations between better infrastructure, such as the presence of fans, indoor classrooms, or access to electricity, and teacher attendance. This result suggests that the government should better invest in this kind of infrastructure. It is crucial for India to start investing in school infrastructure in order to offset the negative effects caused by rising temperatures, especially in a context of climate change. Investing in teachers' comfort is investing in the future.

References

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., and Shapiro, J.S. (2012). Adapting to climate change: the remarkable decline in the US temperature-mobility relationship over the twentieth century. *Journal of Political Economy*, 124(1):105-159.

Barrow, M.W., and Clark, K.A. 1998. Heat-related illnesses. *American Family Physician*, 58(3):749-756.

Carlton, E. J., Cevallos, W., Eisenberg, J. N., Goldstick, J., Levy, K., and Trostle, J. (2014). Heavy rainfall events and diarrhea incidence: the role of social and environmental factors. *American Journal of Epidemiology*, 179(3):344-352.

Chen, X., and Yang, L. (2017). Temperature and industrial output: firm-level evidence from China. *Journal of Environmental Economics and Management*, forthcoming.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007). Are teacher absences worth worrying about in the US? *Education Finance and Policy*, 4(2):115-149.

Costa, H., Floater, G., Hooyberghs, H., Verbeke, S., De Ridder, K. 2016. Climate change, heat stress and labour productivity: A cost methodology for city economies. *Grantham Research Institute on Climate Change and the Environment*, Working Paper 248.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does pollution increase school absences? *The Review of Economics and Statistics*, 91(4):682-694.

Das, J., Holla, A., Mohpal, A., and Muralidharan, K. (2017). The social cost of weak governance: evidence from teachers absence in India. *Journal of Public Economics*, 145:116-135.

Dell, M., Jones, B. F., and Olken, B. A. (2012). Temperature shocks and economic growth: evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3):66-95.

Deryugina, T., and Hsiang, S. M. (2014). Does the environment still matter? Daily temperature and income in the United States. *National Bureau of Economic Research*, working paper 20750.

Deschênes, O., and Moretti, E. (2009). Extreme weather events, mortality, and migration. *The Review of Economics and Statistics*, 91(4):659-681.

Duflo, E., Hanna, R., and Ryan, S.P. (2012). Incentives work: Getting teachers to come to school. *American Economics Review*, 102(4):1241-1278.

Fisk, W. J., Seppänen, O., and Lei, Q.H. (2006). Effect of temperature on task performance in office environment. Ernest Orlando Lawrence Berkeley National Laboratory.

Garg, T., Jagnani, M., and Taraz, V. (2018). Temperature and human capital in India.

Graff Zivin, J., and Neidell, M. (2014). Temperature and the Allocation of Time: Implications for Climate Change. *Journal of Labor Economics*, 32(1):1-26.

Heal, G., and Park, J. (2016). Temperatures stress and the direct impact of climate change: a review of an emerging literature. *Review of Environmental Economics and Policy*, 10(2):1-17.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences of the United States of America,* 107(35):15367-72.

Karjalainen, S. (2007). Gender differences in thermal comfort and use of thermostats in everyday thermal environments. *Building and Environment,* 42(4):1594-1603.

Kremer, M., Chaudhury, N., Rogers, F. H., Muralidharan, K., and Hammer, J. (2005). Teacher absence in India: a snapshot. *Journal of the European Economic Association*, 3(3):658-667.

Miguel, E., Satyanath, S., and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4):725-753.

Miller, R. T., Murnane, R.J., and Willett, J.B. (2008). Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational Evaluation and Policy Analysis,* 30(2):181-200.

Somanathan, E., Sudarshan, A., and Tewari, M. (2015). The impact of temperature on productivity and labor supply: evidence from Indian manufacturing. Centre for Development Economics, Delhi School of Economics, working paper 244.