The Unequal Effect of Temperature on Test Scores:

Evidence from Colombia

Bridget Hoffmann
Xiomara Pulido
Diego Vera-Cossio

Department of Research and Chief Economist

DISCUSSION PAPER Nº
IDB-DP-1000

May 2023

The Unequal Effect of Temperature on Test Scores:

Evidence from Colombia

Bridget Hoffmann*
Xiomara Pulido**
Diego Vera-Cossio*

* Inter-American Development Bank
** University of British Columbia

May 2023

Inter-American Development Bank

Abstract[1]

We investigate the impact of heat shocks on high-stakes test scores across Colombia. We show that exposure to extreme heat in the week leading up to the exam reduces test scores. The negative effect is larger and in many cases concentrated among students with characteristics indicating low-income, indicating one channel through which climate change will exacerbate inequality.

JEL Classifications: Q54, I24, Q56, O15
Keywords: Temperature, Environmental shocks, Test scores, Inequality

[1]Bridget Hoffmann: Inter-American Development Bank. Xiomara Pulido: University of British Columbia. Diego Vera-Cossio: Inter-American Development Bank. All errors are our own. We received funding from an IDB Economic and Sector Works. Declarations of interest: none. Opinions, findings, conclusions, and recommendations expressed here are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank.

I Introduction

Human capital is an important driver of social mobility. Therefore, to understand how climate change will affect inequality, it is crucial to understand the effect of temperature on human capital and how this effect differs across the income distribution. A large literature has documented a negative relationship between temperature and test scores, which both serve as a measure of accumulated human capital and determine access to higher education, allowing further human capital accumulation.

Through its effect on extreme temperature shocks, climate change could exacerbate inequality in human capital through two channels. First, lower-income students could be more exposed to extreme temperatures. Second, extreme temperatures could have unequal effects on test scores, with more disadvantaged students suffering larger negative impacts.

We exploit variation in students' exposure to extreme temperature over time within a municipality to identify the effect of local temperature on SABER 11 test scores in Colombia. SABER 11 is a high-stakes, mandatory high school exit exam that is also a determinant of entry to post-secondary educational institutions. We find that absolute measures of extreme heat in the week leading up to the exam have a negative effect on test scores. The effect is consistent across various thresholds of extreme heat. Similarly, we find that hours in the week leading up to the exam that are unusually hot relative to the municipality-specific distribution of temperature have a negative effect on test scores.

Next, we look at heterogeneity by student characteristics that proxy for income level. We find that the negative effects of heat are typically larger for students with characteristics associated with lower-income, such as attending a public school or a rural school. Similarly, owning a PC and having an internet connection mitigate the negative effect of high temperature on test scores, indicating that wealth or the educational inputs that it can buy mitigate the effect of short-term temperature shocks on test scores. These results provide suggestive evidence that climate change, which will increase the occurrence of extreme temperatures, will exacerbate inequalities in educational outcomes in a country that is already characterized by low social mobility.

A large literature documents a casual effect of temperature on test scores, including high-stakes test scores and test scores in developing countries (Graff Zivin et al., 2018, 2020; Park et al., 2020; Park, 2020; Li and Patel, 2021; Melo and Suzuki, 2023; Vu, 2022). We seek to extend the literature beyond average effects to document one channel through which climate change will exacerbate human capital inequalities.

II Data

We use data from the SABER 11 standardized exam from the second semester of 2009 to the second semester of 2019 as a measure of approximately 5.5 million students' cognitive skills. SABER 11 is a mandatory, high-stakes exam that Colombian students take at the end of high school. SABER 11 is similar to the SAT in the United States in that it evaluates students' general knowledge across subjects and is essential for access to higher education. Many Colombian universities base their admissions on students' SABER 11 scores, and others implement their own admissions exams while requiring completion of SABER 11 as an enrollment requirement (Busso et al., 2020). However, unlike the SAT, taking the SABER 11 test is a graduation requirement and around 90% of students in the last grade of high school take it (Londoño-Velez et al., 2021)[2]. All students across Colombia take SABER 11 on the same day (within a semester-year) and complete the test sections in the same order. SABER 11 is graded centrally, so variations in local temperature should impact students' test scores but not the scoring of the test (i.e., the performance of the test graders).

