EdWorkingPaper No. 24-910

Under the Weather? The Effects of Temperature on Student Test Performance

Deven Carlson
University of Oklahoma

Adam Shepardson
University of Oklahoma

As students are exposed to extreme temperatures with ever-increasing frequency, it is important to understand how such exposure affects student learning. In this paper we draw upon detailed student achievement data, combined with high-resolution weather records, to paint a clear portrait of the effect of temperature on student learning across a six-year period for students in Tulsa, Oklahoma. The detailed, longitudinal nature of our data allows us to estimate the effects of both test-day and longer-term temperature on student test performance, and to examine how the effects of both temperature measures vary across seasons, student background, and the distribution of student achievement. Our results show that test-day temperature has no significant effect on student test performance in fall or winter, but a clear negative effect on students' spring performance, particularly in math. Second, we find that summer temperature has a positive, statistically significant, and substantively meaningful effect on student performance on the fall MAP assessment—these effects appear in both math and reading. The results also illustrate that 90-day temperature affects math performance in winter and spring, but these estimates are modest in substantive magnitude.

VERSION: February 2024

Suggested citation: Carlson, Deven, and Adam Shepardson. (2024). Under the Weather? The Effects of Temperature on Student Test Performance. (EdWorkingPaper: 24-910). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/p9q4-jb65

Under the Weather? The Effects of Temperature on Student Test Performance

Deven Carlson
Adam Shepardson

University of Oklahoma

February 2024

Abstract
As students are exposed to extreme temperatures with ever-increasing frequency, it is important to understand how such exposure affects student learning. In this paper we draw upon detailed student achievement data, combined with high-resolution weather records, to paint a clear portrait of the effect of temperature on student learning across a six-year period for students in Tulsa, Oklahoma. The detailed, longitudinal nature of our data allows us to estimate the effects of both test-day and longer-term temperature on student test performance, and to examine how the effects of both temperature measures vary across seasons, student background, and the distribution of student achievement. Our results show that test-day temperature has no significant effect on student test performance in fall or winter, but a clear negative effect on students' spring performance, particularly in math. Second, we find that summer temperature has a positive, statistically significant, and substantively meaningful effect on student performance on the fall MAP assessment—these effects appear in both math and reading. The results also illustrate that 90-day temperature affects math performance in winter and spring, but these estimates are modest in substantive magnitude.

Introduction

An ever-growing body of research documents that features of the natural environment can play a substantial role in shaping students' achievement and attainment outcomes. Recent work shows that students' educational outcomes are affected by exposure to wildfire smoke (Wen & Burke 2022), industrial pollution (Persico & Venator 2021), and vehicle emissions (Heissel, Persico, & Simon 2022; Austin, Heutel, & Kreisman 2019). Perhaps most important in this line of research, though, is the evidence of local weather conditions, particularly temperature, affecting educational outcomes (Park 2022; Park et al. 2020; Graff Zivin, Hsiang and Neidell 2018; Park et al. 2021). As climate change increases the frequency with which students are exposed to extreme temperatures (Ricke et al. 2018; Robinson et al. 2021), it becomes increasingly important to understand how such exposure affects student learning. And while some initial progress has been made on this topic, data limitations and unmeasured contextual differences have posed hurdles to gaining a nuanced understanding of the impacts of local weather conditions on student achievement.

In this paper we leverage detailed student achievement data combined with high-resolution weather records to paint a clear portrait of the effect of temperature on student learning across a six-year period for students in Tulsa, Oklahoma. Our anonymized student achievement data consists of scores on the math and reading portions of the MAP Growth test for students enrolled in Tulsa Public Schools (TPS) between the 2014-15 and 2019-20 school years. During this six-year period, TPS administered the MAP test three times per school year, once in the fall (i.e. September), a second time in the winter (i.e. January), and a third time in the spring (i.e. May). Thus, our data contain up to 18 observations for each student, with these observations spread across meteorological seasons. We combine these achievement data with weather measures derived from the Oklahoma Mesonet, a network of 120 environmental monitoring stations spread across Oklahoma. These stations measure and record a variety of weather conditions at five-minute intervals, with archived records available back to the mid-1990s (Brock et al. 1995; McPherson et al. 2007).

The detailed, longitudinal nature of our data allows us to conduct a wide range of empirical analyses that explore multiple dimensions of the effects of temperature on student learning. First, we estimate the effects of both test-day temperature and long-term temperature exposure for students in a single empirical context, defining the latter measure as the average temperature over the three-month period leading up to MAP administration. This allows us to distinguish the short- and long-term learning effects of temperature using a design that explicitly accounts for students' prior performance on the MAP assessment. Crucially, by estimating these effects for students within a single empirical context, we minimize the potential for unmeasured contextual factors, such as variation in climate control capabilities or differences in student acclimatization, to enter our estimates.

Second, we exploit the tri-annual administration of the MAP assessment to estimate both test-day and long-term temperature effects separately for the fall, winter, and spring administrations. In addition to providing direct estimates of seasonal heterogeneity in temperature effects, these analyses also shed light on the role of acclimatization as a moderator of student learning, particularly for test-day temperature. In Oklahoma, both May and September—the months of the respective spring and fall MAP administrations—often have exceptionally hot days, but students are much more accustomed to the heat in September than they are in May. Thus, we are uniquely able to assess whether the same temperature in different months differentially impacts student performance.

Third, the richness of our data allows us to explore several potential sources of heterogeneity in the effects of temperature. Motivated by the presence of heterogeneous temperature effects in prior work (Graff Zivin et al. 2020; Park et al, 2020; Vu 2022), we investigate the possibility of variation in temperature effects across students of different racial or ethnic backgrounds, by the median income of the zip code in which students reside, and for students at different points of the achievement distribution. Together, the results of our analyses paint a vivid, multi-dimensional picture of temperature's effect on student learning and, in doing so, propel our understanding of the topic forward.

To briefly summarize our main results, we first show that test-day temperature has no significant effect on student MAP performance in fall or winter, but a clear negative effect on students' spring MAP performance, particularly in math. Second, we find that summer temperature has a positive, statistically significant, and substantively meaningful effect on student performance on the fall MAP assessment—these effects appear in both math and reading. The results also illustrate that 90-day temperature affects math performance in winter and spring, but these estimates are modest in substantive magnitude. Finally, our analyses of heterogeneity show little meaningful variation by student race/ethnicity or the distribution of student achievement. There is some evidence of heterogeneity according to average zip code income and, when such variation exists, the impacts are larger among students residing in relatively less affluent areas than among students in zip codes with above-average income levels. In the concluding section we discuss the implications of these results for research, policy, and practice.

Weather and Student Achievement

Our work builds upon findings from a limited set of prior empirical studies examining the impacts of temperature, both test-day and longer-term, on student learning. The literature on test-day temperature reaches disparate conclusions, with studies showing that test-day heat reduces performance among students taking the New York Regents exam (Park 2022), Peabody math test (Graff Zivin, Hsiang and Neidell 2018), and college entrance exams in both China (Graff Zivin et al. 2020) and Vietnam (Vu 2022), but has no impact on the performance of students taking Brazil's college entrance exam (Li and Patel 2021) or reading and math tests in Australia (Johnston et al. 2021). Studies of longer-term temperatures exhibit a similar dynamic, with long-term heat exposure reducing performance on the Programme for International Student Assessment (PISA) (Park et al. 2021) and college entrance exams in South Korea (Cho 2017), but leaving student performance on the Peabody assessment unaffected (Graff Zivin, Hsiang and Neidell 2018). And analysis of Preliminary Scholastic Aptitude Test (PSAT) data shows student performance to be affected by temperatures during the school year, but not during the summer (Park et al. 2020).

On the surface, the empirical literature linking learning and temperature seems varied and conflicted. However, the studies comprising this literature estimate different parameters, span different meteorological seasons, and are set in disparate educational and geographic contexts. Understanding how these differences connect to the heterogeneity observed in the empirical literature requires engagement with a range of conceptual considerations that span the physiological and behavioral realms, as well as the built environment.

Physiologically, the literature on heat stress makes clear that prolonged exposure to an excessively hot environment disrupts core cognitive abilities (Taylor et al. 2016), including memory (Gaoua et al. 2011; Lee et al. 2015) and decision-making (Froom et al. 1993; Coehoorn et al. 2020). These physiological disruptions, which motivate most existing empirical work on temperature and learning, have the potential to inhibit students' knowledge retention over the course of their school year. Following the same logic, test-day heat could likewise impede students' knowledge retrieval and application during the period in which they sit for an exam. Importantly, though, individuals can become acclimated to environmental conditions after extended exposure (Périard, Racinais, and Sawka 2015; Castellani and Young 2016). Such acclimatization may, at least to some degree, insulate individuals in certain climates from weather impacting their learning, a dynamic that can produce geographic heterogeneity in the effect of temperature on student learning. A small amount of prior empirical work provides support for this phenomenon. For example, in South Korea, heat had a greater impact on the test scores of students in cities with relatively cool summers, compared to their peers residing in warmer locales (Cho 2017). Similarly, in the U.S. context, heat produced a disproportionately large reduction in performance on math and reading tests for students residing in locations with relatively cool average high temperatures (Roach and Whitney 2022). And recent work attributed the negative impacts of cold weather on Australian students' test performance to individuals' acclimatization to hot, but not cold, temperatures (Johnston et al. 2021). More generally, these findings illustrate the importance of considering how acclimatization, or its absence, might mediate the impact of temperature on student learning.

Along with physiological reactions, weather can also induce a range of behavioral responses. Perhaps most relevant are the studies demonstrating that temperature affects time allocation decisions (Graff Zivin and Neidell 2014; Heaney et al. 2019; Fan et al. 2023), with individuals least likely to engage in outdoor activities when the ambient temperature is very cold or very hot, and most likely to find such activities appealing on pleasantly warm days. The literature lacks systematic evidence detailing how weather affects the time-use behavior among elementary and secondary school students, but it is self-evident that such effects exist.[1] From

[1] Evidence from postsecondary contexts indicates that college students reduce study time on hot days and class attendance on cold days (Alberto et al. 2021).

snow days, to outdoor extracurricular activity cancellations, to district policies specifying the weather conditions that trigger indoor recess, there are myriad examples of how weather shapes the way that students spend their time. These factors underscore the importance of considering how temperature might affect student learning through the mechanism of time allocation.

