EdWorkingPaper No. 24-910

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

Deven Carlson
University of Oklahoma

Adam Shepardson
University at Albany

As students are exposed to extreme temperatures with ever-increasing frequency, it is important to understand how such exposure affects student performance. Drawing on detailed Measures of Academic Progress (MAP) achievement data combined with high-resolution weather records, we estimate the effect of temperature on student test performance across a six-year period for students in Tulsa, Oklahoma. Test-day heat reduces student MAP scores in the fall and spring, particularly in math. Relative to a day in the 70s, a test day reaching the 90s reduces math scores by more than 1.5 percentiles, or roughly 0.06 standard deviations, with smaller declines in reading. Heat effects emerge at lower temperatures in the spring than in the fall and effects are larger for male students.

VERSION: September 2026

Suggested citation: Carlson, Deven, and Adam Shepardson. (2026). Under the Weather? Seasonal 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? Seasonal Effects of Temperature on Student Test Performance

Deven Carlson
University of Oklahoma

Adam Shepardson
University at Albany

Abstract
As students are exposed to extreme temperatures with ever-increasing frequency, it is important to understand how such exposure affects student performance. Drawing on detailed Measures of Academic Progress (MAP) achievement data combined with high-resolution weather records, we estimate the effect of temperature on student test performance across a six-year period for students in Tulsa, Oklahoma. Test-day heat reduces student MAP scores in the fall and spring, particularly in math. Relative to a day in the 70s, a test day reaching the 90s reduces math scores by more than 1.5 percentiles, or roughly 0.06 standard deviations, with smaller declines in reading. Heat effects emerge at lower temperatures in the spring than in the fall and effects are larger for male students.
Keywords: Seasonal heterogeneity, Student achievement, Temperature

Introduction

An ever-growing body of research documents that the natural environment can play a substantial role in shaping students' achievement and attainment outcomes (e.g., Austin et al. 2019; Heissel et al. 2022; Persico & Venator 2021; Wen & Burke 2022). Within this broad literature, a series of studies provide compelling evidence that local weather conditions, particularly temperature, affect educational outcomes (e.g., Graff Zivin et al. 2018; Park 2022; Park et al. 2020; Park et al. 2021). And 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 achievement. Although prior work has made notable progress on this topic, data limitations and the potential for 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 thorough, localized weather records to paint a clear portrait of the effect of temperature on student test performance 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 Measures of Academic Progress (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 Growth assessment three times per school year, once in the fall (i.e., Aug.-Sept.), a second time in the winter (i.e., Nov.-Jan.), and a third time in the spring (i.e., Apr.-May) in school buildings with full climate control capabilities.[i] 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 throughout 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).

We use these detailed, longitudinal data to conduct two main sets of analyses. First, we estimate the effect of test-day temperature on student MAP performance. These analyses exploit the fact that, in each season, the MAP administration period in TPS spanned multiple weeks, with schools administering the assessment on different days across this period. The different administration dates generated variation in the test-day weather that students experienced, which we use to identify the effect of test-day temperature on student performance. Notably, we estimate these effects separately for fall, winter, and spring. Second, we apply a similar approach to estimating the effects of temperatures over the three months leading up to test days in each season. In particular, we estimate the effects of a marginal day with a maximum temperature in each of nine ten-degree temperature bins—these bins span maximum temperatures in the teens all the way through those exceeding 100 degrees—relative to a day with a max temp in the 70s. Together, the first set of analyses speaks to the effects of temperature on test performance—fundamentally a question of measurement—whereas the second set of analyses bears on whether sustained temperature exposure affects learning itself.

We highlight five main findings that emerge from these analyses. First, test-day heat reduces students' MAP performance in both math and reading, with the estimated effects larger in math than in reading. In math, we estimate that a day where the max temperature reaches the 90s reduces student MAP performance by more than 1.5 percentiles, which equates to about 0.06 standard deviations, compared to a day with a max temp in the 70s. For context, the full-sample standard deviation of the maximum test-day temperature is roughly 16 degrees, so a move from the 70s to the 90s represents approximately a one standard deviation increase. In reading, the analogous effects are slightly larger than a one-half percentile decline, or about 0.02 standard deviations. Second, we show that, in math, the threshold at which test-day heat decreases MAP scores is lower in the spring than it is in the fall. Compared to a day with a max temp in the 70s, a day that maxes out in the 80s reduces student scores on the spring MAP administration, but has no effect on student scores in the fall—scores in that season decline only when the max temperature exceeds 90 degrees. Third, we provide evidence that test-day heat reduces the scores of male students more than it does female students, particularly in math. Fourth, we provide evidence that test-day heat reduces the amount of time that students spend taking the MAP assessment and explore whether test duration may be a mechanism by which the effects of heat operate. Fifth, our analyses return relatively little evidence of systematic temperature effects over the three-month period leading up to the MAP assessment, although there are indications that very cold days (maximum temp below 30 degrees) reduce student scores on the winter MAP assessments.

Our work builds on prior literature in multiple ways. First, our study is unique in its ability to separately estimate temperature effects on student test performance in the fall, winter, and spring. And the resulting evidence of seasonal heterogeneity in temperature effects has important implications for our understanding of the topic. For example, from the standpoint of practice, this seasonal heterogeneity suggests that inferences regarding student learning gains are likely sensitive to the temperature on the day of the spring MAP administration, particularly given relatively lower sensitivity to test-day temperatures in the fall, when the performance baseline is established. Second, our ability to observe students' MAP test duration sheds light on potential mechanisms by which the effects of test-day temperature may operate. Specifically, we show that conditioning on test duration reduces, but does not fully eliminate, the negative effects of test-day heat, which is consistent with these effects operating via both behavioral and physiological/cognitive mechanisms. Third, our estimation of meaningful test-day temperature impacts alongside the general absence of systematic longer-term temperature effects informs discussions of the impact of temperature on student test performance versus learning more broadly. Our results provide clear evidence that temperature affects test performance, but little evidence that it affects learning, at least in our empirical context. This distinction is key as it suggests that temperature can distort the measurement of achievement while having little effect on the underlying learning that the achievement measures are intended to capture.

All told, our results contribute important depth and nuance to our understanding of the relationship between temperature and student test performance. In doing so, this work speaks to the validity of the conclusions that researchers and policymakers draw from test scores, which are generated under uneven environmental conditions that lie beyond the control of students and educators.

Background and Context

Temperature and Student Achievement: The Role of Physiological, Behavioral, and Environmental Factors

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 performance and learning. The literature on test-day temperature reaches conflicting 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 et al. 2018), primary school-level math and reading exams in India (Garg, Jagnani, and Taraz 2020), 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 & Patel 2021). Studies of longer-term temperatures exhibit a similar dynamic, with long-term heat exposure reducing scores on the Programme for International Student Assessment (PISA) (Park et al. 2021) and college entrance exams in South Korea (Cho 2017), but only producing small—and imprecisely estimated—score reductions for U.S. students taking the Peabody assessment (Graff Zivin et al. 2018). And Johnston et al.'s (2021) work in Australia found no evidence that hot days leading up to National Assessment Program literacy and numeracy tests reduced scores on those assessments, but more recent work from the Australian context has provided evidence of negative long-term heat impacts (Srivastava et al. 2025).

On the surface, the empirical literature linking student test performance to temperature seems varied and, at times, contradictory. 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 (Coehoorn et al. 2020; Froom et al. 1993; Yi et al. 2021), with children "more susceptible to heat illness" than adults (Hyndman et al. 2023, p. 23-24). These physiological disruptions, which motivate most existing empirical work on temperature and education, 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. For example, Melo & Suzuki (2023) find a meaningful, negative heat impact on low-stakes college entrance exams in Brazil but substantially smaller effects on high-stakes exams, suggesting that student effort may mediate performance in hot conditions.

Importantly, though, individuals can also become acclimated to environmental conditions after extended exposure (Castellani & Young 2016; Périard et al. 2015). Such acclimatization has two related, yet distinct, implications for analyses of the effect of temperature on test performance. First, acclimatization may, to some degree, insulate individuals in relatively warm climates from the effects of heat on test performance, a dynamic that could produce geographic heterogeneity in the effect of temperature on student test performance. A small amount of prior empirical work provides support for this phenomenon. 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 & Whitney 2022), with a similar effect found for American players of an online arithmetic game (Krebs 2024). And recent work attributes the negative impacts of cold weather on Australian students' test performance to the heat, but not cold, acclimatization in the country's warmest regions (Johnston et al. 2021; Srivastava et al. 2025), with testing in colder regions more strongly impacted by heatwaves during school years (Srivastava et al. 2025). Second, acclimatization provides a conceptual basis for expecting the effects of heat—and temperature more broadly—to vary by season. In locations with meaningful seasonal variation in temperature, students may become acclimated to heat throughout the course of the summer, suggesting that hot temperatures during the spring may have a larger effect on student test performance and learning than similar heat in the fall. In general, these considerations illustrate the usefulness of considering how acclimatization, or its absence, might moderate the impact of temperature on student test performance.