The exam includes 50 math, 50 social science, 58 natural science, 41 reading, and 55 English questions. The exam includes both questions to test basic knowledge and questions to test higher-level problem-solving, interpretation,

[2]We restrict our sample to test takers between the ages of 11 and 25, with students aged 15 and over making up 99.8% of the sample.

and analysis. The final combined score, which ranges from 0 to 500, is called the global score. We standardized the global score (mean zero and variance one) by year and semester to make the test scores comparable over time.

In addition, as part of SABER 11, students complete a socioeconomic questionnaire that includes information on their mother's education, whether the student is enrolled in a public or private school, whether the school is in a rural or urban area, whether the student attends full-day, morning, afternoon or weekend school, whether the school is calendar A or B, whether the student took the exam in the first or second period of the year, whether the student has a PC and whether they have access to the internet.[3] Additional details about the SABER 11 data set that we use are included in the Data Appendix.

We merge the SABER 11 data with weather data from ERA5 at the municipality level. ERA5 is a gridded reanalysis data set (latitude-longitude grid of 0.25 x0.25 degrees) of hourly humidity and temperature. We calculate the hourly municipality-level weather series from 2001-2019 as the area weighted average of the temperature and humidity of grids that overlap with the administrative boundaries. We use the hourly municipality-level data series to create variables representing students' short-term (6 days prior to the exam and the day of the exam) temperature exposure in the week leading up to the SABER 11 exam. First, we create the mean temperature over the week leading up to the exam. Second, to focus on temperature extremes, we create a set of variables capturing the number of hours that the temperature in the student's municipality exceeds 32 degrees, 33 degrees, 34 degrees, and 35 degrees Celsius in the week leading up to the exam. Figure 1 shows the mean number of hours students are exposed to above 32 degrees Celsius in the week leading up to the exam by semester-year.[4] Third, we create variables capturing the

[3]In Colombia, schools choose between two calendars. In calendar A, students start school in February and finish in November. The B calendar begins in September and ends in June, similar to the school year in the United States. Most Colombian students study in schools with calendar A and take the Saber 11 test in the second semester of the year, while students in calendar B take the Saber 11 test in the first semester of the year. In our data, 97% of students are in Calendar A.

[4]Most students take the exam in the second semester. The systematic difference in the mean number of hours above 32C in the week leading up to the exam between the first and

number of hours that the temperature is below the 5th percentile, below the 10th percentile, above the 90th percentile, and above 95th percentile of the municipality's 2001-2019 temperature distribution in the week leading up to the exam.

Colombia is crossed by the Andes mountains. Therefore, despite Colombia's equatorial location, there is a wide range of temperatures across the country's regions due to its mountainous topography. Between 2009 and 2019, the years in our SABER 11 sample, the average temperature for the 1,120 municipalities was 20.8 degrees Celsius, with a standard deviation of 5.03 degrees. The minimum average temperature in our sample was 11.15 degrees Celsius, which occurred in the municipality of Betéitiva in 2012, and the maximum average temperature in our sample was 30.4 degrees Celsius, which occurred in the municipality of El Paso in 2015.

III Empirical Strategy

Our empirical strategy exploits variation in students' exposure to temperature over time within a municipality to estimate the effect of temperature on test scores. Specifically, we use regressions with semester-year and municipality fixed effects and a rich set of student-level controls to identify the impact of short-term temperature exposure on the standardized test score. First, we estimate the following specification that documents the effect of short-term weather on test scores.

$$Y_{imt} = \alpha + \beta_1 temp_{mt} + \rho X_{imt} + \gamma_t + \delta_m + \epsilon_{imt} \tag{1}$$

where Y_{imt} is the global score for student i in municipality m in semester-year t. $temp_{mt}$ is the mean temperature in municipality m in the week leading up to the exam in semester-year t. X_{imt} is a vector of student-level controls including age, mother's education level (indicator for high school or less), an indicator for attends a rural school, an indicator for the school calendar (A or

second semesters reflects the different locations of students taking the exam in semester 1 and 2, which is correlated with the school calendar.

B), an indicator for public school, school hours (morning, evening, night, Saturday and unique time), an indicator of household strata (strata 1-2 or strata 3+), an indicator for owning a PC, and an indicator for internet connection at home. γ_t is a semester-year fixed effect, and δ_m is a municipality fixed effect. ϵ_{imt} are standard errors clustered at the municipality level.