Finally, the built environment interacts with weather conditions in ways that, theoretically, can both mitigate and exacerbate the impact of weather on student learning. In terms of mitigation, the vast majority of classrooms across the country have climate control capabilities—heat and air conditioning—and there is evidence that these capabilities can significantly reduce the effects of temperature on student performance (Park et al. 2020). Classrooms lacking air conditioning are mostly confined to the Northeast, the Upper Midwest, and the Pacific Coast. A similar pattern is evident for home air conditioning, with the bulk of homes in the South, Southwest, and Plains states having air conditioning while most dwellings in the Northeast, Appalachia, Upper Midwest, Mountain West, and Pacific Coast lack cooling capabilities (Park et al. 2020). Other features of the built environment also have the potential to amplify the impacts of weather conditions. For example, the large amount of pavement and roofing material in developed areas contribute to the Urban Heat Island effect (Mohajerani et al. 1997), a phenomenon in which urban areas exhibit temperatures several degrees higher than those observed in outlying areas (Oke 1982). On the flip side, tall buildings can amplify wind, resulting in lower wind chills and greater levels of discomfort during the winter (Soligo et al. 1998). Together, the combination of physiological, behavioral, and physical considerations form a strong foundation for expecting the relationship between weather and student learning to vary across geographic contexts.

Weather and Learning in Tulsa, Oklahoma

Tulsa, Oklahoma serves as the setting for our empirical inquiry into the effects of weather on student learning. Several features of the Tulsa context render it well-suited for this purpose. First, in contrast to most past studies on the topic, our focus on a single geographic context allows us to clearly characterize the area's weather and climate and systematically consider how the climate is likely to interact with the physiological, behavioral, and physical considerations described in the prior section. Second, Tulsa's weather is highly variable, both within and across seasons, which provides an ideal empirical context for analyzing the effects of both test-day and longer-term weather. Third, the school district that serves the Tulsa area, Tulsa Public Schools (TPS), has long administered benchmark assessments to its students at multiple points during the school year. Specifically, since 2014 TPS has administered the MAP Growth assessment, one of the most common formative assessments in school districts across the country. In TPS, assessment administrations occur in the fall, winter, and spring, a schedule that facilitates our analysis of seasonal heterogeneity in the impact of weather on student learning outcomes.

Tulsa's climate is characterized as moderate, with long, hot summers that tend to peak in late July or early August and generally mild winters. Summers average highs of 93-94 in July and August, with almost 75 days with a high above 90 and 11 days that hit the century mark. Winter is at its coldest in January and although it can get quite cold—the record low in Tulsa history is 16 degrees below zero—winters are generally mild. The average high in January is well above freezing at 49 degrees, and Tulsa typically gets less than 9 inches of snow annually. Spring and fall can be quite pleasant, but are characterized by their variability, with both very hot and cold temperatures possible in these seasons. Moreover, spring is often referred to as "tornado season" as April and May often bring severe thunderstorms and the associated hazards of tornados, hail, and damaging wind.

Tulsa's climatological profile interacts in important ways with the physiological, behavioral, and physical considerations discussed above. Physiologically, the excessive summer heat could have the potential to inhibit student cognition and knowledge acquisition and retention, but such an impact is mitigated by three major factors. First, and most basically, school is not in session for much of the summer, meaning that little in the way of formal learning is taking place for many students. Second, during the years spanning our analysis, all schools in TPS had renovated air conditioning systems (Davis 2012) and the vast majority of residences also had cooling capabilities. These features of the built environment are likely to minimize the impact of excessive heat on student learning (Park et al. 2020). Third, excessive heat during the summer may, on the margin, spur students to substitute indoor activities for outdoor ones. Given the relatively large amount of free time during the summer, it is possible that the indoor activities involve relatively more engagement with academically oriented content and knowledge. Together, these considerations support scenarios where summer heat could either boost or depress student learning—there is no clear theoretical prediction.

Beyond the summer, excessively hot temperatures are also possible during the fall and spring administrations, and we highlight two considerations relevant to these seasons. First, as noted above, all TPS schools and almost all student residences have air conditioning, a feature of the built environment that mitigates the impacts of heat. Second, even recognizing the importance of cooling capabilities, students are much more likely to be acclimated to heat during the fall MAP administration than the spring one. Thus, to the extent that temperature impacts manifest at all in these seasons, they are much more likely to occur in spring than in fall.

For winter MAP administrations, many of the same considerations outlined above are relevant, including climate control capabilities in schools and residences. Students' substitution of indoor activities for outdoor ones during periods of cold weather is also potentially pertinent, although perhaps less so than in summer. Because school is in session for most of the winter, students have relatively less discretionary time to make these time-use tradeoffs. Again, the various conceptual considerations do not point to a clear expectation—the question is ultimately an empirical one.

Data and Measures

We construct our analytic dataset with records from two main sources. First, we draw upon administrative records from TPS. These records contain a wide range of information on the universe of TPS' MAP administrations from the 2014-15 to 2019-20 school years.[2] Separately for reading and math, and structured in a student-by-administration format, the records contain an anonymized unique student identifier and detailed information about the student at the time of the administration, including an identifier for the school they attend, their sex, racial/ethnic identification, disability status, and zip code of their residence. The zip code information is particularly important as we use it to connect students to the weather conditions they experience. Notably, the use of student identifiers renders these data untraceable to any one individual, leaving the test takers themselves anonymous.

The records also contain information about the administration itself. In addition to the school year, the records indicate the season of administration (i.e. Fall, Winter, Spring), the date of the administration, students' scale score and percentile on the assessment, and the amount of time taking the test. Our empirical analyses use students' percentile score on the MAP assessment as the primary outcome variable. Together, the TPS records contain comprehensive information on all MAP administrations, including information on the student sitting for each

[2] For the 2014-15 through 2018-19 school years, TPS records contain information on the fall, winter, and spring administrations. For the 2019-20 school year, the records only contain information from the fall and winter administrations—the spring administration did not take place because of COVID.

administration, the circumstances of the administration, and student performance on the assessment.

Second, we take advantage of detailed weather records archived at the Oklahoma Mesonet. Established in the mid-1990s, the Mesonet is a network of 120 environmental monitoring stations spread across the entire state of Oklahoma, with each of the state's 77 counties containing at least one station. For Tulsa in particular, we draw upon data from 18 stations within 50 miles of the city. Each of these stations continually measures a variety of weather conditions each day, computing and recording observations at five-minute intervals that contain the five-minute averages for a variety of weather conditions, including temperature, wind speed, precipitation, dew point, and humidity, among dozens of other measurements. These five-minute observations are used to calculate a single daily value for each monitored weather condition. For example, the archived daily summary of a Mesonet station's "average air temperature" reports the mean of all five-minute averaged air temperature readings on a given date.

We use a four-step process that leverages these detailed, high-frequency Mesonet records, which are available back to the mid-1990s, to construct the weather measures in our analyses below. First, for each of the 18 stations within 50 miles of Tulsa, we obtained the archived daily summary records of average daily air temperature and daily maximum air temperature. We pulled these station-level measures for each day from January 1, 2000 through the last observed date of MAP administration in the TPS records. Second, we overlayed a grid of one-by-one kilometer cells on the land within a 50 mile radius of Tulsa's geographic centroid and use an inverse distance weighting[3] (IDW) technique that leverages the station-by-day measures to interpolate a value of each weather measure to each grid cell for each day. That is, we generate weather measures spanning the entire Tulsa metropolitan area with high spatial resolution. The left-hand panel of Figure 1 illustrates the location of Mesonet stations within 50 miles of Tulsa. It also situates the region within the broader geography of Oklahoma. The right-hand panel of the figure zooms in on that 50-mile radius, illustrating the one-by-one kilometer grid cells, the borders of zip codes containing at least one student residence, and the geographic centroid of each zip code.

[Insert Figure 1 about here]

Third, for each zip code containing at least one student residence, we extract the full set of interpolated daily weather measures for the grid cell containing the geographic centroid of the zip code. This extraction provides us with the flexibility to calculate zip code-level weather measures at any temporal scale. Fourth, we combine the set of student residential zip codes with information on each zip code's set of MAP testing dates, which we glean from the TPS records, and calculate the two main weather measures used in our analysis below: 1) Average test day temperature, and 2) The average daily temperature over the 91 days leading up to, and including, the day of test administration—we refer to this as our 90-day temperature measure.

[3] IDW interpolation predicts the unobserved value of a variable at a spatial location in inverse proportion to its distance from other, known values. The IDW estimate at any given location is, accordingly, a weighted average of surrounding observed values where spatially closer values exert more influence over the ultimate estimate than those farther away (Weber and Englund 1992). In practice, we leveraged the idw function from R package gstat (Pebesma 2004) in conjunction with other spatial data science tools (Pebesma and Bivand 2023) to interpolate zip code level weather predictions from archived Mesonet station data. IDW interpolation via gstat assigns weights as the absolute value of the difference between the desired prediction location (a grid cell) and a given observed value, with this difference raised to the power of –2 by default (Pebesma and Bivand 2023, Ch. 12). Prior to use in the model, our weather observations were linked to station longitudes and latitudes likewise obtained directly from Oklahoma Mesonet, thus putting their data structure in conversation with one-by-one kilometer grid cell coordinates in the Tulsa area.

As the final step in creating our analytic dataset, we merge the zip code-by-test day weather measures back into the TPS records of MAP administrations. The end result of this process is a dataset structured in a student-by-MAP administration format that contains information on student characteristics, performance on the MAP assessment, and measures of both test-day average temperature and the mean average daily temperature over roughly three months prior to each student MAP administration. Table 1 presents descriptive statistics for our dataset. The table illustrates that TPS is a racially diverse district, with about a quarter of students identifying as Black, a quarter white, and one-third Hispanic. And about 15 percent of students identify as Native American or as multiple races. The MAP administrations are about evenly split between the three seasons, although the proportion of spring administrations is somewhat lower due to the COVID pandemic precluding spring administrations in 2020. Finally, Table 1 illustrates that, on average, TPS students score at the 40th percentile of the national MAP distribution.

Figure 2 presents the distributions of our weather measures for each of the three seasons of MAP administration.[4] The top panel presents the distributions of average test-day temperature separately for the fall, winter, and spring administration. For each season, the panel depicts substantial variation in average test-day temperature. For example, average test-day temperatures during fall administrations range from just over 60 degrees to well over 80—the high temperature on days with an average temperature of 80 is typically in the mid- to high-90s, perhaps even reaching 100. The distribution of test-day temperatures for spring administrations is, on average, somewhat cooler than fall, but there is still clear overlap. Each of these two

[4] Figure 2 presents the distributions of average test-day temperature and average daily temperature for the three months leading up to MAP administrations in math. Although the analytic samples for reading and math MAP administrations differ slightly, the distributions are substantively identical.

seasons had MAP administrations on both hot and cold days, but the coldest days were in spring and the hottest ones in fall. Unsurprisingly, winter administrations have, on average, the coldest average temperature, but the range is substantial. Some winter administrations occur on days with an average temperature down in the 20s while others take place on days where the average temperature is well into the 60s.