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 (Fan et al. 2023; Graff Zivin & Neidell 2014; Heaney et al. 2019), 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. Weather can also motivate behavioral changes among students, such as hot and cold temperatures encouraging absenteeism and, specifically for heat, disciplinary infractions (McCormack 2023). It is self-evident that other temperature effects exist,[ii] from snow days, to outdoor extracurricular activity cancellations, to district policies specifying the weather conditions that trigger indoor recess—weather clearly shapes how students spend their time. Hyndman & Vanos (2023) observe that extreme weather events can create logistical hurdles to administering education, which may disrupt routines that contribute to students' schooling engagement. These factors underscore the importance of considering how temperature might affect students' test scores through the mechanism of time allocation leading up to assessments.

Finally, the built environment interacts with weather conditions in ways that, theoretically, can both mitigate and exacerbate the impacts of weather on student performance and learning. In terms of mitigation, classrooms across the country typically have climate control capabilities—heat and air conditioning—and there is evidence that these capabilities can significantly reduce the effects of temperature on student test scores (Park et al. 2020) and student behavior (McCormack 2023). 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 contributes to the Urban Heat Island effect (Mohajerani et al. 2017), a phenomenon in which urban areas exhibit temperatures several degrees higher than those observed in outlying areas (Oke 1982). Tall buildings, meanwhile, can amplify wind, resulting in lower wind chills and greater levels of discomfort during the winter (Soligo et al. 1998). These factors render schools in urban areas especially vulnerable to experiencing "uncomfortable microclimates" driven by the abundance of artificial surfaces (e.g., asphalt playgrounds) (Hyndman et al. 2023, p. 21). Together, the combination of physiological, behavioral, and physical considerations forms a strong theoretical foundation for expecting the relationship between weather and test scores to vary across geography.

Weather and Student Achievement in Tulsa, Oklahoma

Tulsa, Oklahoma serves as the setting for our empirical inquiry into the effects of weather on student test-day performance and longer-term learning. Several features of the Tulsa context render it well-suited for this purpose. First, 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 setting for analyzing temperature effects. 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 test performance.

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 max temperatures of 93-94 in July and August, with almost 75 days with a max above 90 and 11 days that hit the century mark (NWS 2024). 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 max temperature 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" given that April and May often bring severe thunderstorms and the associated hazards of tornados, hail, and damaging straight-line winds.

Tulsa's climatological profile interacts in important ways with the physiological, behavioral, and physical considerations discussed above. Physiologically, the excessive summer heat could inhibit student cognition and knowledge acquisition and retention, but such an impact is mitigated by three major factors. First, 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 nearly all residences had cooling capabilities—recent data from the U.S. Energy Information Administration (2023) indicate that 96 percent of Oklahoma households use air conditioning. Although these features of the built environment help minimize the impact of excessive heat on student learning (Park et al. 2020), the energy poverty literature demonstrates that the quality and use of residential climate control systems varies by household income level (e.g., Bednar & Reames 2020; Romitti et al. 2022). These inequalities may impede students' test performance through channels such as sleep loss from nighttime heat exposure leading up to exams (Obradovich et al. 2017). As we describe in greater detail below, this possibility motivates an analysis of heterogeneity in the effects of temperature by neighborhood income level. 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, hot temperatures are also possible during actual test administrations in fall and spring, and we highlight two considerations relevant to these seasons. First, as noted above, all TPS schools and almost all student residences have air conditioning, although the quality of residential cooling capabilities likely varies across TPS households (Romitti et al. 2022). 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 on either performance or learning manifest in these seasons, they may be more likely to occur in spring than in fall. On the flip side, it is more common for Tulsa to experience very hot temperatures during the fall administration period than during the spring.

For winter MAP administrations, many of the same considerations outlined above are relevant, including climate control capabilities in schools and residences, as well as possible negative impacts of outdoor cold temperatures on indoor exam performance despite such climate control (Cook & Heyes 2020). 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 Empirical Strategy

Data

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.[iii] 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.

The records also contain information about the test 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 spent taking the test. Our empirical analyses use students' percentile score—normed off the national population of test-takers in each grade and season—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 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 each day. We complement this Mesonet data with air pollution records from the Environmental Protection Agency (EPA)'s monitoring station in Tulsa, which provides daily measures of PM2.5.

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 continuously operating stations within 50 miles of Tulsa, we obtain the archived daily summary records of average daily air temperature and daily maximum and minimum air temperature. We pull 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 overlay 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[iv] (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 locations of included 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 residential 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 core weather measures used in our test-day and three-month analyses below: 1) A series of variables indicating the ten-degree temperature bin containing the maximum test-day temperature (i.e., 50-59 degrees, 60-69 degrees, etc.), and 2) A series of variables counting the number of days during the prior three-month period when the maximum daily temperature fell within each ten-degree bin.

As the final step in creating our analytic dataset, we merge the zip code-by-test day weather measures 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 temperature and binned counts of maximum daily temperature over the three months prior to each student MAP administration. Table 1 presents descriptive statistics on the students comprising 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. About 15 percent of students identify as either 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 smaller due to the COVID pandemic precluding spring administrations in 2020.

[Insert Tables 1 and 2 about here]

Table 2 presents descriptive statistics for the temperature and test-based outcome measures underlying our analysis, both overall and separately by season. For test-day temperature, the table shows, across all seasons, an average maximum daily temperature of 72.5 degrees Fahrenheit. This overall average masks significant seasonal variation, however, with average maximum temps above 86 degrees on the day of the fall MAP administration and below 55 degrees in winter—the average maximum temperature for the spring administration is about 77 degrees. The ten-degree test-day temperature bins further illustrate the variability of Tulsa's climate, with max temperatures ranging from the 20s to above 100 degrees. For fall administrations, maximum temperatures are typically in the 70s, 80s, or 90s while maximum temperatures for spring administrations generally span the 60s to the 80s. Maximum temperatures are most variable for the winter administrations, which commonly see maxes from the 30s all the way up to the 70s. The three-month temperature statistics illustrate broad similarity in the temperature profiles leading up to winter and spring administrations, with the period preceding fall administrations notably hotter—most days are in the 80s or 90s at peak daily temperatures. Finally, the bottom panel of the table illustrates that students score at about the 40th percentile in each season, with performance slightly higher in the spring than in the fall or winter. Students spent an average of 42 minutes on the fall MAP test, with that number increasing to 46 minutes in winter and spring.

Estimating the Effect of Test-Day Temperature on Student Performance

Empirically, we begin by estimating the effect of test-day temperature on student MAP performance. As described earlier, each MAP administration season in TPS spanned multiple weeks, with schools administering the assessment on different days across the period. The different administration dates generated meaningful variation in the weather students experienced on the day they took the test, which we use as the basis for estimating the effect of test-day temperature on student MAP performance. We leverage this variation to first estimate the effect of test-day temperature on student MAP performance using a model of the form:

$$P_{icsgt} = \sum_{b=1}^{b=7} \gamma_b M_{icst} + \sum_{s=1}^{s=3} \delta_s D_{ist} + \sum_{s=1}^{s=3} W_{ist} \phi_s + \rho_{ic} + \theta_s + \lambda_{gt} + \varepsilon_{icsgt} \quad (1)$$

where we specify the outcome measure as the MAP percentile, *P*, for student *i* attending school *c* at the time of the MAP administration in season *s* during grade *g* in school year *t*. Our treatment specification, $\sum_{b=1}^{b=7} \gamma_b M_{icst}$, consists of a series of dummy variables indicating the ten-degree temperature bin containing the maximum temperature on the day the student took the MAP test. The first of the seven temperature bins, which we index with *b*, encompasses test days with a maximum temperature below 40 degrees while the seventh bin captures days with max temps of 90 or above—the remaining five bins capture days that max out in the 40s, 50s, 60s, 70s, and 80s, respectively. We specify the reference category as test days with a max temp in the 70s, resulting in coefficient γ_b representing the difference in student test performance on days where the maximum temperature falls in bin *b*, relative to performance on days that max out in the 70s.

In addition to the treatment specification, the model contains a control for the number of days, *D*, between the test date and a fixed point of each seasonal MAP administration—the beginning of the school year for fall, the end of winter break for winter, and the end of the school year for spring. We also include controls for air pollution levels, average humidity, and cumulative precipitation on the test day,[v] which are represented by W. The coefficients on all these controls are allowed to vary by season. Finally, the model contains fixed effects for student-by-school, ρ_{ic}, grade-by-school year, λ_{gt}, and season, θ_s; ε_{icsgt} represents the error term. We include *D* to account for differences across schools in MAP administration dates and, thus, the number of instructional days students are exposed to prior to sitting for the MAP. Separately for reading and math, we estimate this model via ordinary least squares (OLS) with standard errors clustered at the school-by-test day level. The empirical model in equation (1) constrains the estimated effect of test-day temperature from varying across seasons. In the prior section, however, we discussed how seasonal differences in student acclimatization, among other factors, provide a conceptual basis for expecting temperature effects to vary across seasons. Accordingly, we extend the specification in equation (1) by estimating season-specific temperature effects with a model of the form:

$$P_{icsgt} = \sum_{s=1}^{s=3}\left(\sum_{b=1}^{b=7}\gamma_{bs}M_{icst}\right) + \sum_{s=1}^{s=3}\delta_s D_{ist} + \sum_{s=1}^{s=3}W_{ist}\phi_s + \rho_{ic} + \theta_s + \lambda_{gt} + \varepsilon_{icsgt} \qquad (2)$$

where we specify the series of ten-degree temperature bins separately for each of the three MAP administration seasons. To illustrate, the model contains three separate bins for days with maximum temperatures in the 60s, one for fall test days, a second for winter, and the third for spring. We again specify the reference category as test days with max temps in the 70s, a choice motivated by max daily temperatures falling into this bin in each of the three MAP administration seasons.[vi] The remaining contents of the model are identical to those in equation (1) and we again estimate the model via OLS separately by subject.