Second, we estimate the following specification that focuses on extreme heat.

$$Y_{imt} = \alpha + \beta_1 Above32C_{mt} + \rho X_{imt} + \gamma_t + \delta_m + \epsilon_{imt} \tag{2}$$

All variables are defined as in equation 1 except that we replace $temp_{mt}$ with $Above32C_{mt}$. $Above32C_{mt}$ is the number of hours that the municipality m's temperature was above 32 degrees Celsius in the week leading up to the exam taken in semester-year t. We estimate this specification using variables that capture the number of hours in the week leading up to the exam that are above 32 degrees Celsius, 33 degrees Celsius, 34 degrees Celsius, and 35 degrees Celsius.

Third, we estimate the following specification that focuses on the effect of unusually cold and hot temperatures.

$$Y_{imt} = \alpha + \beta_1 Above95_{mt} + \beta_2 Below5_{mt} + \rho X_{imt} + \gamma_t + \delta_m + \epsilon_{imt} \tag{3}$$

Equation 3 replaces $Above32C_{mt}$ with two variables capturing unusually cold and hot hours in the week leading up to the exam relative to the municipality's temperature distribution. $Above95_{mt}$ is the number of hours in the week before the exam that were in the 95th percentile or above and $Below5_{mt}$ is the number of hours that were in the 5th percentile or below for municipality m and semester-year t. We estimate the same specification for the number of hours in the week before the exam that were in the 90th percentile or above and the number of hours that were in the 10th percentile or below.

We estimate heterogeneous effects using the following two specifications

$$Y_{imt} = \alpha + \beta_1 Above32C_{mt} + \theta_1 Above32C_{mt} * C_{imt} + \rho X_{imt} + \gamma_t + \delta_m + \epsilon_{imt} \tag{4}$$

$$Y_{imt} = \alpha + \beta_1 Above90_{mt} + \beta_2 Below10_{mt} + \theta_1 Above90_{mt} * C_{imt} + \theta_2 Below10_{mt} * C_{imt} + \rho X_{imt} + \gamma_t + \delta_m + \epsilon_{imt} \tag{5}$$

where all variables are defined as in 2 and 3. The estimators θ_1 and θ_2 capture the differential effect of extreme temperatures by a student characteristic, included in X_{imt}, that could proxy for students' income level.

IV Results

First, we investigate the effect of temperature on test scores. To begin, we explore the effect of mean temperature over the week leading up to the exam on test scores. Panel A of table 1 displays the results of estimating equation 1. Column (1) shows that there is no significant effect of mean temperature of the week leading up to the exam on the global test score. Next, we focus on investigating the effect of high temperatures. Panel B presents the results of estimating equation 2. Column (2)-(5) shows that there is a significant negative effect of high temperature on test scores and that the magnitude of this effect is increasing with the intensity of the heat. For instance, an additional 10 hours above 32 degrees Celsius in the week leading up to the exam reduces test scores by 0.01 standard deviations and an additional 10 hours above 35 degrees Celsius in the week leading up to the exam reduces test scores by 0.03 standard deviations. Finally, we focus on unusual temperatures. Panel C presents the results of estimating equation 3. Column (6) shows that there is a negative but insignificant effect of hours above the 95th percentile of the temperature distribution, and column (7) shows that there is a negative, significant effect of hours above the 90th percentile of the temperature distribution. Overall, we find a robust negative effect of high temperature on the global test score.

Second, we investigate whether short-term temperature shocks have a larger impact for low-income students. We are able to provide only suggestive evidence because we look for heterogeneity along dimensions that proxy for lower-income. Panel A of table 2 shows the results of estimating equation 4, and panel B of table 2 shows the results of estimating equation 5. Column (1) shows heterogeneous results by attending a rural school. In panel A, we find no significant heterogeneity in the effect of hours above 32 degrees Celsius for students who attends rural schools. In panel B, we find no effect of hours above the 90th percentile or below the 10th percentile for students who do not attend rural schools. In contrast, relative to students who do not attend rural schools, we find significant negative effects of hours above the 90th percentile and hours below the 10th percentile for students who attend rural schools. We investigate heterogeneity by attending a public vs. private school in column (2). In panel A, we find no effect of hours above 32 degrees Celsius for student attending private schools. However, relative to students who attend private schools, we find a significant negative effect of hours above 32 degrees Celsius for students attending public schools. In panel B, we find no significant heterogeneity in the effect of hours that are unusually hot whether a student attends a public school. In column (3), we find no significant heterogeneity in the effect of extreme temperature by the education level of a students' mother using either specification (panel A and panel B). In columns (4) and (5), we explore heterogeneity by ownership of a personal computer (PC) and an internet connection. In panel A, we find no heterogeneous effects by ownership of a PC or internet connection for hours above 32 degrees Celsius. However, in panel B, we find that ownership of a PC and ownership of an internet connection reduce the magnitude of the negative impact of hours that are unusually hot. Considered together, the results suggest that the negative impact of extreme temperature on test scores is larger in magnitude for low-income students.