[Insert Figure 2 and Table 1 about here]

The bottom panel of Figure 2 presents distributions of the mean of the average daily temperatures over each 90-day period leading up to fall, winter, and spring MAP administrations. Of the three seasonal administrations, fall has far and away the highest average temperatures in the three months leading up to the test, but the range is still about ten degrees—some summers are hotter than others. Spring and, especially, winter exhibit substantial variation in the average temperature over the three months prior to MAP administration. For spring, the wide range of temperatures simply reflects the significant variability of Oklahoma weather. For winter, however, it reflects both the innate variability of Oklahoma weather and variation in administration dates. Some winter administrations occurred as early as December while others took place as late as February. Thus, for some administrations, the data are capturing temperatures over September, October, and November—this is the relatively warmer part of the distribution—while for others the data are capturing temperatures in November, December, and January, which are much colder months. As we describe below, our empirical strategy accounts for variation in MAP administration dates.

Empirical Strategy

We estimate the effect of weather on student MAP performance using a model of the form:

$$P_{its} = \beta P_{its-1} + \gamma W_{its} + \delta D_{its} + \varepsilon_{its} \tag{1}$$

where we model the MAP percentile, P, for student i in school year t and season s as a function of their lagged MAP percentile, P_{its-1}, our weather measure of interest, W, and a measure of the number of calendar days since the prior MAP administration, D, which we specify as a fixed effect. Finally, ε is an i.i.d. error term. We estimate this model via ordinary least squares (OLS) with standard errors clustered by student. We estimate this model separately for reading and math and, for each of these two subjects, separately for the fall, winter, and spring MAP administrations. Finally, for each subject and season, we estimate the model separately for our two weather measures—average test-day temperature and the mean of the average daily temperature over the three months leading up to and through MAP administration. In addition to the specification in equation (1), we also estimate a variant that conditions on observable student characteristics, namely sex, race/ethnicity, and disability status.

Several features of the model presented in equation (1) warrant discussion. First, by specifying students' lagged MAP percentile on the right-hand side of the model we estimate the effect of weather on the change in students' MAP percentile from one administration to the next. Importantly, our analytic sample only includes MAP administrations for which the student has a score from the prior seasonal administration. For example, our sample only includes observations for the fall MAP administration if the student has a score from the spring administration in the prior school year. Similarly, observations of winter (spring) MAP administrations only enter our analytic sample if the student has an observation from the fall (winter) administration.

Second, contrary to several prior studies on the topic (e.g., Park et al. 2020), our model does not contain fixed effects for any unit of time, namely school year. Our decision to omit school year fixed effects is motivated by the nature of the variation in our empirical setting. Whereas much prior work draws on samples containing a significant degree of geographic dispersion, our sample is concentrated in a small geographic area. Accordingly, in our application a large majority of variation in temperature is temporal rather than spatial, particularly for our three-month temperature measure. Thus, specifying the model with a fixed effect for school year would absorb almost all of the variation in weather, resulting in the coefficient for our weather measure being estimated almost exclusively with the small amount of spatial variation. Of course, the omission of school year fixed effects does not imply a lack of concern about validity threats posed by unobserved, time-varying factors correlated with both weather and MAP performance. Indeed, this concern motivates our inclusion of students' lagged MAP performance on the right-hand side of the model, which is a common approach for addressing unobserved heterogeneity.

Third, as we noted in our discussion of Figure 2, there were some within-season differences in the timing of MAP administration, particularly for the winter administration. These differences in administration timing generated a scenario where students differed in the number of instructional days between MAP administrations, which could clearly affect the amount of progress that students make from one administration to the next. We account for these differences by explicitly conditioning on the number of calendar days between MAP administrations, represented by D in equation (1) above.

The coefficient of interest in the model presented in equation (1) is γ, which under plausible assumptions represents the causal effect of weather on the change in students' MAP percentile from one administration to the next. Specifically, γ can be interpreted causally as long as W is exogenous conditional on the contents of the model. Although not directly testable, there is little reason to think that either test-day weather or average weather over the three months leading to MAP administration would be correlated with factors shaping MAP performance, particularly after conditioning on both students' performance on the prior MAP administration and the number of days since that administration.

Results

Table 2 presents the estimated effects of test-day temperature on student MAP performance. The top panel of the table presents the effects for math while the bottom panel presents reading effects. For each of the three seasons, the left-hand column presents results from the model depicted in equation (1), which does not contain observable student characteristics, while the right-hand column presents results from a variant containing those characteristics. The fact that the two sets of results are nearly identical to one another provides some indirect evidence in support of our main identifying assumption, that our weather measures are exogenous conditional on students' lagged MAP performance and a fixed effect for the number of days since the prior administration.

[Insert Table 2 and Figure 3 about here]

The primary takeaway from Table 2 is that test-day temperature has no significant effect on student MAP performance in fall or winter, but a clear negative effect on student performance, particularly in math, on the spring MAP administration. Figure 3 helps convey the substantive magnitude of the statistical results in Table 2. Specifically, separately for math and reading, Figure 3 presents binned scatterplots of the relationship between average test-day temperature for spring MAP administrations (x-axis) and student MAP percentile (y-axis). To create these scatterplots, we first order the analytic sample according to test-day temperature, divide it into 50 equally-sized bins, and then, for each bin, plot the mean of test-day temperature and student MAP performance. Importantly, prior to plotting the means, both test-day temperature and student MAP are residualized on students' lagged MAP performance and the fixed effect for the number of days since the prior administration. Accordingly, the visual depictions in Figure 3 are identical to their statistical analogs in Table 2.

Figure 3 illustrates that, in math, average student performance declines by about three percentiles—from about the 44th percentile to the 41st—as we move from the lower part of the distribution, where the average test-day temperature is about 50 degrees, to the upper reaches, where average test-day temperature is closer to 80 degrees. Although an effect of this size is considered relatively small in the scheme of educational interventions (Kraft 2020), there is clear value in recognizing that student test performance is perceptibly affected by a factor over which educational stakeholders have no influence. The bottom panel of Figure 3 illustrates a similar pattern for reading, albeit one more modest in magnitude. Average student performance only declines by about one percentile across the range of the distribution.

An interesting feature of Table 2 is the presence of test-day temperature effects on students' spring MAP performance, but not on their fall MAP performance. Such a pattern is particularly notable given Figure 2's illustration that the distribution of average test-day temperature for fall administrations lies noticeably to the right of the spring distribution, although there is some overlap. Prior literature, however, suggests that this pattern could be explained by acclimatization. At the time of the spring MAP administration, TPS students have had very little recent exposure to hot temperatures, leaving them more susceptible to the physiological impacts of heat. By the fall MAP administration, however, students may have endured three straight months where high temperatures exceed 90 degrees almost every day, allowing them to adapt to the heat.

[Insert Table 3 and Figure 4 about here]

Table 3 presents the estimated effects of the average temperature over the 90 days leading up to MAP administration on student test performance, with the top panel again presenting math results and the bottom panel detailing the reading impacts. Within each season, the right- and left-hand columns present results from models with and without, respectively, student background characteristics. Figure 4 provides a visual complement to the statistical results in Table 3, presenting binned scatterplots for each set of results presented in column (1).

We highlight three takeaways from the results presented in Table 3 and Figure 4. First, it is clear that summer temperature has a statistically significant and substantively meaningful effect on student performance on the fall MAP assessment. Perhaps contrary to expectations, though, the effect is positive. That is, student performance increases as the average temperature over the 90 days leading up to the fall MAP administration rises, with the magnitude of the impact larger in reading than in math. Figure 4 illustrates that, in math, average student performance increases by about 3 percentiles as we move from relatively cooler summers to relatively hot ones. In reading, the effect is even larger, with student performance about four percentiles higher in the upper parts of the three-month temperature distribution, compared to lower parts of the distribution. Although the mechanisms driving these effects are not fully clear, one plausible explanation, consistent with prior literature (e.g., Graff Zivin and Neidell 2014; Heaney et al. 2019; Fan et al. 2023), is students spending relatively more time indoors during hot summers and engaging in activities, such as reading, that boost performance on their fall MAP assessment.

Second, the results make clear that long-term temperature has a positive, statistically significant impact on student math performance in all three seasons. However, the impact on student scores for the spring and, especially, winter administrations are substantively modest. In spring, Figure 4 demonstrates that the difference in average student performance between the top and bottom of the three-month temperature distribution is about two percentiles. In winter, the difference is only about a percentile. Third, Figure 4 illustrates that, in reading, the average temperature over the 90 days leading up to either winter or spring has no meaningful effect on student MAP performance. Together, the results in Table 3 and Figure 4 illustrate that, in a district with recently updated climate control capabilities, long-term temperature matters most for student learning during the summer, a period when students typically have significant discretionary time. This suggests that, along with the test-day temperature impacts likely operating through physiological mechanisms, weather also affects student learning through behavioral mechanisms, specifically time allocation.

Supplemental Analyses and Robustness Checks

Our main results present the average effect of temperature—both test-day and 90-day—on student academic performance. These average impacts may mask variability in the effects of temperature along several dimensions. In this section, we assess the potential for heterogeneity in temperature impacts along three different dimensions: 1) Student race/ethnicity, 2) The distribution of student achievement, and 3) Average income of students' zip code. Together, this series of analyses will contribute important detail to our understanding of the impact of temperature on student learning.

Along with his exploration of heterogeneity, this section also presents results from a series of alternative empirical specifications. Specifically, we estimate: 1) A specification that replaces the measure of lagged achievement with a student fixed effect, 2) A variant of the empirical model that contains measures of both test-day and long-term temperature, and 3) A specification that adds a school fixed effect to the model presented in equation (1). The results of these analyses will shed light on the robustness of our primary results presented in Tables 2 and 3 to alternative analytic choices.

Heterogeneity by Student Race/Ethnicity

It is possible that the academic performance of students from different racial/ethnic backgrounds could be differentially affected by temperature, either test-day or longer-term. Such variability could occur because students from different racial or ethnic groups could experience distinctive built environments—both at home and in school—that could, in turn, generate different physiological or behavioral responses relevant to academic performance. To assess the possibility of such heterogeneity, we estimate the model in equation (1) separately for students from five racial or ethnic groups—Black, Hispanic, white, Native American, and students who identify as multiple races. We again estimate the models separately by subject, season, and temperature type (i.e. test-day vs. 90-day). After estimating the model for each racial/ethnic group for a given subject, season, and temperature measure, we conducted a test of coefficient equality for the temperature measure across the five racial/ethnic groups. That is, we ask whether we can reject the null hypothesis that the temperature coefficients for the five racial/ethnic groups are equal. Rejecting this null would provide evidence of statistical variation in temperature impacts across racial/ethnic groups.

[Insert Tables 4 and 5 about here]

We present the results of this analysis in Tables 4 (math) and 5 (reading). In each table, the top panel presents results for the fall MAP administration while the middle and bottom panels present results for winter and spring, respectively. The left-hand side presents the estimated effect of test-day temperature separately for each of the five racial/ethnic groups while the right-hand side presents results for 90-day temperature. In addition to the coefficient estimates for each racial/ethnic group, the table also presents the *p*-value from the test of coefficient equality.