Results: Effect of Test-Day Temperature on Student Performance

Table 3 and Figure 2 present the estimated effects of maximum test-day temperature on student MAP performance resulting from estimation of equation (1)—Figure 2 plots the coefficients in Table 3 to provide a clear visual presentation of performance patterns. As noted above, we specify the 70-degree bin as the reference category in our models as it allows for a common reference point across all three seasons. Starting with math, the results show that, compared to days with max temps in the 70s, student performance declines as max temperatures rise into the 80s and, especially, the 90s. Our results show that students score about 0.5 and 1.5 percentiles lower on these days, respectively—these differences are statistically significant. To facilitate comparison of these effects with the broader literature, days with max temps in the 80s reduce student performance by about 0.02 standard deviations, compared to days with max temps in the 70s, and a day in the 90s reduces performance by about 0.06 standard deviations.[vii] As noted above, the full-sample standard deviation of the maximum test-day temperature is roughly 16 degrees, so a change in temperature from the mid-70s to the low 90s represents about a one standard deviation increase. Accordingly, these effects are meaningful in magnitude, in the range of what Kraft (2020) would deem “medium” impacts in his empirical benchmarking of effect sizes of educational interventions.

[Insert Table 3 and Figure 2 about here]

In addition to heat-driven reductions in student math performance, the results also indicate that students perform worse on math tests during very cold days, relative to days with maximum temperatures in the 70s. More specifically, average student performance declines by about half a percentile on days that fail to reach 40 degrees. This difference, which corresponds to about 0.02 standard deviations, is statistically significant at $p<0.05$. Together, the math results in Table 3 make clear that, relative to days in the 70s, student performance is reduced by both cold and, especially, hot test days.

The reading results are broadly similar to those for math, although there are a couple notable differences. First, whereas student math performance declined on cold days, relative to days with max temps in the 70s, the reading results demonstrate improved MAP performance on days with a maximum temperature below 40 degrees. And the magnitude of the performance increase—approaching a percentile—is statistically significant. Second, although the reading results also show performance declines on hot days, they are only apparent for days that reach the 90s and, at about 0.6 percentiles, are more modest in magnitude. Broadly, though, the math and reading results are more similar than different.

[Insert Figure 3 about here]

Table A1 and Figure 3 present results from estimating the model in equation (2), which returns the effect of binned maximum test-day temperature separately for each of the three MAP administration seasons. Table A1 presents coefficients and standard errors for all estimable bins. Seasonal temperature variation results in some bins being sparsely populated in specific seasons, with impacted bins demonstrating coefficients estimated with very little precision. To prevent such imprecision from overwhelming the coefficient plots, Figure 3 only presents coefficient estimates and confidence intervals for bins containing at least one percent of the maximum test-day temperatures in the corresponding MAP administration season. We again estimate the seasonal specifications from Table A1 and Figure 3 relative to test days with maximum temperatures in the 70s, as all seasons have numerous test days captured by that bin.

The estimates presented in Figure 3 add important nuance to the overall results.[viii] We highlight four main takeaways. First, the results make clear that the math performance decline on

80-89 degree days previously seen in Table 3 is driven entirely by the spring MAP administration. In math, students score about a percentile lower on spring days that max out in the 80s, compared to days in the 70s. These results stand in sharp contrast to those for fall, where the point estimates for 80-89 degree days are close to zero and statistically insignificant. These seasonal differences are reinforced by results of a formal test of equality presented in Table A1, which rejects the null hypothesis that the estimated effects of 80-89 degree days in the fall and spring are equal to one another.

Second, while days in the 80s have minimal impact on student performance in the fall, Figure 3 demonstrates that 90+ degree days drive performance declines in this season, with larger effects for math than reading. Specifically, the results indicate that fall days in the 90s (or above) reduce student math performance by a bit more than half a percentile, relative to a test day in the 70s, but have little impact on reading performance.[ix] Third, in contrast to the pronounced seasonal variation in the effects of 80-89 degree days, there is little meaningful variation in the effects of days with max temps in the 50s or 60s, compared to a day in the 70s. Fourth, Figure 3 makes clear that, unsurprisingly, the effects for the relatively cooler ten-degree temperature bins—the 30s and 40s—are primarily driven by the winter MAP administration. Thus, the overall results for bins at the lower end of the temperature spectrum broadly reflect the winter results. Together, the results in this section tell a clear, if nuanced, story about the role of test-day temperature in shaping student performance—warm temperatures reduce student MAP scores, with the size of these effects larger in spring than in fall.

Estimating the Effect of Three-Month Temperature on Student Performance

To this point, our analyses provide evidence that test-day temperature can affect student MAP performance. The effects of temperature, however, may not be confined to the day that students sit for the assessment. Prior work suggests that, through both behavioral and physiological mechanisms, temperature has the potential to affect student learning over the course of the school year (Park et al. 2020), with these effects manifesting on students' test performance. Several features of our data—the six-year time span, the seasonal administration of the MAP assessment, and the detailed weather measures—position us well to estimate any longer-term temperature impacts. In our analyses below, we operationalize longer-term temperature as the maximum daily temperatures over the three months leading up to each test day, including weekends and breaks. This window reflects the seasonality of MAP, with students' successive MAP administrations separated by about three months. In effect, our measure is designed to broadly capture the temperatures students have experienced since their prior MAP administration.

The middle panel of Table 2 above presents descriptive statistics for our three-month temperature measures. Specifically, it presents the mean of the average daily temperatures over the prior three months, as well as the average number of days in that period with a maximum temperature falling within each ten-degree temperature bin. The table illustrates that, across all MAP administrations, max temperatures in the 50s through 90s were relatively common in the three months leading up to test administration—each bin in this range contains, on average, more than ten days where the maximum temperature fell within the bin. However, the table also makes clear that there were days with max temps as low as the teens and as high as the 100s. The second, third, and fourth panels of Table 2 highlight seasonal variation in temperatures preceding MAP administration. In the fall, the three-month period leading up to testing was dominated by max temps in the 80s and 90s, but winter and spring administrations saw more varied temperature distributions. During the winter period, a plurality of days had max temps in the 60s, though they ranged from the 40s to the 80s. Prior to the spring MAP, students most frequently experienced days that maxed out in the 50s, 60s, and 70s, with occasional days outside this range.

We estimate the effects of three-month temperature on student MAP performance using a model conceptually similar to the one we use to estimate test-day temperature effects. Specifically, we estimate:

$$P_{icsgt} = \sum_{h=1}^{h=10} \varphi_h N_{icst} + \sum_{s=1}^{s=3} \delta_s D_{ist} + \sum_{b=1}^{b=7} \gamma_b M_{icst} + \sum_{s=1}^{s=3} W_{ist} \phi_s + \rho_{ic} + \theta_s + \lambda_{gt} + \varepsilon_{icsgt} \quad (3)$$

where we again specify the outcome measure as the MAP percentile, *P*, for student *i* attending school *c* at the time of the MAP administration in season *s* during grade *g* in school year *t*. Our treatment specification, $\sum_{h=1}^{h=10} \varphi_h N_{icst}$, consists of a series of variables counting the number of days over the three months preceding MAP administration with a maximum temperature in the ten-degree temperature bin, which we index with h. The first of these variables counts the number of days that max out in the teens while the tenth bin contains a count of the days with a maximum temperature exceeding 100—the remaining eight bins count the number of days with max temps in the 20s through 90s, respectively. We specify the reference category as days with a maximum in the 70s, resulting in coefficient φ_h representing the effect of an additional day during the three months preceding MAP administration with a max temperature in bin h, relative to a day with a max in the 70s. Importantly, we control for test-day temperature via a measure mirroring that in equation (1). Further, the $\sum_{s=1}^{s=3} W_{ist} \phi_s$term includes covariates measuring the respective means of daily average air pollution, daily total precipitation, and daily average humidity over the three-month period. The remaining contents of the model are identical to those in equations (1) and (2), above.

Our analyses of test-day temperature demonstrated clear variation in the effects across MAP administration seasons. These empirical results, coupled with the conceptual considerations outlined above, point to the potential for seasonal variation in the effects of three-month temperature as well. Accordingly, we estimate:

$$P_{icsgt} = \sum_{s=1}^{s=3}(\sum_{h=1}^{h=9} \varphi_{hs} N_{icst}) + \sum_{s=1}^{s=3} \delta_s D_{ist} + \sum_{b=1}^{b=7} \gamma_b M_{icst} + \sum_{s=1}^{s=3} W_{ist} \phi_s + \rho_{ic} + \theta_s + \lambda_{gt} + \varepsilon_{icsgt} \quad (4)$$

where we specify the variables counting the number of days over the three months preceding MAP administration with a maximum temperature in the ten-degree temperature bin separately for each season.[x] The remaining contents of the model are identical to those in equation (3) and we again estimate the model via OLS separately by subject.