V Conclusions

We find that extreme temperature reduces high-stakes test scores in Colombia. In particular, we find negative effects of extreme heat and unusual heat in the week leading up to the test. We provide suggestive evidence that these effects may be larger for lower-income students. Together, these preliminary results suggest that socioeconomic inequalities are reflected in human capital inequalities. These preliminary results indicate that climate change, which will increase the occurrence of extreme temperatures, will exacerbate inequalities in educational outcomes in a country that is already characterized by low social mobility.

References

Graff Zivin, J., S. Hsiang, and M. Neidell (2018). Temperature and human capital in the short and long run. *JAERE 5*(1), 77–105.

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

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

Melo, A. P. and M. Suzuki (2023, March). Temperature, effort, and achievement: Evidence from a large-scale standardized exam in brazil. Unpublished.

Park, R. J. (2020). Hot temperature and high-stakes performance. *Journal of Human Resources 57*(2), 400–434.

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

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

Figures

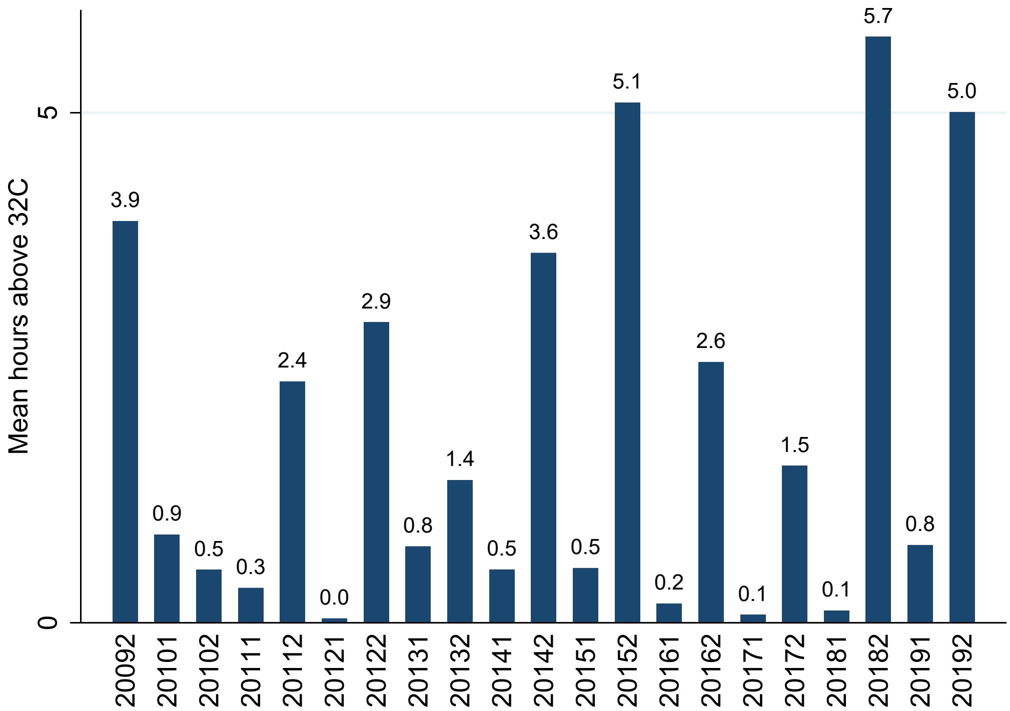

Figure 1: Mean Number of Hours Above 32 Degrees Celsius by Semester-Year

Note: The figure reports the mean number of hours above 32 degrees Celsius by semester-year.