The results show that, in math, there is no signification variation across groups in the impact of test-day temperature. For the 90-day temperature measure, however, the results demonstrate statistically significant heterogeneity in the effect of temperature for the fall administration, with marginally significant variation apparent in winter and spring. In fall, the positive impact of summer heat is noticeably larger for Hispanic and, to a lesser extent, white students than it is for Black students and those who identify as multiple races. The winter results suggest small positive temperature impacts for Black and Hispanic students and null results for the other groups. In spring, the results demonstrate a null relationship between temperature and MAP performance for Hispanic students while the estimates are positive and significant for the four other groups.

The reading results, presented in Table 5, only reveal significant variation in the impacts of temperature for test-day temperature during the spring MAP administration. Here, test-day heat significantly reduces the performance of white and Hispanic students, as well as those who identify as multiple races, but has no impact on the performance of Black or Native American students. In all other cases, our test of coefficient equality cannot reject the null that the coefficient estimates for the five racial/ethnic groups are equal to one another.

Considered as a whole, although there are a handful of contexts in which the effects of temperature vary across students from different racial/ethnic backgrounds, the major story emerging from Tables 4 and 5 is that such variation is the exception, rather than the rule.

Heterogeneity Across the Distribution of Student Achievement

Our primary results demonstrated that test-day temperature affected student performance on the spring MAP administration while longer-term temperature impacted students' fall MAP scores. We further explore these results by analyzing whether these effects vary for students at different points of the achievement distribution. To perform this analysis we estimate—separately for each combination of season, subject, and temperature measure—a series of regressions where we specify the outcome as an indicator that a student's MAP percentile is equal to or exceeds percentile *p*, with *p* ranging from 1 to 99 in increments of 1. Thus, for a given combination of season, subject, and temperature measure (e.g., test-day temperature and spring MAP performance in reading) we estimate 99 separate regressions. In each regression, we model the outcome as a function of the temperature measure, a fixed effect for a student's percentile on the prior administration, and the number of days since the prior MAP administration. Then, for each of those regressions, we recover both the coefficient estimate for the temperature measure and the 95 percent confidence interval and plot those points in order from *p*=1 to *p*=99. That is, we plot the coefficient estimate for the temperature measure—and its attendant confidence interval—for the probability of scoring at or above *p*=1, then do the same for *p*=2 through *p*=99. Together, these plots shed light on whether the impacts of temperature are larger (or smaller) at some points of the achievement distribution than at others.

[Insert Figure 5]

Figure 5 presents these plots. The top row of the figure presents the impact of test-day temperature on spring MAP performance, with math on the left and reading on the right. The bottom row presents the impact of 90-day temperature on fall MAP performance—math and reading are again on the left and right, respectively. We highlight two takeaways from the top-row plots. First, test-day heat negatively impacts math performance across the full distribution, but the effects are largest the 60th percentile of the distribution. Second, the impacts in reading are uniformly smaller in magnitude than those in math, and are generally modest in magnitude.

The impacts that do exist, though, do not meaningfully differ across the achievement distribution.

Turning to the bottom row of Figure 5, which presents the impact of 90-day temperature on fall MAP performance, we see positive effects across the full distribution for both reading and math. There are differences between the two subjects, however, in the points of the distribution at which the impacts are relatively larger or smaller. In math, the impacts are largest between about the 20th and 50th percentiles of the distribution. By contrast, the reading impacts are relatively larger above the 50th percentile. On the whole, though, and as was the case with our analyses of potential heterogeneity by student race/ethnicity, the story emerging from Figure 5 is one of broad similarity in the impact of temperature across the achievement distribution.

Heterogeneity by Zip Code Income

Motivated by the likelihood of differences in the built environment between zip codes with lower and higher income levels—and variability in students' physiological or behavioral responses spurred by these environmental differences—we examine whether the impact of temperature on academic performance differs by average zip code income. We construct our measure of zip code income via a two-step process. First, we obtained annual IRS tax statistics at the zip code level from 2014 to 2020. Then, separately for each year, we used these statistics to calculate the adjusted gross income (AGI) per tax return filed in each zip code. We use mean AGI as our measure of zip code income in the analyses in this section. Using this measure, we estimate the model in equation (1) separately for students residing in zip codes with average AGIs below and above the median for all zip codes. We again estimate the models separately by subject, season, and temperature type (i.e. test-day vs. three-month). After estimating the model for each group for a given subject, season, and temperature measure, we conduct a test of coefficient equality for the temperature measure across the low- and high-income zip codes.

[Insert Table 6 about here]

We present the results of this analysis in Table 6. The left-hand panel of the table presents math results while the right-hand panel presents reading results—each panel contains separate columns presenting the estimated effects for test-day and 90-day temperatures. Further, the top, middle, and bottom panels present results for fall, winter, and spring MAP administrations, respectively.

For test-day temperature, only the results for the spring MAP administration provide any evidence of variability. Those results, however, suggest that the negative effect of test-day heat is larger for students residing in relatively less affluent zip codes. In reading, test-day temperature only affects the performance of students in lower-income zip codes—the estimate for students in relatively affluent zip codes is insignificant. In math, estimates for both lower- and higher-income zip codes are negative and significant, but the estimate for lower-income areas is larger in magnitude than the one for higher-income zip codes.

Turning to three-month temperature, the results only demonstrate significant variation by zip code income for the fall MAP administration in math. Again, the estimated effects are larger for students residing in relatively less affluent zip codes—the estimated effect for students in lower-income neighborhoods is about twice as large than for students in higher-income areas. The reading results exhibit a similar pattern, but the difference between the two sets of estimates does not reach statistical significance. Together, the results in Table 6 illustrate that, in most cases, temperature has no differential impact on the MAP performance of students in lower- and higher-income neighborhoods. In cases when there is a difference, though, the impacts are more pronounced for students in relatively lower-income zip codes, compared to their peers residing in more affluent areas.

Alternative Empirical Specifications

The model presented in equation (1) represents our preferred approach for estimating the effect of temperature on student MAP performance—we discuss our rationale for that preference above. We recognize, however, that there are other reasonable, if flawed, specifications of an empirical model, and here we present results from three alternative specifications.

First, we present results from a specification that models student MAP performance solely as a function of a temperature measure—either test-day or 90-day—and a student fixed effect. The student fixed effect addresses validity threats from time-invariant factors correlated with both the temperature measure and student MAP performance, thereby performing a function similar to that of lagged MAP percentile in equation (1). However, the student fixed effects approach does not exploit the seasonal administration of the MAP assessment as effectively as our preferred empirical strategy. Because we estimate the model separately by season, there is a full calendar year between MAP administrations. Our longer-term temperature measure, by contrast, is a 90-day measure, meaning that roughly 275 days of weather could, in theory, affect student learning but are not reflected in our temperature measure. Further, this approach does not account for the varying administration dates within season, or for differences in timing between administrations. With these caveats in mind, the results from a model containing a student fixed effect still provide a useful point of comparison to our primary results.

Second, we present results from a variant of the specification in equation (1) in which the model contains both temperature measures, test-day and 90-day temperature. If the measures are indeed exogenous, a key identifying assumption in our analysis, then the coefficient estimates should not be materially changed through simultaneous inclusion of the measures. Third, we present results from a variant of equation (1) where we include a school fixed effect in the model. In doing so, we leverage within-school variation in weather exposure to estimate the effect of temperature on student MAP performance. This within-school variation is both temporal in nature—variation occurring across years—and spatial, due to differences in students' residential locations. Together, these two analyses provide important robustness checks for our primary results.

[Insert Tables 7 & 8 about here]

We present the results of these three alternative specifications in Tables 7 (test-day temperature) and 8 (90-day temperature). In each table, the left-hand panel presents results from the fall MAP administration while the middle and right-hand panels present results from the winter and spring administrations, respectively. The three alternative specifications described above are presented in the top, middle, and bottom panels.

We highlight three main takeaways from these results. First, the two major substantive findings from our primary results—positive impacts of summer heat on fall MAP performance and negative impacts of test-day temperature on a student's spring MAP percentile—are also apparent in the results of each alternative specification, providing further confidence in these impacts. Second, the results from almost all alternative specifications in Table 7 show a significant, negative impact of test-day heat on a student's fall MAP percentile.[5] Although our primary results return insignificant estimates of the impact of test-day temperature on fall MAP performance, the consistency of the alternative results raise the prospect that the impacts of test-day heat play out in both the spring and fall. Third, in addition to the aforementioned positive impacts of summer heat, all three alternative specifications for the 90-day temperature results show a positive and statistically significant effect of warmer temperatures on spring math

[5] The exception is the reading result from the specification containing a school fixed effect—this estimate is positive and statistically significant.

performance. This finding is consistent with our primary results, further instilling confidence in a statistically significant, if substantively modest, impact.

Considered as a whole, the results from the alternative empirical specifications are remarkably similar in sign and significance to our preferred estimates presented in Tables 2 and 3 above.

Discussion and Conclusion

As our changing climate exposes students to extreme temperatures with increasing frequency, it becomes ever more important to gain further understanding of how such exposure affects student learning. In this paper we use detailed data from MAP administrations over a six-year period in Tulsa, Oklahoma, combined with high-resolution temperature records, to paint a clear portrait of the effect of temperature on student learning. Exploiting the tri-annual MAP administration—TPS administers the assessment in fall, winter, and spring—our study is uniquely able to shed light on how the effects of temperature on student learning vary across seasons. Moreover, the richness of our data allows us to gain insight about heterogeneity on the basis of a student's racial or ethnic background, the average income of the zip code in which they reside, and by their place in the distribution of student achievement.

These analyses produced a few main takeaways. First, we show that test-day temperature has no significant effect on student MAP performance in fall or winter, but a clear negative effect on students' spring MAP performance, particularly in math. Second, summer temperature has a positive, statistically significant, and substantively meaningful effect on student performance on the fall MAP assessment—these effects appear in both math and reading. The results also illustrate that 90-day temperature affects math performance in winter and spring, but these estimates are modest in substantive magnitude. Third, our analyses of heterogeneity show little meaningful variation by student race/ethnicity or the distribution of student achievement.

There is some evidence of heterogeneity according to average zip code income and, when such variation exists, the impacts are larger among students residing in relatively less affluent areas than among students in zip codes with above-average income levels. Together, these findings have implications for research, policy, and practice.

The existing research literature identifies two main mechanisms through which the effects of temperature operate—physiological reactions and behavioral responses. Further, these mechanisms interact with the built environment in ways that shape those responses. For this study, the most important feature of schools' built environment is climate control, primarily air conditioning, but also heat. In contrast to many prior studies with samples spanning a broad geography, and thus broad variation in climate control capabilities, our focus on a single educational context minimizes variation on this front. The HVAC systems in all TPS schools were renovated shortly before our study period. Of course, there are likely differences in the physical characteristics of students' home environments, but variability in the built schooling environment is unlikely to explain the estimated effects.