Results: Effect of Three-Month Temperature on Student Performance

Table A2 and Figure 4 present the effects of three-month temperatures derived from estimating equation (3)—Figure 4 plots the coefficients in Table A2 to provide a clear visual presentation of performance patterns. Starting with math, the results suggest a marginal pre-test day with a max temp below 30 degrees in the three months leading up to the MAP assessment reduces student test performance by a bit more than half of a percentile. However, the relative infrequency of these days results in imprecisely estimated coefficients, with the confidence intervals extending very close to zero. Beyond the negative effects of cold days, there is little evidence of systematic effects across other temperature bins; estimates on both sides of the reference category—a day with a max temp in the 70s—are close to zero and few of the estimates are statistically different from one another. The story is broadly similar for reading, with nearly all of the estimates very close to zero and statistically insignificant.

[Insert Figure 4 about here]

Table A3 and Figure 5 present results from estimating the model in equation (4), which returns the season-specific effect of an additional day during the three months preceding MAP administration with a max temperature in bin *h*, relative to a day that tops out in the 70s. Importantly, our three-month temperature measure counts the number of days in each temperature bin over a fixed window but does not capture the timing of those occurrences. Because a given temperature's typical position in the window can differ across seasons, these coefficients potentially reflect not only the occurrence of a temperature, but also its timing relative to the assessment—future work might attempt to empirically partition these two factors.

Starting with fall, the results provide little evidence that longer-term temperature affects student MAP performance—all estimates in both reading and math are insignificant relative to a 70-79 degree day. Turning to winter, the math results indicate that an additional day with a max temp below 30 in the three month lead-up to the MAP meaningfully reduces student performance—a pattern consistent with snow-induced school cancellations (e.g., Goodman 2014)—but there is little evidence of an effect across other temperature bins. The reading results broadly align with those for math. Finally, the spring results in both math and reading provide little indication of systematic temperature effects in the three months leading up to MAP administration. Together, the three-month results provide some evidence that cold days in the lead-up to the winter MAP administration reduce student scores, but the general takeaway is one of few systematic temperature effects.

[Insert Figure 5 about here]

Supplemental Analyses and Robustness Checks

Heterogeneity in the Effects of Test-Day Temperature on Student Performance

Prior work provides a basis for exploring whether the effects of temperature on student performance vary by student sex, where physiological differences between males and females are a possible source of heterogeneity—this possibility has been empirically supported in a controlled lab setting (Chang & Kajackaite 2019) and is consistent with multiple studies that link heat or cold exposure to student test scores by sex (e.g., Cook & Heyes 2020; Johnston et al. 2021; Srivastava et al. 2025), although other research finds little evidence of such patterns (e.g., Cho 2017; Park et al. 2020; Li & Patel 2021). Separately, we also recognize that near-universal access to air conditioning in Oklahoma households does not imply universally high-quality air conditioning, which may be more prevalent in higher income areas. We examine the possibility that temperatures' effects on student performance vary by income, drawing upon annual Internal Revenue Service (IRS) income statistics at the zip code-level[xi] and flagging students residing in zip codes below ("low-income") and above ("high-income") the sample median.

[Insert Figures 6 and 7 about here]

Figure 6 presents results from estimating the model shown in equation (1) separately for boys and girls, mirroring the coefficients in Table A4. Overall, we find the results for male and female students to be more alike than different. However, similar to Chang and Kajackaite (2019), there is some evidence that girls' scores are generally less affected by heat than their male peers, particularly in math. Relative to a day with a maximum temperature in the 70s, testing on a day with a max of 90 or above is estimated to reduce boys' scores by more than two percentiles, on average, compared to less than a 1.5 percentile reduction for girls. Season-specific analyses corroborate that, in math, boys' greater heat sensitivity is apparent on both the fall and spring MAP administrations.[xii]

Regarding income, we do not observe significant differences in temperature effects between students residing in lower- vs. higher-income zip codes (see Figure 7 and Table A5). Although our income measure exhibits substantial cross-zip code variation—about $28,000 vs. $105,000 at the 10th and 90th percentiles, respectively—we do not observe either within-zip code variation or, more consequentially, student-level household income. Accordingly, our data are imperfectly suited to analyzing income-based heterogeneity, which would be a valuable focus for future work able to draw upon high-quality student-level income measures.

Effect of Test-Day Temperature on Student Test Duration

Our main results make clear that test-day heat in the fall and spring reduces student performance. To gain insight into potential mechanisms generating these impacts, we take advantage of information in our data measuring the amount of time that students spent taking the MAP assessment. Specifically, we estimate a variant of the model in equation (1) where we specify the outcome as the number of minutes the students spent taking the test.

[Insert Figure 8 about here]

We present the results of this analysis in Figure 8, which correspond to coefficients from Table A6. The estimates make clear that, in both reading and math, test-day heat reduces the amount of time students spend taking the MAP test. Specifically, the results indicate that, relative to a day with a max temp in the 70s, an 80-89 degree day reduces student test duration on both subjects by, on average, an imprecisely estimated 30 seconds while a 90+ degree day results in a reduction of more than two minutes. For context, students spent an average of about 45 minutes on each subject of the assessment. The season-specific estimates make clear that test-day heat reduces test duration across spring and fall MAP administrations.

[Insert Figure 9 about here]

To gain some initial insight into the degree to which these temperature-driven changes in test duration could be responsible for the estimated effects of temperature presented above, we estimate a variant of equation (1) that contains a covariate measuring student test duration. We present results from estimating this model in Figure 9 (which reflects Table A7). To facilitate comparison, we plot these coefficients alongside our primary estimates from Figure 2 (Table 3). The results show that, in both reading and math, conditioning on duration drives the negative coefficients for the 80-89 and 90+ degree temperature bins toward zero. In reading, the coefficients for both bins are very close to zero and insignificant. The math estimates are also reduced in magnitude, but remain negative and significant. Although these patterns are consistent with the negative effects of test-day heat partially operating through reductions in the time spent on the test, other dynamics could produce similar results, leading us to exercise caution in interpreting test duration as a mechanism by which temperature may operate.

Placebo Test-Day Temperature Analysis

To provide additional confidence that our results are not spurious and driven by, for example, school or neighborhood characteristics, we conduct a placebo analysis where we replace the test-day temperature treatment specification in equation (1) with a placebo specification based on the maximum temperature one year—365 days—after the student took the MAP test. Put differently, we re-code the series of test-day temperature bin indicators according to the maximum temperature 365 days after the test date. With the measure of future temperature in place, we estimate its "effect" on student MAP performance using the model presented in equation (1). We present the results from estimating this model in Table A8 in the appendix. As expected, the placebo results are broadly insignificant and display no systematic pattern with respect to temperature effects—these results stand in clear contrast to our main results presented above. As a whole, the results in Table A8 instill a degree of confidence in our identifying assumptions and, ultimately, in our substantive conclusions.

Exploiting Within-Season, Across-Subject Variation in Test-Day Temperature

Our primary analyses estimate the effect of temperature separately by MAP subject. In this analysis we combine the two subjects into a single dataset and exploit differences in test-day temperature for a given student in a particular season in a given year. We do this via the model:

$$P_{iust} = \left[\sum_{u=1}^{u=2}\left(\sum_{b=1}^{b=7}\gamma_{ub}M_{iust}\right)\right] + \sum_{s=1}^{s=3}\delta_s D_{iust} + \sum_{s=1}^{s=3}W_{iust}\phi_s + \rho_{ist} + \theta_{iu} + \varepsilon_{iust} \qquad (5)$$

where we specify the outcome measure as the MAP percentile, *P*, for student *i* on the assessment of subject *u* during season *s* in school year *t*. Our treatment specification again consists of a series of dummy variables indicating the ten-degree temperature bin containing the maximum temperature on the day the student took the MAP test. However, because our analytic sample contains student scores in both math and reading, we estimate the coefficient for each bin separately by subject. In addition to this treatment specification, the model also contains a measure of the number of calendar days since the prior MAP administration in the same subject, *D*, the coefficient for which we also allow to vary by season. Finally, the model contains fixed effects for student-by-season-by-year, ρ_{ist}, and student-by-subject, θ_{iu}; ε_{iust} represents the error term. We estimate the model via OLS and cluster standard errors at the school-by-test date level.

We present the results from estimating this model in Table A9 in the appendix. In math, the estimated test-day temperature effects are broadly similar to the primary estimates presented above—test days with max temps in the 80s or 90s reduce student performance by about one-half of a percentile, compared to a day that maxes out in the 70s. The reading results, by contrast, return no evidence of negative effects of test-day heat. If anything, there is some evidence of test-day maximum temperatures in the 80s or 90s producing a slight increase in student performance, compared to a day with a max temp in the 70s. Together, the results provide further indication that test-day heat affects math performance more than it does reading.

Discussion and Conclusion

As our changing climate exposes students to extreme temperatures with increasing frequency, it becomes ever more important to gain a deeper understanding of how such exposure affects students' learning and performance. 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 detailed portrait of the effect of temperature on student test performance.