Tables

Table 1: Effects of Temperature on SABER 11 Score

Variables	(1) Global z-score	(2) Global z-score	(3) Global z-score	(4) Global z-score	(5) Global z-score	(6) Global z-score	(7) Global z-score
Panel A							
Mean temperature (7 days)	-0.0131						
	(0.0083)						
Panel B							
Above 32C		-0.0011***					
		(0.0002)					
Above 33C			-0.0016***				
			(0.0003)				
Above 34C				-0.0020***			
				(0.0003)			
Above 35C					-0.0035***		
					(0.0008)		
Panel C							
Below 5th						0.0004	
						(0.0004)	
Above 95th						-0.0005	
						(0.0004)	
Below 10th							0.0002
							(0.0003)
Above 90th							-0.0005*
							(0.0003)
Mean	20.07	3.018	1.572	0.700	0.291		
Mean Below						9.690	18.69
Mean Above						13.01	23.09
Controls							
FE municipality	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE year-term	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	5,514,840	5,514,840	5,514,840	5,514,840	5,514,840	5,514,840	5,514,840

Notes: In panel A, mean temperature is the average temperature for the 6 days before and the day of the test. In panel B, above 32C is the number of hours in the week before the test that were equal to 32C or above. Similarly for 33C, 34C, and 35C. In panel C, Below 5th percentile is the number of hours that are below the 5th percentile of the distributions of temperatures from 2001 to 2019 for each municipality. Above 95th percentile is the number of hours that are above the 95th percentile of the distributions of temperatures from 2001 to 2019 for each municipality. Similarly for Below 10th and Above 90th precentile. All columns include controls by age, mother's education (indicator for high school or less), rural school (indicator), calendar, public school (indicator) and school hours (morning, evening, night, Saturday and unique time), an indicator of household strata (strata 1-2 or strata 3+), own PC (indicator), and internet connection at home (indicator). All columns include semester-year and municipality fixed effects. Standard errors clustered at municipality level in parentheses. The row labeled mean displays the mean temperature or the mean number of hours above each temperature threshold in the week leading up to the exam. The rows labeled mean above and mean below show the mean number of hours above the 90th and 95th percentiles and below the 5th and 10th percentiles. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2: Heterogeneous Effects of Temperature on SABER 11 Score

Variables	(1) Global z-score	(2) Global z-score	(3) Global z-score	(4) Global z-score	(5) Global z-score
Panel A					
Above 32C	-0.0010***	0.0021	-0.0022**	-0.0011***	-0.0011***
	(0.0003)	(0.0015)	(0.0011)	(0.0003)	(0.0003)
Rural School	-0.1652***				
	(0.0150)				
Above 32C X Rural School	-0.0008				
	(0.0007)				
Public School		-0.0160			
		(0.0217)			
Above 32C X Public School		-0.0037**			
		(0.0016)			
Mom has high school or less			-0.4439***		
			(0.0089)		
Above 32C X Mom HS			0.0012		
			(0.0012)		
Own PC				0.1037***	
				(0.0029)	
Above 32C X Own PC				-0.0001	
				(0.0007)	
Own Internet					0.0973***
					(0.0043)
Above 32C X Own Internet					0.0000
					(0.0008)
Mean	0.134	0.732	0.807	0.570	0.494
Mean Above 32	3.018	3.018	3.018	3.018	3.018
Panel B					
Below 10th	0.0004	-0.0008	0.0003	-0.0000	0.0001
	(0.0004)	(0.0005)	(0.0007)	(0.0003)	(0.0003)
Above 90th	-0.0003	-0.0006	-0.0008*	-0.0008**	-0.0008***
	(0.0003)	(0.0004)	(0.0005)	(0.0003)	(0.0003)
Rural School	-0.1065***				
	(0.0231)				
Below 10th X Rural school	-0.0016**				
	(0.0007)				
Above 90th X Rural school	-0.0015***				
	(0.0005)				
Public School		-0.0524*			
		(0.0292)			
Below 10th X Public school		0.0014*			
		(0.0008)			
Above 90th X Public school		0.0001			
		(0.0006)			
Mom has high school or less			-0.4493***		
			(0.0262)		
Below 10th X Mom HS or less			-0.0001		
			(0.0006)		
Above 90th X Mom HS or less			0.0004		
			(0.0007)		
Own PC				0.0825***	
				(0.0094)	
Below 10th X Own PC				0.0004	
				(0.0003)	
Above 90th X Own PC				0.0005**	
				(0.0002)	
Own Internet					0.0782***
					(0.0110)
Below 10th X Own Internet					0.0002
					(0.0004)
Above 90th X Own Internet					0.0007***
					(0.0002)
Mean	0.134	0.732	0.807	0.570	0.494
Mean Below	18.69	18.69	18.69	18.69	18.69
Mean Above	23.09	23.09	23.09	23.09	23.09
Controls					
FE municipality	Yes	Yes	Yes	Yes	Yes
FE year-term	Yes	Yes	Yes	Yes	Yes
Obs	5,514,840	5,514,840	5,514,840	5,514,840	5,514,840