Similar considerations apply to acclimatization. In studies that draw their samples from different geographic or climatological contexts, heterogeneity in the effects of temperature on student learning could be attributable to differences in acclimatization across those contexts. In our study, the compact geography minimizes the potential for differences in student acclimatization for any given MAP administration. Although students almost certainly differ in their degree of acclimatization across administrations, the within-administration variability is likely to be relatively small.

So, if the estimated effects are unlikely shaped by physical differences in students' educational environment or within-administration variation in acclimatization, what are the likely explanations for the results described above? Although data limitations prevent us from identifying the mechanisms at play with certainty, we discuss a couple candidate explanations. The positive effect of summer heat on fall MAP performance is consistent with a behavioral response where the uncomfortably hot temperatures induce students to spend relatively more time indoors engaging in activities that, on the margin, improve performance on the fall MAP administration. Data limitations prevent us from observing students' specific activities or time use, but such questions would be a natural topic for future inquiry. More generally, future research would do well to explore how weather shapes students' educationally-relevant time use, both in and out of formal schooling settings.

We find it notable that 90-day temperature meaningfully matters for fall MAP performance, but has either modest or no effects on winter or spring MAP scores. Such a pattern could be explained by the imbalance in students' discretionary time across these periods. Students typically have a relatively large amount of flexibility in their time use during the 90 days preceding the fall MAP administration, relative to the winter and spring administrations. It is perhaps unsurprising, then, that effects of longer-term temperatures are disproportionately apparent during periods when students have relatively more control over their time use. Consequently, building on the suggestions for future research discussed above, additional work would do well to examine how the effect of weather on student time use varies across the school year.

Our finding that test-day heat reduces students' performance on the spring MAP administration is consistent with research showing that hot temperatures can disrupt core cognitive abilities (Taylor et al. 2016), including memory (Gaoua et al. 2011; Lee et al. 2015) and decision-making (Froom et al. 1993; Coehoorn et al. 2020). These functions have obvious implications for test performance. Notably, though, evidence of negative effects of test-day heat on fall MAP performance—an administration for which temperatures are often hotter than in spring—is somewhat less consistent. Such a pattern is consistent with students being less acclimated to heat during the spring administration than during the fall one. More generally, this combination of findings suggest that student academic performance is most likely to be notably affected when test-day weather differs substantially from the conditions to which students have become acclimated.

Through its simultaneous exploration of both test-day and longer-term temperature, coupled with its analysis of heterogeneity on the basis of season, student background, and other dimensions, this paper meaningfully advances our understanding of how temperature affects student learning. In doing so, it contributes to ever-growing literature documenting how features of the natural environment play a substantial role in shaping students' educational outcomes. Along with answering several important questions, however, it also raises many others. For example, it does little to address how other meteorological phenomena—precipitation, humidity, and wind are just three examples—might affect student learning. And, as we note above, our data are limited in their ability to shed light on the specific behaviors or mechanisms generating the effects. Finally, we have no way of knowing whether our results are specific to the TPS context, or whether they have broader applicability. Given scholars' increasing recognition of the importance of the natural environment for students' educational outcomes, such issues will likely be addressed as the literature continues to evolve.

References

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

Austin, Wes, Garth Heutel, and Daniel Kreisman. 2019. School bus emissions, student health and academic performance. *Economics of Education Review*, 70: 109-126. https://doi.org/10.1016/j.econedurev.2019.03.002.

Brock, F. V., Crawford, K. C., Elliott, R .L., Cuperus, G. W., Stadler, S. J., Johnson, H. L. and Eilts, M. D. 1995. The Oklahoma Mesonet: A Technical Overview. *Journal of Atmospheric and Oceanic Technology*, 12: 5-19. https://doi.org/10.1175/1520-0426(1995)012%3C0005:TOMATO%3E2.0.CO;2

Castellani J. W., and Young A. J. 2016. Human physiological responses to cold exposure: Acute responses and acclimatization to prolonged exposure. *Auton Neurosci*, 196: 63-74. DOI: 10.1016/j.autneu.2016.02.009.

Cho, H. 2017. The effects of summer heat on academic achievement: A cohort analysis. *Journal of Environmental Economics and Management* 83: 185-196. doi:10.1016/j.jeem.2017.03.005.

Coehoorn, C. J., L. A. Stuart-Hill, W. Abimbola, J. P. Nealy, and O. E. Krigolson. 2020. Firefighter neural function and decision-making following rapid heat stress. *Fire Safety Journal* 118: 103240. https://doi.org/10.1016/j.firesaf.2020.103240

Davis, D. "16 Years, $65 Million Later, Tulsa Public Schools Near End of Air Conditioning Upgrade." News on 6. August 3rd, 2012. https://www.newson6.com/story/5e364a142f69d76f62063403/16-years-65-million-later-tulsa-public-schools-near-end-of-air-conditioning-upgrade (Accessed July 29th, 2023).

Fan, Y., J. Wang, N. Obradovich, and S. Zheng. 2023. Intraday adaptation to extreme temperatures in outdoor activity. *Sci Rep* **13:** 473. https://doi.org/10.1038/s41598-022-26928-y

Froom, P., Y. Caine, I. Shochat, J. Ribak. 1993. Heat Stress and Helicopter Pilot Errors. *Journal of Occupational Medicine* 35 (7): 720-724. https://www.jstor.org/stable/45013929

Gaoua, N., S. Racinais, J. Grantham, F. E. Massioui. 2011. Alterations in cognitive performance during passive hyperthermia are task dependent. *International Journal of Hyperthermia* 27 (1): 1-9. https://doi.org/10.3109%2F02656736.2010.516305

Graff Zivin, J. and M. Neidell. 2014. Temperature and the Allocation of Time: Implications for Climate Change. *Journal of Labor Economics*, 32 (1): 1-26. https://doi-org.ezproxy.lib.ou.edu/10.1086/671766

Graff Zivin, J., S. M. Hsiang, M. Neidell. 2018. Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists,* 5 (1), 77-105. Available at: https://gps.ucsd.edu/_files/faculty/graff-zivin/graff-zivin-publications-2018.pdf.

Graff Zivin, J., Y. Song, Q. Tang, 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. https://doi.org/10.1016/j.jeem.2020.102365.

Heaney, A. K., Daniel Carrión, Katrin Burkart, Corey Lesk and Darby Jack. 2019. Climate Change and Physical Activity: Estimated Impacts of Ambient Temperatures on Bikeshare Usage in New York City. *Environmental Health Perspectives*, 127 (3). https://doi.org/10.1289/EHP4039

Heissel, Jennifer A., Claudia Persico and David Simon. 2022. Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance. *Journal of Human Resources*, 57 (3): 747-776. https://doi.org/10.3368/jhr.57.3.1218-9903R2

Johnston, D. W., R. Knott, S. Mendolia, P. Siminski. 2021. Upside-Down Down-Under: Cold Temperatures Reduce Learning in Australia. *Economics of Education Review,* 85: 102172. https://doi.org/10.1016/j.econedurev.2021.102172

Lee, W., M. Moon, H. G. Kim, T. H. Lee, M. Sook Oh. 2015. Heat stress-induced memory impairment is associated with neuroinflammation in mice. *Journal of Neuroinflammation,* 12: 102. https://doi.org/10.1186/s12974-015-0324-6

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. https://doi.org/10.1016/j.econlet.2020.109698

Mohajerani A., J. Bakaric, and T. Jeffrey-Bailey. 2017. The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete. *J Environ Manage,* 197: 522-538. DOI: 10.1016/j.jenvman.2017.03.095.

Oke, T. R. 1982. The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108 (455): 1-24. DOI: http://dx.doi.org/10.1002/qj.49710845502

Park, J. R. Hot Temperature and High Stakes Performance. 2022. *Journal of Human Resources,* 57 (2): 400-434. https://doi.org/10.3368/jhr.57.2.0618-9535R3

Park, J. R., J. Goodman, M. Hurwitz, and J. Smith. Heat and Learning. 2020. *American Economic Journal: Economic Policy* 12(2): 306-39. https://www.aeaweb.org/articles?id=10.1257/pol.20180612.

Park, J. R., A. P. Behrer, and J. Goodman. 2021. Learning is inhibited by heat exposure, both internationally and in the United States. *Nature Human Behavior*. 5: 19-27. https://doi.org/10.1038/s41562-020-00959-9

Pebesma, Edzer J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30 (7): 683-691. https://doi.org/10.1016/j.cageo.2004.03.012

Pebesma, Edzer J. and R. Bivand. 2023. Spatial Data Science: With Applications in R. First Edition, Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Périard, J. D., Racinais S, and Sawka M.N. 2015. Adaptations and mechanisms of human heat acclimation: Applications for competitive athletes and sports. *Scand J Med Sci Sports*, 25 Suppl 1: 20-38. DOI: 10.1111/sms.12408.

Persico, Claudia L. and Joanna Venator. 2019. The Effects of Local Industrial Pollution on Students and Schools. *Journal of Human Resources*. https://doi.org/10.3368/jhr.56.2.0518-9511R2

Renee A. McPherson, Christopher A. Fiebrich, Kenneth C. Crawford, James R. Kilby, David L. Grimsley, Janet E. Martinez, Jeffrey B. Basar, Bradley G. Illston, Dale A. Morris, Kevin A. Kloesel, Andrea D. Melvin, Himanshu Shrivastava, J. Michael Wolfinbarger, Jared P. Bostic, David B. Demko, Ronald L. Elliott, Stephen J. Stadler, J. D. Carlson, and Albert J. Sutherland. 2007. Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, 24 (3): 301-321. https://doi.org/10.1175/JTECH1976.1

Ricke, K., L. Drouet, K. Caldeira, and M. Tavoni. 2018. "Country-level social cost of carbon." *Nature Clim Change*, **8**: 895–900. https://doi.org/10.1038/s41558-018-0282-y

Roach, Travis and Jacob Whitney. 2022. Heat and learning in elementary and middle school. *Education Economics*, 30 (1): 29-46, DOI: 10.1080/09645292.2021.1931815

Robinson, Alexander, Jascha Lehmann, David Barriopedro, Stefan Rahmstorf and Dim Coumou. 2021. Increasing heat and rainfall extremes now far outside the historical climate. *npj Clim Atmos Sci*, **4:** 45. https://doi.org/10.1038/s41612-021-00202-w

Soligo, Michael J., Peter A. Irwin, Colin J. Williams, and Glenn D. Schuyler. 1998. A comprehensive assessment of pedestrian comfort including thermal effects. *Journal of Wind Engineering and Industrial Aerodynamics*, 77-78: 753-766. https://doi.org/10.1016/S0167-6105(98)00189-5.