Consistent with prior work, we show that test-day heat reduces student test performance, even in a district with updated climate control capabilities. However, our study adds important depth and nuance to our understanding of these effects. Specifically, we highlight three main ways that this paper extends prior work on the topic, discussing the implications of each of these extensions for both research and practice. First, our study is uniquely positioned to separately estimate temperature effects on student test performance in the fall, winter, and spring. And our results provide clear evidence of seasonal variation in temperature effects. Indeed, relative to a day that maxes out in the 70s, a day with a maximum temperature in the 80s significantly reduces student performance in the spring but has no effect on fall test performance—student performance in the fall declines only when temperatures reach the 90s.

From a research standpoint, this points to the importance of further inquiry into acclimatization as a potential moderator of heat effects. In doing so, it provides important context for considering our results alongside prior work investigating the effects of temperature on student test performance. For example, Park (2022) finds that a test day where the temperature maxes out in the 90s, relative to the mid-70s, reduces student performance on the New York Regents Exam by more than 10 percent of a standard deviation while our main results (see Table 3) fall below that size. However, our work illuminates several potential explanations for this difference. Notably, the climate control infrastructure (i.e., air conditioning) in TPS was superior to that in New York City, which likely accounts for some of the difference. Beyond that feature of the built environment, given the state's relatively cooler climate, students in New York were unlikely to be well-acclimatized to 90+ degree heat at the time of the exam in June. In the TPS context, by contrast, nearly all of students' test-day exposure to 90+ degree heat occurred during the fall MAP administration, which followed Oklahoma's hot summer months that regularly exposed students to 90+ degree days and may have contributed to relatively higher levels of acclimatization. More generally, our work points to the importance of systematically analyzing the roles of seasonality and student acclimatization levels in moderating temperature effects.

In terms of practice, our evidence of seasonal heterogeneity has clear implications for conclusions that teachers and administrators draw regarding student learning gains over the course of the school year. The smaller fall effects indicate that test-day temperature plays only a modest role in establishing students' performance baseline. By contrast, the greater sensitivity of spring test scores to temperature implies that administering the MAP on a hot spring day may lead educators to infer that students made fewer gains, relative to the inferences that would have been drawn had the assessment been administered on a cooler date. Furthermore, there is reason to think that temperature effects will extend beyond the MAP test, meaning that student performance on state tests used for accountability purposes—typically administered in the spring of each school year—may be meaningfully affected by test-day temperature.

Second, our study is among the first to incorporate student test duration into analyses of temperature effects on test performance. We first show that, in both reading and math, a day with a maximum temperature in the 90s reduces student test duration by almost a tenth of a standard deviation, relative to a day that maxes out in the 70s. Then, we show that the estimated effects of test-day heat are smaller in magnitude when we specify student test duration as a covariate in the model. The estimated effects in math are smaller in magnitude, but remain negative and significant while the reading effects are rendered insignificant. Together, these results advance our academic understanding of the effects of test-day heat by demonstrating that they could partially operate through the behavioral mechanism of students reducing the amount of time they spend on the test. However, it also suggests that, at least in math, the effects operate through additional mechanisms, perhaps physiological or cognitive ones.

Future research would do well to further explore the mechanisms by which temperature effects manifest. A potentially fruitful line of inquiry might examine whether and how temperature differentially affects the cognitive processes that promote math versus reading achievement, which may partially account for the observed heterogeneity in temperature effects across the two subjects. And from a practical standpoint, our results suggest that educators might consider strategies and incentives to encourage student focus, particularly on hot test days.

Third, our estimation of meaningful test-day temperature impacts alongside the general absence of systematic longer-term temperature effects informs discussions around the impact of temperature on student test performance versus learning more broadly. Our results provide clear evidence that temperature affects test performance, but much more limited evidence that it impacts learning, at least in our empirical context. From a research perspective, these results demonstrate the importance of conceptually separating student test performance from student learning *per se*. At the same time, we recognize the challenges of credibly estimating each set of effects. Indeed, our test-day temperature effects (i.e., performance effects) are better identified than our three-month temperature effects (i.e., learning effects). Still, future research should work to create and implement designs that can credibly disentangle short-term performance effects from longer-term impacts on learning.

Along with its scholarly relevance, the distinction between performance effects and learning impacts has clear implications for policy and practice. A reasonable interpretation of our results holds that temperature can systematically distort the measurement of achievement while having little effect on the underlying learning that the achievement measures are intended to capture. Such an interpretation suggests that policymakers should use these scores as inputs into high-stakes decisions only with substantial caution, and punitive decisions (e.g., grade retention) carry even greater equity risks from misattributing environmental impacts to student effort. It also points toward concrete actions school and district personnel could take to mitigate performance effects. For example, the district could work to implement flexible test scheduling protocols that would allow educators to avoid testing students on unusually hot days. This policy could be paired with building operations procedures, like systematically evaluating and re-adjusting classrooms' temperatures prior to test-taking, designed to ensure that students sit for the MAP in a comfortable environment.

To better contextualize our estimated effects, it is useful to benchmark them against the estimated impacts from the literature on a separate, but related, environmental hazard—air pollution. Estimates across that literature—examining the effects of pollution from highways (Heissel et al. 2022), industrial airborne toxins (Persico & Venator 2021), and coal-fired power generation (Duque & Gilraine 2022)—generally range from about 0.02 to 0.06 standard deviations, although a recent analysis of air filter installation finds larger effects of 0.1 to 0.2 standard deviations (Gilraine 2025). In general, though, the magnitude of our estimated effects is comparable to that in the air pollution literature. This suggests that a 90-plus degree MAP administration produced performance declines comparable to those produced by a year of exposure to coal-fired pollution, or from attending a school downwind of a major highway.

All told, this paper meaningfully advances our understanding of how temperature affects student performance and, in doing so, contributes to an ever-growing literature documenting how aspects of both the natural environment and anthropogenic hazards play a substantial role in shaping students' educational outcomes. Along with answering several important questions, however, it is also subject to multiple limitations—we highlight four. First, like other studies on the topic, we cannot observe indoor temperatures in the school when students sit for the MAP, leaving open how the conditions students actually experience during testing affect their performance. Future work would do well to examine the interplay of indoor and outdoor temperatures with respect to student academic performance.

Second, we use the daily maximum temperature as the primary measure in our empirical analysis, which may differ from conditions at the time students sat for the assessment, which is unobservable in our data. Of course, any measurement error resulting from such differences would attenuate our estimates. Relatedly, our three-month measures capture the number of days in each temperature bin over a fixed window but not the timing of those days within it. Because the typical recency of a given temperature differs across seasons, we interpret the season-specific three-month estimates as reflecting cumulative seasonal exposure rather than exposure at a fixed distance from the test, and we see value in future work that leverages richer within-window variation to assess whether the timing of exposure independently shapes learning.

Third, it is possible that TPS administrators use the flexibility that the NWEA MAP accords to avoid administering the assessment on days with extreme temperatures. Informal conversations with district personnel, coupled with MAP administrations spanning a broad range of temperatures, suggest that such a scenario is unlikely, but we cannot rule it out as our data only contain information on administration dates—we do not observe the scheduling process. Finally, although there are significant benefits to analysis of a single, localized empirical context, we have no way of knowing the degree to which our results generalize beyond Tulsa. Still, our study draws upon data from a diverse school district near the middle of the national MAP achievement distribution, and with much temperature variation across seasons and years. These characteristics, along with the benefit of repeated test administrations per student, facilitate our addition of important depth and nuance to a rapidly evolving literature.

---

[i] Fall MAP administrations occurred almost exclusively in September during the 2014-15 through 2016-17 school years. In the 2017-18 through 2019-20 school years, the first administrations occurred in late August and extended through mid-September. Winter administrations were more variable in their timing. They occurred during January and February in 2014-15 and 2016-17, but only February in 2015-16. In 2017-18 the administration occurred earlier, with almost all tests taken in November and December. In 2019-20 the winter MAP administrations occurred during December and January. Across all years, Spring administrations occurred entirely in late-April through mid-May.

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

[iii] 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.

[iv] 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.

[v] We obtain humidity and precipitation controls from the Oklahoma Mesonet, interpolating them using the same approach we applied to the temperature measures. We use the daily average of relative humidity as our humidity measure. Our precipitation measure gauges inches of daily precipitation, which includes frozen precipitation (e.g., snow) upon thawing. We supplement these weather records with daily PM2.5 air pollution readings from the Tulsa EPA monitor, which reflect average daily readings across Tulsa and thus cannot be interpolated to student zip codes.

[vi] Seasonal temperature variation results in some bins not being populated in certain seasons. Table 2 demonstrates that, for the fall administration, there are no days with maximum temperatures in the 20s, 30s, 40s, or 50s and only a very small number of days with highs above 100. Accordingly, for fall we combine days with highs in the 90s or 100s into a single bin that we refer to as 90 or above. For winter, there are no days with high temperatures above 80 and only a handful of days with highs in the 20s. We combine days with highs in the 20s or 30s into a single bin that we refer to as the 39 or below bin. In spring, there are no days with maximum temperatures below 50 or above 100. Although we created each bin for each season, bins unpopulated in a particular season are omitted from the empirical model.