Notes: In panel A, above 32C is the number of hours in the week before the test that were equal to 32C or above. In panel B, Below 10th percentile is the number of hours that are below the 10th percentile of the distributions of temperatures from 2001 to 2019 for each municipality. Above 90th percentile is the number of hours that are above the 90th percentile of the distributions of temperatures from 2001 to 2019 for each municipality. All columns include controls by age, mother's education (indicator high school or less), rural school (indicator), calendar, public school (indicator) and school hours (morning, evening, night, Saturday and unique time), an indicator of household strata (strata 1-2 or strata 3+), own PC (indicator), internet connection at home (indicator). Standard errors clustered at the municipality level are in parentheses. The row labeled mean displays the mean of the variable by which we approximate low income. In panel A, the row labeled mean Above 32 displays the mean number of hours above 32C. In panel B, the rows labeled mean above and mean below show the mean number of hours above the 90th and below the 10th percentiles. *** p<0.01, ** p<0.05, * p<0.1

Data Appendix

In order to access the information for Saber 11 from 2009-2 to 2019-2, the reader must follow the next steps:

1. Go to https://www.icfes.gov.co/web/guest/data-icfes
2. Click on "Register to DataIcfes."
3. Fill out the form requiring information such as name, email address, type of user (researcher, student, etc.), the organization of affiliation, the name of the organization, etc.
4. Wait for an email that indicates the link where the data can be downloaded and the password. The email is automatically and immediately sent after registration.
5. Click on "Access here to the data repository."
6. Introduce the password from registration.
7. Click on the folder "04. Saber11"
8. Click on the folder "3. Resultados Saber 11"
9. Click on each folder from 2009-2 to 2019-2 and select download.

After downloading all the available information, each data set is imported to Stata. The variables that we keep for the analysis from 2009-2 to 2014-1 are:

- `periodo`
- `estu_genero`

- `estu_estudiante`
- `estu_fechanacimiento`
- `cole_cod_mcpio_ubicacion`
- `cole_jornada`
- `cole_naturaleza`
- `fami_estratovivienda`
- `fami_educacionmadre`
- `fami_nivelsisben`
- `fami_tienecomputador`
- `fami_tieneinternet`
- `cole_area_ubicacion`
- `cole_jornada`
- `cole_naturaleza`

- `punt_lenguaje`
- `punt_ciencias_sociales`
- `punt_quimica`
- `punt_filosofia`
- `punt_biologia`
- `punt_fisica`
- `punt_ingles`
- `cole_calendario`

From 2014-2 to 2019-2 we keep:

- `periodo`
- `estu_genero`
- `estu_estudiante`
- `estu_fechanacimiento`

- cole_cod_mcpio_ubicacion
- cole_jornada
- cole_naturaleza
- fami_estratovivienda
- fami_educacionmadre
- fami_nivelsisben
- fami_tienecomputador
- fami_tieneinternet
- cole_area_ubicacion
- cole_jornada
- cole_naturaleza
- punt_lectura_critica
- punt_matematicas

- `punt_c_naturales`
- `punt_sociales_ciudadanas`
- `punt_ingles cole_calendario`

Then, we calculate the global index score. From 2009-2 to 2014-1, the equation that the government used to calculate the global score was:

```
GI =  (3*matematicas + 3*lenguaje +2*ciencias_sociales
+ quimica + filosofia + biologia + fisica + ingles)/13
```

From 2014-2 to 2019-2, the equation that is used to calculate the global index score is:

```
GI =  (3*matematicas + 3*lectura_critica + 3*c_naturales +
3*sociales_ciudadanas + ingles)/13
```

The scores are standardized at the year-period level. Although the variables names can vary, we harmonized all years to create a repeated panel at the individual level. The Sisben group is not included in the regressions because the variable disappeared in 2016-2. However, we use as a proxy the variable strata, a categorical variable that goes from 1 to 6; the first category, 1, indicates the most vulnerable students in the sample, and the last category, 6, marks the wealthier students. This variable is extensively used in the country for charging differential tariffs on the cost of public services.