Taylor, L., S. L. Watkins, H. Marshall, B. J., Dascombe, and J. Foster. 2016. The Impact of Different Environmental Conditions on Cognitive Function: A Focused Review. *Frontiers in Physiology*, 6. https://doi.org/10.3389/fphys.2015.00372

Vu, T. M. Effects of Heat on Mathematics Test Performance in Vietnam. 2022. *Asian Economic Journal*, 36 (1): 72-94. https://doi.org/10.1111/asej.12259

Weber, D., and E. Englund. 1992. Evaluation and comparison of spatial interpolators. *Math Geol*, **24:** 381–391. https://doi-org.ezproxy.lib.ou.edu/10.1007/BF00891270

Wen, J., and M. Burke. 2022. Lower test scores from wildfire smoke exposure. *Nat Sustain*, **5:** 947–955. https://doi.org/10.1038/s41893-022-00956-y

Tables and Figures

Figure 1. Map of Tulsa Public School District and Mesonet Stations within 50 Miles

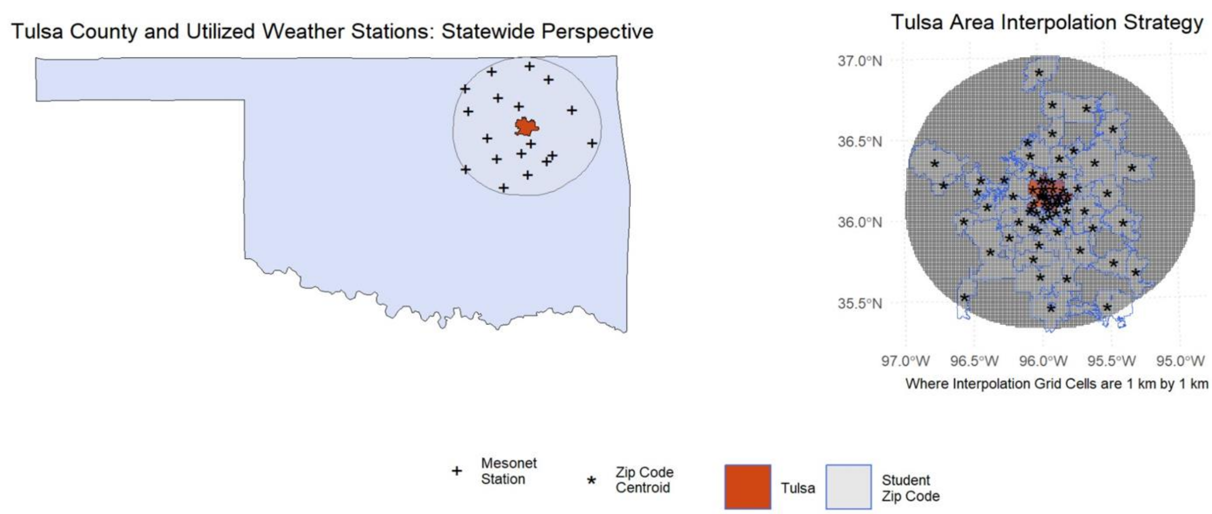


Table 1. Descriptive statistics

Characteristic	N	Mean	Std. Dev.	Min.	Max.
Female	339,034	0.492	0.500	0	1
Male	339,034	0.508	0.500	0	1
Asian	339,034	0.014	0.118	0	1
Black	339,034	0.239	0.426	0	1
Hispanic	339,034	0.348	0.476	0	1
American Indian/Alaska Native	339,034	0.051	0.219	0	1
Multiple Races	339,034	0.100	0.301	0	1
Pacific Islander	339,034	0.006	0.077	0	1
White	339,034	0.242	0.428	0	1
Disability	339,034	0.172	0.377	0	1
Kindergarten	339,034	0.144	0.351	0	1
Grade 1	339,034	0.147	0.354	0	1
Grade 2	339,034	0.150	0.357	0	1
Grade 3	339,034	0.166	0.372	0	1
Grade 4	339,034	0.070	0.255	0	1
Grade 5	339,034	0.065	0.246	0	1
Grade 6	339,034	0.060	0.237	0	1
Grade 7	339,034	0.050	0.219	0	1
Grade 8	339,034	0.048	0.214	0	1
Grade 9	339,034	0.048	0.214	0	1
Grade 10	339,034	0.043	0.203	0	1
Grade 11	339,034	0.007	0.081	0	1
Grade 12	339,034	0.003	0.051	0	1
Fall administration	339,034	0.361	0.480	0	1
Spring administration	339,034	0.281	0.450	0	1
Winter administration	339,034	0.357	0.479	0	1
MAP Percentile	339,030	40.212	28.881	1	99
Average 90-day Temperature	337,764	61.735	13.571	37.473	81.701
Test Day Temperature	337,764	61.426	15.669	21.971	87.916

Figure 2. Distribution of Average Test-Day Temperature and 90-Day Temperature, by Season

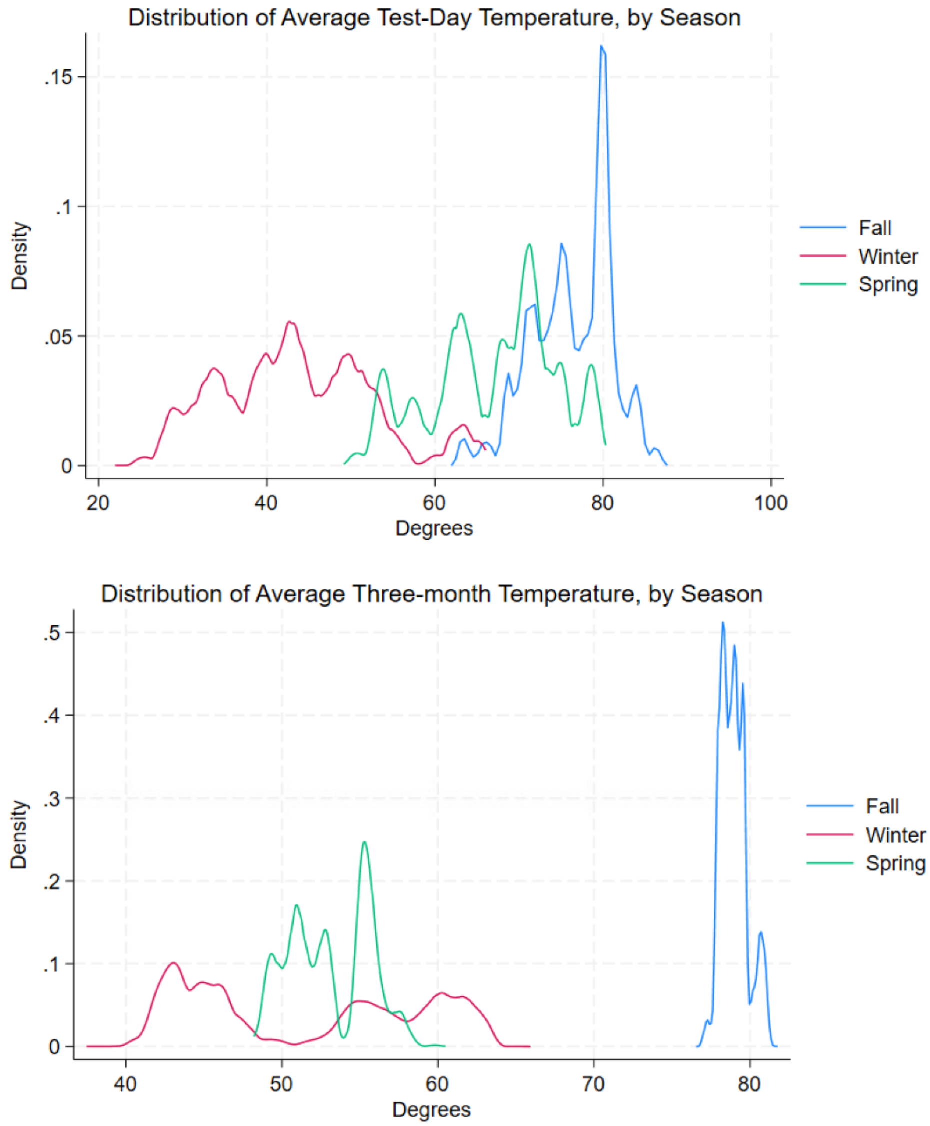


Table 2. Effect of Average Test-Day Temperature on Change in Student MAP Percentile From Prior MAP Administration, by Season and Subject

	MAP Administration					
	Fall		Winter		Spring	
	(1)	(2)	(1)	(2)	(1)	(2)
	Math					
Average Test-Day Temperature	-0.015 (0.014)	-0.016 (0.014)	0.002 (0.005)	-0.0001 (0.005)	-0.096*** (0.009)	-0.093*** (0.009)
N	71,642	71,642	112,490	112,490	89,033	89,033
Observable student chars		X		X		X
	Reading					
Average Test-Day Temperature	0.006 (0.015)	0.009 (0.014)	-0.009* (0.005)	-0.008 (0.005)	-0.037*** (0.010)	-0.036*** (0.010)
N	72,069	72,069	113,411	113,411	89,567	89,567
Observable student chars		X		X		X

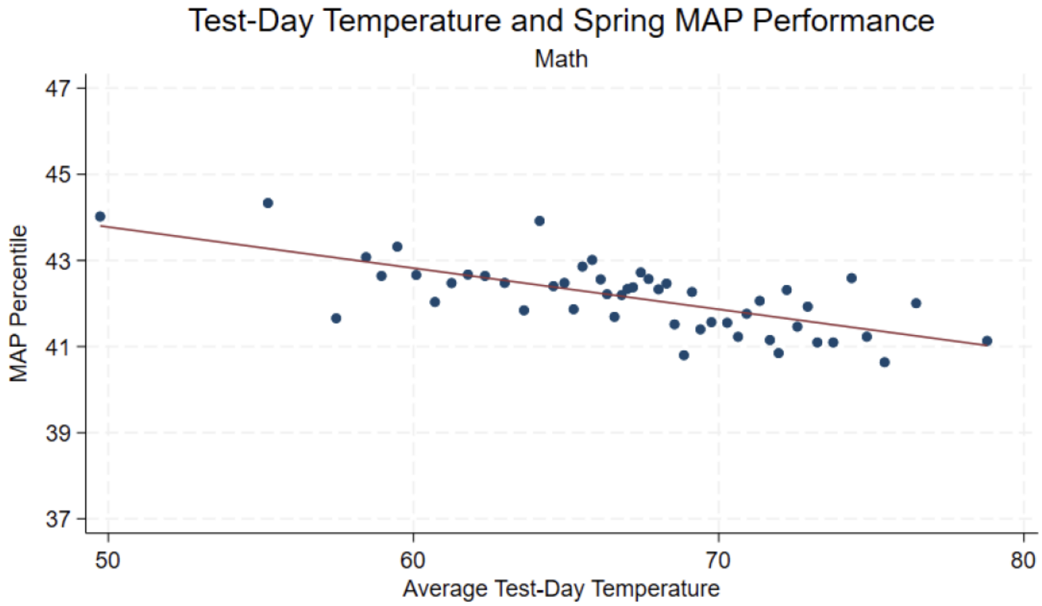
NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by student in parentheses below coefficient estimate Each coefficient from a separate regression estimated via OLS predicting student MAP percentile. In addition to the measure of average test-day temperature, all regressions contain a measure of student's MAP percentile in the same subject from the prior administration and fixed effects for the number of days since the prior MAP administration. Regressions in column (2) for each season also contain measures of student sex, race/ethnicity, and disability status.