[vii] We transform percentiles into the reported standard deviations by standardizing the MAP percentile measure in our data and re-estimating each regression for which we report effects in standard deviation units with this standardized measure as the outcome—we perform this standardization separately for math and reading. Further, because percentile ranks are bounded, we plotted the distribution of our percentile measure for both math and reading. We also compared the percentiles against the underlying scale scores. These exercises returned no evidence of a skewed percentile distribution.

[viii] Point estimates and standard errors for all populated bins are presented in Table A1 in the appendix.

[ix] Across the five school years with spring MAP administrations, there are only four test days with maximum temperatures exceeding 90 degrees—May 8, 9, 29 and 30 in 2018—and May 8 and 9 have only one observation each, meaning that the coefficient estimates are informed almost entirely by MAP administrations on May 29 and 30 in 2018. Notably, Oklahoma teachers walked off the job in spring of 2018, which caused many schools in the state, including those in TPS, to shut down for two weeks in early April. TPS revised its academic calendar in response to this shutdown, moving the last day of school from May 23 to May 31. Accordingly, the coefficient estimates for spring days in the 90s are informed by student performance on MAP assessments administered just a day or two before the end of a school year that had been extended due to the teacher walkout. We report these results in Table A1 for purposes of transparency, but we caution against drawing any substantive conclusions from them.

[x] Seasonal temperature variation again results in some bins being either unpopulated or sparsely populated in certain seasons—the specific set of unpopulated or sparsely populated bins is indicated by Table 2. Bins unpopulated in a particular season are omitted from the empirical model. Regarding sparsely populated bins, there are relatively few days in both the 10-19 and 20-29 degree bins for the winter and spring MAP administrations—there are no such days for the fall administration. Accordingly, separately for each season, we combine the two bins into a single bin that we label "29 or below." This combining of bins accounts for h spanning nine bins in equation (4) but ten bins in equation (3).

[xi] Specifically, we obtain data from the IRS website detailing adjusted gross income (AGI) per tax return filed in each zip code in each year and merge these records onto the student panel dataset to serve as our income measure.

[xii] Results from the season-specific analyses are available from the authors upon request.

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, W., G. Heutel, and D. 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.

Bednar, D. J., & Reames, T. G. 2020. Recognition of and response to energy poverty in the United States. *Nature Energy*, 5(6): 432-439.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts. 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 A. J. Young. 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.

Chang T. Y., and A. Kajackaite. 2019. Battle for the thermostat: Gender and the effect of temperature on cognitive performance. *PLOS ONE*, 14 (5): e0216362. https://doi.org/10.1371/journal.pone.0216362

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

Cook, N., and A. Heyes. 2020. Brain freeze: outdoor cold and indoor cognitive performance. *Journal of Environmental Economics and Management*, 101: 102318. https://doi.org/10.1016/j.jeem.2020.102318.

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 7/29/2023).

Duque, V., and M. Gilraine. 2022. Coal use, air pollution, and student performance. *Journal of Public Economics*, 213: 104712. https://doi.org/10.1016/j.jpubeco.2022.104712

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, and 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, and 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

Garg, T., M. Jagnani, and V. Taraz. 2020. Temperature and human capital in India. *Journal of the Association of Environmental and Resource Economists*, 7 (6): 1113-1150. DOI: doi.org/10.1086/710066

Gilraine, M. 2025. Air Filters, Pollution, and Student Achievement. *Journal of Human Resources*, 60 (5): 1469-1506. DOI: 10.3368/jhr.0421-11642R2

Goodman, J. (2014). *Flaking out: Student absences and snow days as disruptions of instructional time* (No. w20221). National Bureau of Economic Research.

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://www.jstor.org/stable/10.1086/671766

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

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. https://doi.org/10.1016/j.jeem.2020.102365.

Heaney, A. K., D. Carrión, Katrin Burkart, C. Lesk and D. 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, J. A., C. Persico and D. 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

Hyndman, B., A. Shortridge, and J. Vanos. 2023. The Impact of Indoor and Outdoor Heat Extremes on Schoolchildren. In *The Impact of Extreme Weather on School Education: Protecting School Communities*, edited by B. Hyndman and J. Vanos, 18–37. London: Routledge. https://doi.org/10.4324/9781003103165-2

Hyndman, B., and J. Vanos. 2023. The Impact of Extreme Weather on School Communities. In *The Impact of Extreme Weather on School Education: Protecting School Communities*, edited by B. Hyndman and J. Vanos, 1–17. London: Routledge. https://doi.org/10.4324/9781003103165-1
Johnston, D. W., R. Knott, S. Mendolia, and 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

Kraft, M. A. 2020. Interpreting Effect Sizes of Education Interventions. *Educational Researcher,* 49(4): 241-253. https://doi.org/10.3102/0013189X20912798

Krebs, B. 2024. Temperature and Cognitive Performance: Evidence from Mental Arithmetic Training. *Environmental and Resource Economics*, 87: 2035–2065. https://doi.org/10.1007/s10640-024-00881-y

Lee, W., M. Moon, H. G. Kim, T. H. Lee, and 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

McCormack, K. 2023. Education under extremes: temperature, student absenteeism, and disciplinary infractions. Unpublished Manuscript: https://kristen-mccormack.com/files/mccormack_jmp.pdf (Accessed 7/1/2025).

McPherson, R. A., C. A. Fiebrich, K. C. Crawford, J. R. Kilby, David L. Grimsley, J. E. Martinez, J. B. Basar, B. G. Illston, D. A. Morris, Kevin A. Kloesel, A. D. Melvin, H. Shrivastava, J. M. Wolfinbarger, J. P. Bostic, D. B. Demko, R. L. Elliott, S. J. Stadler, J. D. Carlson, and A. 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

Melo, A. P., and M. Suzuki. 2023. Temperature, effort, and achievement. Unpublished Manuscript: https://www.apmelo.com/files/manuscript_MM2.pdf (Accessed 7/1/2025).

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.

National Weather Service (NWS). 2024. NWS Tulsa Main Climate Page. *National Oceanic and Atmospheric Administration: National Weather Service*. https://www.weather.gov/tsa/ClimateMain (Accessed 10/10/2024).

Obradovich, N., R. Migliorini, S. C. Mednick, and J. H. Fowler. 2017. Nighttime temperature and human sleep loss in a changing climate. *Sci. Adv.*, 3: e1601555. DOI: 10.1126/sciadv.1601555

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, E. 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., S. Racinais, and M. N. Sawka. 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, C. L. and J. Venator. 2021. The Effects of Local Industrial Pollution on Students and Schools. *Journal of Human Resources*. https://doi.org/10.3368/jhr.56.2.0518-9511R2

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

Robinson, A., J. Lehmann, D. Barriopedro, S. Rahmstorf and D. 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

Romitti, Y., I. Sue Wing, K. R. Spangler, and G. A. Wellenius. 2022. Inequality in the availability of residential air conditioning across 115 US metropolitan areas. *PNAS nexus*, 1 (4): pgac210. https://doi.org/10.1093/pnasnexus/pgac210

Soligo, M. J., P. A. Irwin, C. J. Williams, and G. 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.

Srivastava, P., T. A. Trinh, and X. Zhang. 2025. Weather effects on academic performance: an analysis using administrative data. *Popul Environ* **47**: 25. https://doi.org/10.1007/s11111-025-00495-6

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

United States Energy Information Administration. 2023. "Highlights for air conditioning in U.S. homes by state, 2020." *USEIA*. https://www.eia.gov/consumption/residential/data/2020/state/pdf/State%20Air%20Conditioning.pdf (Accessed 7/6/2025).

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

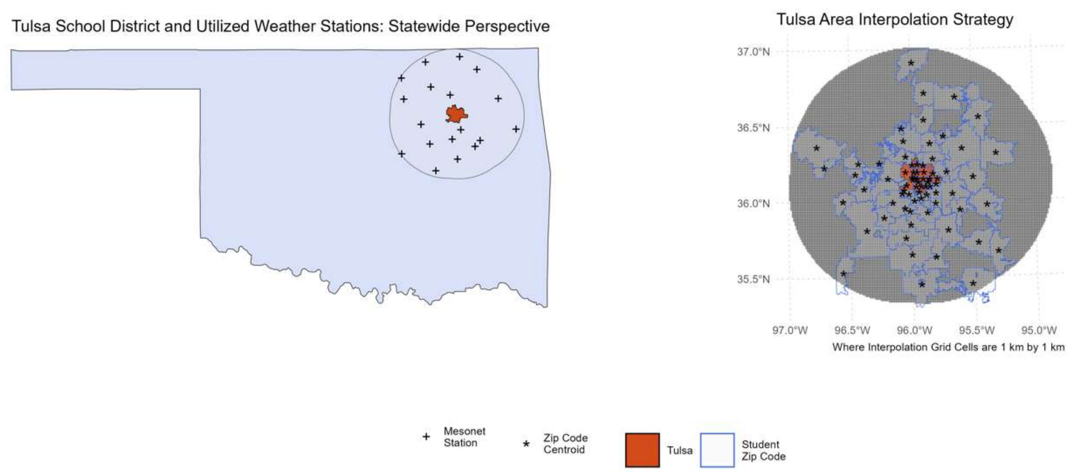
Weber, D., and E. Englund. 1992. Evaluation and comparison of spatial interpolators. *Math Geol,* **24:** 381–391. https://doi-org.ezproxy.lib.ou.edu/10.1007/BF00891270

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

Yi, F., T. Zhou, L. Yu, B. McCarl, Y. Wang, F. Jiang, and Y. Wang. 2021. Outdoor heat stress and cognition: Effects on those over 40 years old in China. Weather and Climate Extremes, 32: 100308. https://doi.org/10.1016/j.wace.2021.100308

Tables and Figures

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



NOTE: The left-hand map situates Tulsa Public School District (TPS) within Oklahoma, a southwestern U.S. state. Plus signs (+) denote Oklahoma Mesonet weather monitoring station locations inside a 50-mile radius around TPS, as bounded by the rendered circle. The right-hand map illustrates how students' zip codes were imposed on an interpolation grid to facilitate linking our 1km-by-1km weather raster to Tulsa area student locations. We exaggerate the zip code centroids (center points) as enlarged stars (*) for added visual clarity.

Table 1. Descriptive statistics for students

Characteristic	Mean
Female	0.492
Male	0.508
Asian	0.014
Black	0.239
Hispanic	0.348
American Indian/Alaska Native	0.051
Multiple Races	0.100
Pacific Islander	0.006
White	0.242
Disability	0.172
Grade K-5	0.742
Grade 6-8	0.158
Grade 9-12	0.101
Fall administration	0.361
Spring administration	0.281
Winter administration	0.357
N Observations	339,034
N Unique students	50,141
N Unique schools	88
N School years	6

NOTE: The spring MAP administration contains fewer testing days than other seasons due to the start of the COVID-19 pandemic.

Table 2. Descriptive statistics for temperature and test-based outcome measures, by season

	Full Sample		Fall		Winter		Spring	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
	Temperature Measures-Test Day							
Maximum Temp- Test day	72.5	16.1	86.3	5.5	54.7	11.1	77.3	7.8
Max Temp 20-29- Test day	0.000	NA	.	.	0.000	NA	.	.
Max Temp 30-39- Test day	0.051	NA	.	.	0.141	NA	.	.
Max Temp 40-49- Test day	0.059	NA	.	.	0.166	NA	.	.
Max Temp 50-59- Test day	0.141	NA	.	.	0.385	NA	0.010	NA
Max Temp 60-69- Test day	0.121	NA	0.006	NA	0.199	NA	0.168	NA
Max Temp 70-79- Test day	0.198	NA	0.139	NA	0.108	NA	0.388	NA
Max Temp 80-89- Test day	0.307	NA	0.513	NA	.	.	0.432	NA
Max Temp 90-99- Test day	0.124	NA	0.342	NA	.	.	0.001	NA
Max Temp 100-109- Test day	0.000	NA	0.000	NA	.	.	.	.
	Temperature Measures- Three Month							
Avg. Temp- Three months	61.7	13.6	78.7	1.0	51.4	7.4	53.1	2.6
Max Temp 10-19-Three months	0.1	0.4	.	.	0.2	0.6	0.1	0.2
Max Temp 20-29-Three months	0.6	1.0	.	.	0.4	0.7	1.5	1.3
Max Temp 30-39-Three months	3.6	5.1	.	.	6.7	6.0	4.4	4.0
Max Temp 40-49-Three months	6.2	6.6	.	.	12.5	6.3	6.2	2.5
Max Temp 50-59-Three months	12.3	10.7	0.0	0.0	19.6	7.8	18.7	5.2
Max Temp 60-69-Three months	15.1	11.4	0.3	0.9	22.0	3.5	25.2	2.3
Max Temp 70-79-Three months	14.4	8.7	5.8	3.1	15.3	5.8	24.4	4.1
Max Temp 80-89-Three months	21.0	16.1	39.5	8.9	10.6	9.4	10.4	4.0
Max Temp 90-99-Three months	17.5	21.4	44.8	9.2	3.7	3.8	0.1	0.4
Max Temp 100-109-Three months	0.2	0.5	0.5	0.8	.	.	.	.
N Observations	337,764		122,049		120,752		94,963	
N Unique students	49,981		46,825		46,076		41,224	
N Unique test dates	336		113		131		92	
N Unique schools	88		84		87		88	
N School years	6		6		6		6	
	Test-based Outcome Measures							
Test Percentile	40.2	28.9	39.8	28.3	39.7	28.7	41.4	29.8
Test Duration	44.7	25.1	42.4	22.5	46.0	26.5	46.0	26.3
N Observations	339,030		122,510		121,203		95,317	
N Unique students	50,141		46,972		46,220		41,356	
N Unique test dates	336		113		131		92	
N Unique schools	88		82		87		88	
N School years	6		6		6		6	

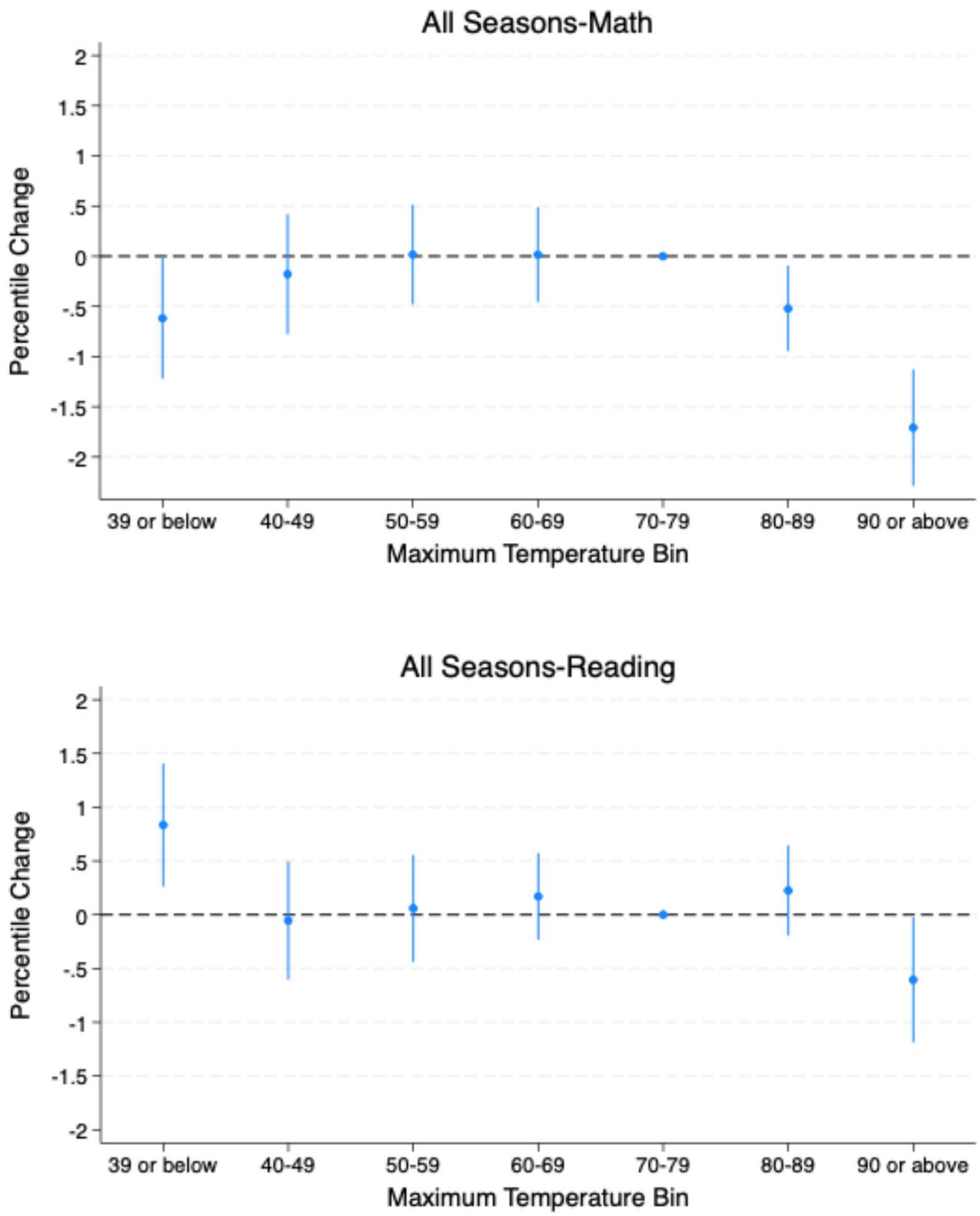
NOTE: Due to seasonal variation in temperatures, some 10-degree bins are not populated during seasons lacking days with relevant temperature maxima. All temperatures are provided in degrees Fahrenheit consistent with the main text. The mean values provided for maximum test-day temperature bins reflect the proportion of all test days containing a maximum daily temperature within the specified bin. For three month temperature bins, the mean values are the average number of days falling into that temperature bin over the three months before MAP tests. Test durations in the bottom portion of the table are provided in minutes.