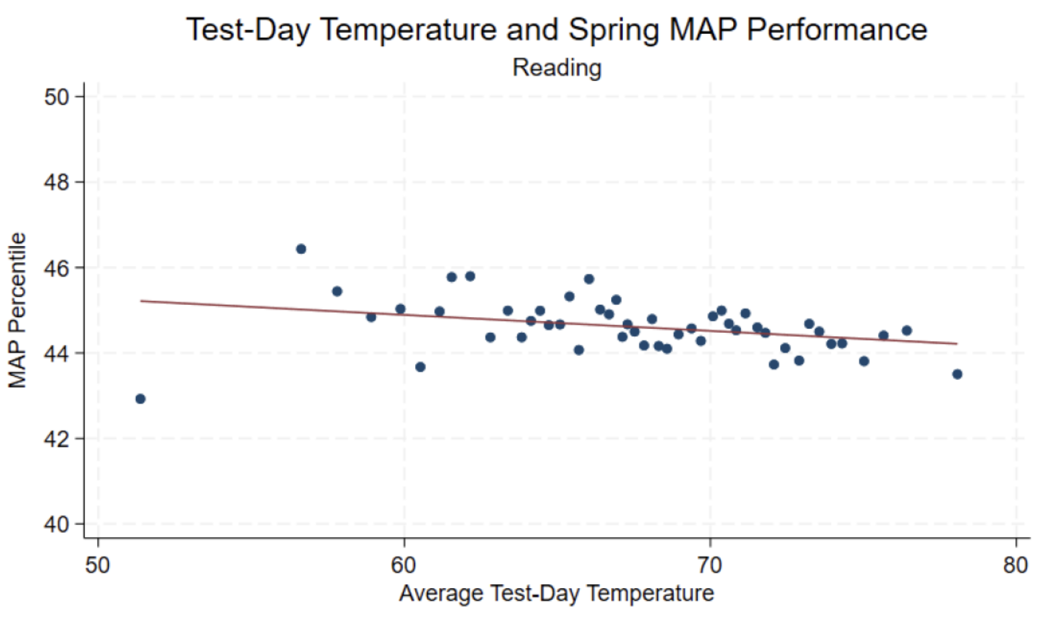
Table 3. Effect of Average 90-Day Temperature on Change in Student MAP Percentile From Prior MAP Administration, by Season and Subject

	MAP Administration					
	Fall		**Winter**		**Spring**	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
	Math					
Average 90-Day	0.864***	0.801***	0.061***	0.060***	0.181***	0.188***
Temperature	(0.085)	(0.083)	(0.020)	(0.020)	(0.028)	(0.028)
N	71,642	71,642	112,490	112,490	89,033	89,033
Observable student chars		X		X		X
	Reading					
Average 90-Day	1.117***	1.052***	-0.041**	-0.038*	-0.029	-0.030
Temperature	(0.078)	(0.076)	(0.020)	(0.020)	(0.029)	(0.029)
N	72,069	72,069	113,411	113,411	89,567	89,567
Observable student chars		X		X		X

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by student in parentheses below coefficient estimate Each coefficient from a separate regression estimated via OLS predicting student MAP percentile. In addition to the measure of average 90-day temperature, all regressions contain a measure of student's MAP percentile in the same subject from the prior administration and fixed effects for the number of days since the prior MAP administration. Regressions in column (2) for each season also contain measures of student sex, race/ethnicity, and disability status.

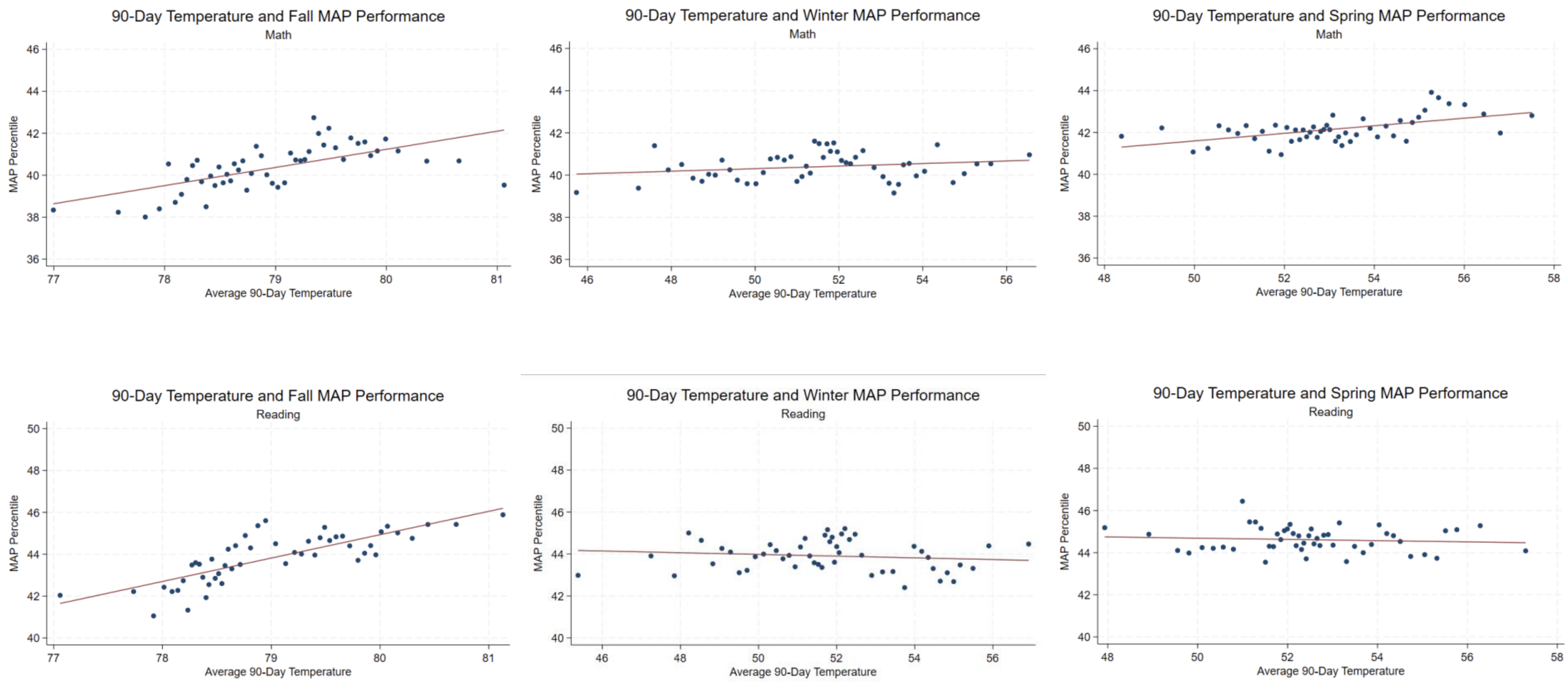
Figure 3. Binned Scatterplot of Average Test-Day Temperature and Spring MAP Percentile, by Subject





Note: Binned scatterplots created by: 1) Ordering the analytic sample according to test-day temperature, 2) Dividing it into 50 equally-sized bins, and 3) Plotting the mean of test-day temperature and student MAP performance. Prior to plotting the binned means, both test-day temperature and student MAP percentile are residualized on students' lagged MAP performance and the fixed effect for the number of days since the prior administration. Accordingly, the visual depictions in the figure are identical to their statistical analogs in Table 2.

Figure 4. Binned Scatterplot of Average 90-Day Temperature and Student MAP Percentile, by Season and Subject



Note: Binned scatterplots created by: 1) Ordering the analytic sample according to 90-day temperature, 2) Dividing it into 50 equally-sized bins, and 3) Plotting the mean of 90-day temperature and student MAP performance. Prior to plotting the binned means, both 90-day temperature and student MAP percentile are residualized on students' lagged MAP performance and the fixed effect for the number of days since the prior administration. Accordingly, the visual depictions in the figure are identical to their statistical analogs in Table 3.

Table 4. Effect of Average Temperature on Change in MAP Math Percentile From Prior MAP Administration, by Race/Ethnicity and Season

	Test Day Temperature					90-Day Temperature				
Math	**Black**	**Hispanic**	**Native American**	**Multiple**	**White**	**Black**	**Hispanic**	**Native American**	**Multiple**	**White**
					Fall					
Average Temperature	-0.026	-0.027	-0.032	-0.069	0.044	0.414***	1.021***	0.877**	0.553**	0.848***
	(0.027)	(0.023)	(0.066)	(0.047)	(0.030)	(0.152)	(0.131)	(0.376)	(0.266)	(0.166)
p: Test of equality			0.230					0.064		
N	16,799	25,870	3,556	7,096	16,914	16,799	25,870	3,556	7,096	16,914
					Winter					
Average Temperature	0.013	-0.003	0.023	-0.031*	-0.007	0.070*	0.096***	-0.090	0.007	-0.025
	(0.011)	(0.009)	(0.026)	(0.018)	(0.012)	(0.038)	(0.034)	(0.094)	(0.066)	(0.043)
p: Test of equality			0.214					0.090		
N	26,263	39,928	5,553	11,283	27,198	26,263	39,928	5,553	11,283	27,198
					Spring					
Average Temperature	-0.079***	-0.117***	-0.093**	-0.088***	-0.091***	0.214***	0.068	0.249**	0.314***	0.262***
	(0.018)	(0.015)	(0.042)	(0.029)	(0.019)	(0.056)	(0.047)	(0.127)	(0.091)	(0.055)
p: Test of equality			0.587					0.031		
N	21,331	30,941	4,469	8,764	21,784	21,331	30,941	4,469	8,764	21,784

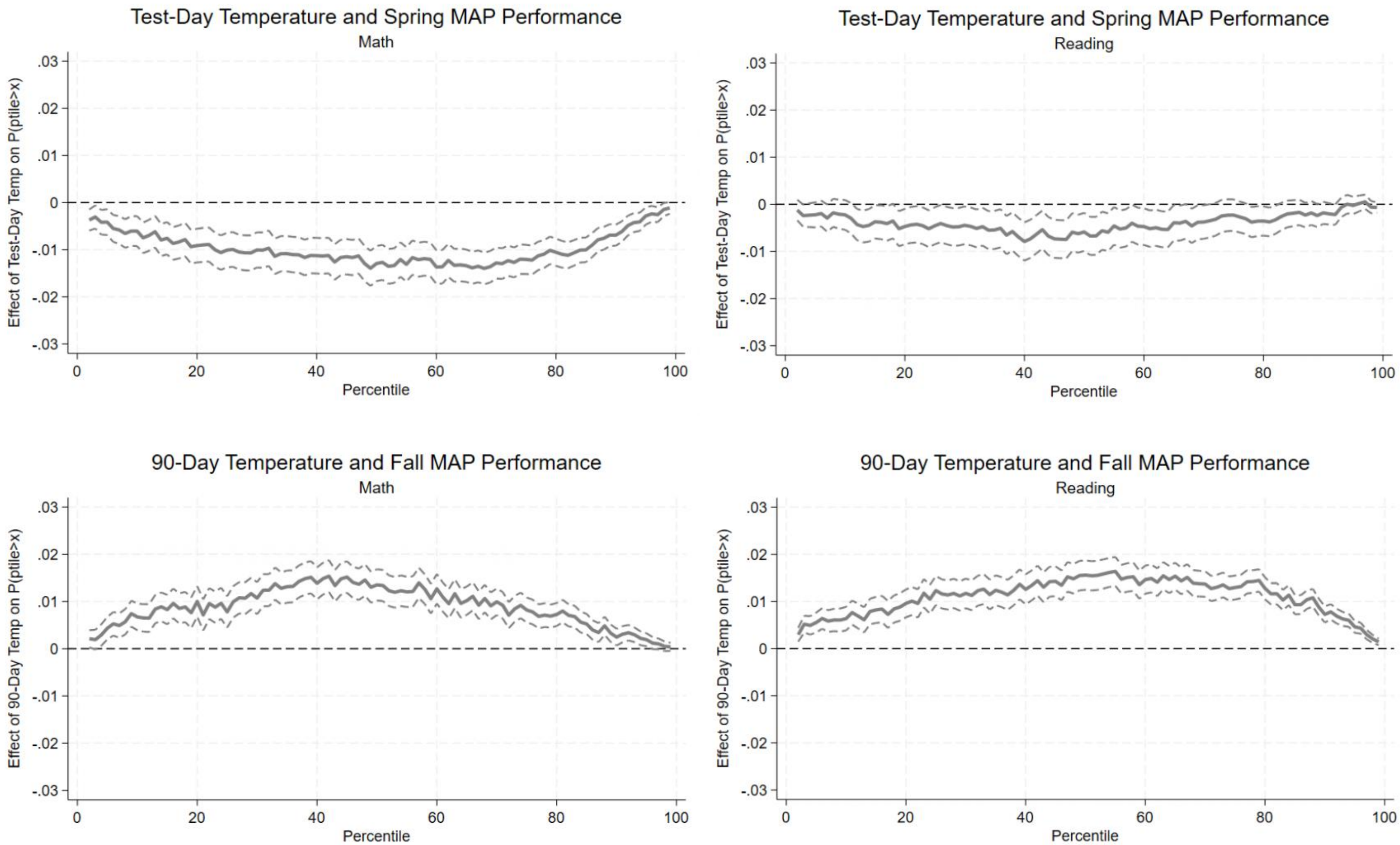
NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by student in parentheses below coefficient estimate. Each coefficient from a separate regression estimated via OLS predicting student MAP percentile. In addition to the measure of average temperature, all regressions contain a measure of student's MAP percentile in the same subject from the prior administration and fixed effects for the number of days since the prior MAP administration.

Table 5. Effect of Average Temperature on Change in MAP Reading Percentile From Prior MAP Administration, by Race/Ethnicity and Season

	Test Day Temperature					90-Day Temperature				
Reading	**Black**	**Hispanic**	**Native American**	**Multiple**	**White**	**Black**	**Hispanic**	**Native American**	**Multiple**	**White**
					Fall					
Average Temperature	-0.020	0.006	-0.064	0.078	0.039	0.967***	0.991***	1.007***	1.362***	1.168***
	(0.030)	(0.024)	(0.068)	(0.047)	(0.029)	(0.157)	(0.126)	(0.368)	(0.254)	(0.158)
p: Test of equality			0.261					0.646		
N	17,085	25,810	3,576	7,142	17,037	17,085	25,810	3,576	7,142	17,037
					Winter					
Average Temperature	-0.012	-0.009	0.009	0.003	-0.017	-0.029	-0.025	-0.136	0.037	-0.130***
	(0.011)	(0.009)	(0.025)	(0.018)	(0.011)	(0.040)	(0.034)	(0.093)	(0.067)	(0.042)
p: Test of equality			0.814					0.125		
N	26,656	40,140	5,577	11,389	27,362	26,656	40,140	5,577	11,389	27,362
					Spring					
Average Temperature	0.015	-0.060***	0.006	-0.060*	-0.052***	-0.026	-0.077	-0.027	0.049	-0.028
	(0.020)	(0.016)	(0.044)	(0.032)	(0.020)	(0.060)	(0.048)	(0.134)	(0.096)	(0.057)
p: Test of equality			0.028					0.816		
N	21,637	30,965	4,486	8,826	21,894	21,637	30,965	4,486	8,826	21,894

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by student in parentheses below coefficient estimate. Each coefficient from a separate regression estimated via OLS predicting student MAP percentile. In addition to the measure of average temperature, all regressions contain a measure of student's MAP percentile in the same subject from the prior administration and fixed effects for the number of days since the prior MAP administration.

Figure 5. Estimated Effect of Temperature on the Distribution of MAP Performance, by Temperature Measure, MAP Administration, and Subject



Notes: Each panel of the figure plots point estimates (solid line) and 95% confidence intervals (dashed lines) of the effect of test-day temperature (top row) or 90-day temperature (bottom row) on the probability that that students' MAP percentile exceeds the value on the x-axis, where x ranges from 1 to 99 in increments of 1. This results in 99 total regressions. In each regression, we model the outcome as a function of the temperature measure, a fixed effect for a student's percentile on the prior administration, and the number of days since the prior MAP administration.

Table 6. Effect of Average Temperature on Change in MAP Percentile From Prior MAP Administration, by Zip Code Income, Temperature Measure, Subject, and Season

	Math				Reading			
	Test Day		90-Day		Test Day		90-Day	
	Below Median Income	Above Median Income	Below Median Income	Above Median Income	Below Median Income	Above Median Income	Below Median Income	Above Median Income
	Fall							
Average Temperature	-0.030	-0.0003	0.946***	0.531***	0.013	0.014	1.110***	0.942***
	(0.019)	(0.020)	(0.110)	(0.115)	(0.022)	(0.020)	(0.108)	(0.111)
p: Test of equality	0.290		0.014		0.988		0.284	
N	36,282	35,265	36,282	35,265	36,571	35,400	36,571	35,400
	Winter							
Average Temperature	-0.004	0.0001	0.044	0.028	-0.017**	-0.006	-0.051*	-0.069**
	(0.008)	(0.008)	(0.028)	(0.029)	(0.008)	(0.008)	(0.028)	(0.029)
p: Test of equality	0.709		0.692		0.300		0.649	
N	56,452	55,581	56,452	55,581	57,229	56,024	57,229	56,024
	Spring							
Average Temperature	-0.116***	-0.077***	0.187***	0.170***	-0.053***	-0.018	-0.072*	-0.049
	(0.013)	(0.013)	(0.039)	(0.040)	(0.014)	(0.014)	(0.040)	(0.041)
p: Test of equality	0.037		0.762		0.078		0.697	
N	44,727	44,192	44,727	44,192	45,293	44,160	45,293	44,160

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by student in parentheses below coefficient estimate. Each coefficient from a separate regression estimated via OLS predicting student MAP percentile. In addition to the measure of average temperature, all regressions contain a measure of student's MAP percentile in the same subject from the prior administration and fixed effects for the number of days since the prior MAP administration.

Table 7. Effect of Average Test-Day Temperature on Change in Student MAP Percentile, by Alternative Specification, Season, and Subject

	MAP Test Percentile					
	Fall		Winter		Spring	
	Math	Reading	Math	Reading	Math	Reading
	Student Fixed Effect					
Test-day Temperature	-0.114***	-0.032**	0.056***	-0.024***	-0.119***	-0.120***
	(0.014)	(0.014)	(0.006)	(0.006)	(0.009)	(0.009)
N	71,643	72,070	112,490	113,434	89,033	89,033
	Both Temperature Measures					
Test-day Temperature	-0.072***	-0.116***	-0.001	-0.006	-0.102***	-0.037***
	(0.015)	(0.016)	(0.006)	(0.006)	(0.009)	(0.010)
N	71,642	72,069	112,490	113,411	89,033	89,567
	Lag & School Fixed Effect					
Test-day Temperature	-0.071***	0.046***	-0.009	-0.015***	-0.079***	-0.029***
	(0.014)	(0.015)	(0.006)	(0.005)	(0.009)	(0.010)
N	71,642	72,069	112,490	113,411	89,033	89,567

NOTE: **p*<0.10, ***p*<0.05, ****p*<0.01. Robust standard error clustered by student in parentheses below coefficient estimate. Each coefficient from a separate regression estimated via OLS predicting student MAP percentile. In addition to the measure of average test-day temperature, all regressions in the "Both Temperature Measure" and "Lag & School Fixed Effects" panels contain a measure of student's MAP percentile in the same subject from the prior administration and fixed effects for the number of days since the prior MAP administration.

Table 8. Effect of Average 90-Day Temperature on Student MAP Percentile, by Alternative Specification, Season, and Subject

	MAP Test Percentile					
	Fall		Winter		Spring	
	Math	**Reading**	**Math**	**Reading**	**Math**	**Reading**
	Student Fixed Effect					
90-Day Temperature	0.869***	1.204***	-0.104***	-0.148***	0.433***	0.297***
	(0.069)	(0.064)	(0.007)	(0.008)	(0.025)	(0.025)
N	71,643	72,070	112,490	113,434	89,033	89,567
	Both Temperature Measures					
90-Day Temperature	0.998***	1.402***	0.061***	-0.033	0.212***	-0.020
	(0.083)	(0.087)	(0.020)	(0.022)	(0.028)	(0.029)
N	71,642	72,069	112,490	113,411	89,033	89,567
	Lag & School FE					
90-Day Temperature	0.647***	1.015***	0.009	-0.090***	0.163***	0.010
	(0.078)	(0.077)	(0.020)	(0.021)	(0.029)	(0.029)
N	71,642	72,069	112,490	113,411	89,033	89,567

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by student in parentheses below coefficient estimate. Each coefficient from a separate regression estimated via OLS predicting student MAP percentile. In addition to the measure of average 90-day temperature, all regressions in the "Both Temperature Measure" and "Lag & School Fixed Effects" panels contain a measure of student's MAP percentile in the same subject from the prior administration and fixed effects for the number of days since the prior MAP administration.