Table 3. Effect of Test-Day Temperature on Student MAP Percentile, by Temperature Bin and Subject

	All Seasons	
Temperature Bin	**Math**	**Reading**
Max Temp 39 or Below	-0.619**	-0.835***
	(0.299)	(0.292)
Max Temp 40-49	-0.179	-0.055
	(0.305)	(0.281)
Max Temp 50-59	0.016	0.059
	(0.253)	(0.254)
Max Temp 60-69	0.016	0.170
	(0.241)	(0.206)
Max Temp 70-79	OMITTED	OMITTED
Max Temp 80-89	-0.520**	0.226
	(0.217)	(0.214)
Max Temp 90 or Above	-1.708***	-0.605**
	(0.296)	(0.297)
N	326,565	329,616
N Student-School	62,666	63,390
N School-Test Date	10,961	11,418

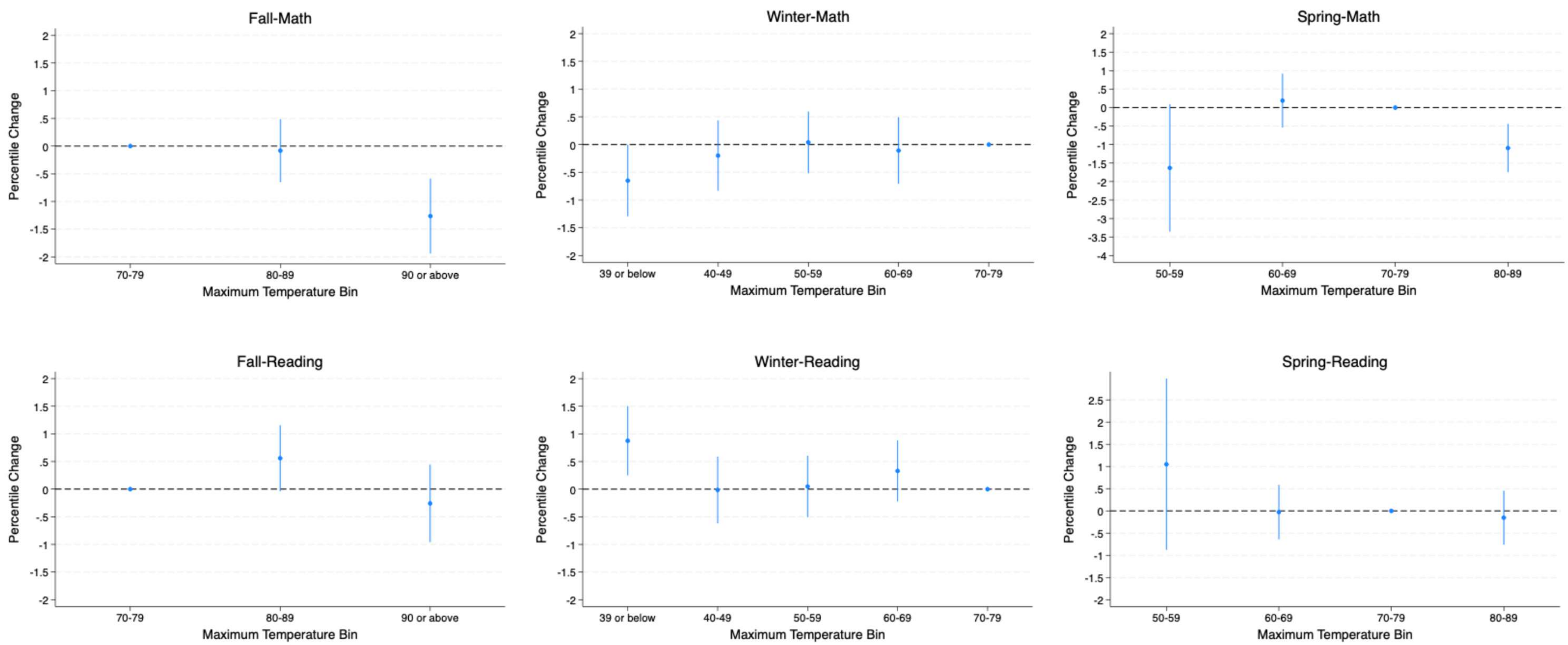
NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard errors clustered by school-by-test date are shown in parentheses below the coefficient estimates. Coefficients in each column come from a single regression estimated via OLS predicting student MAP percentile. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring test-day pollution, humidity, and precipitation—the coefficients on these measures are allowed to vary by season—as well as fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

Figure 2. Effect of Test-Day Temperature on Student MAP Percentile, by Temperature Bin and Subject



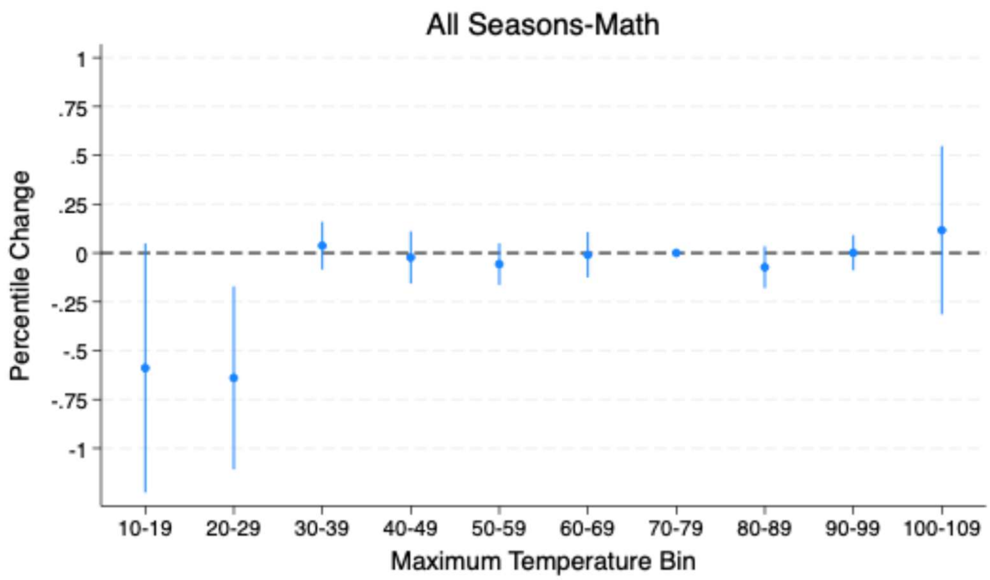
NOTE: Results in each panel of the figure come from a single regression estimated via OLS predicting student MAP percentile. The figure presents coefficient estimates and 95% confidence intervals for a series of variables indicating that the test-day temperature fell within the 10-degree temperature bin specified along the x-axis. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring test-day pollution, humidity, and precipitation—the coefficients on these measures are allowed to vary by season—as well as fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

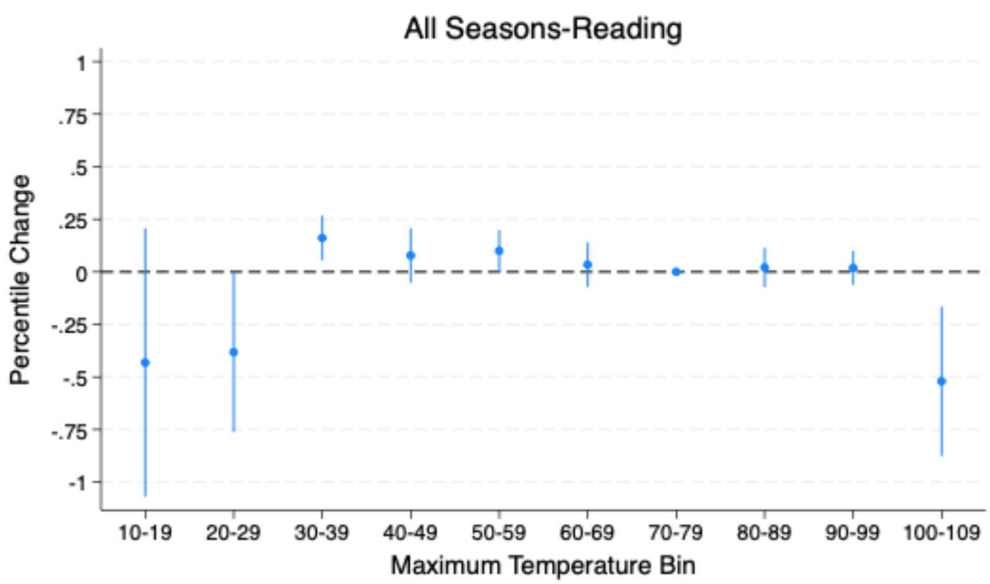
Figure 3. Effect of Test-Day Temperature on Student MAP Percentile, by Temperature Bin, Season, and Subject



NOTE: Results in each row of the figure come from a single regression estimated via OLS predicting student MAP percentile. The top panel displays math results while the bottom panel displays reading results. Each plot presents coefficient estimates and 95% confidence intervals for a series of variables indicating that the test-day temperature fell within the 10-degree temperature bin specified along the x-axis. In addition to this treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring test-day pollution, humidity, and precipitation—the coefficients on these measures are allowed to vary by season—as well as fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

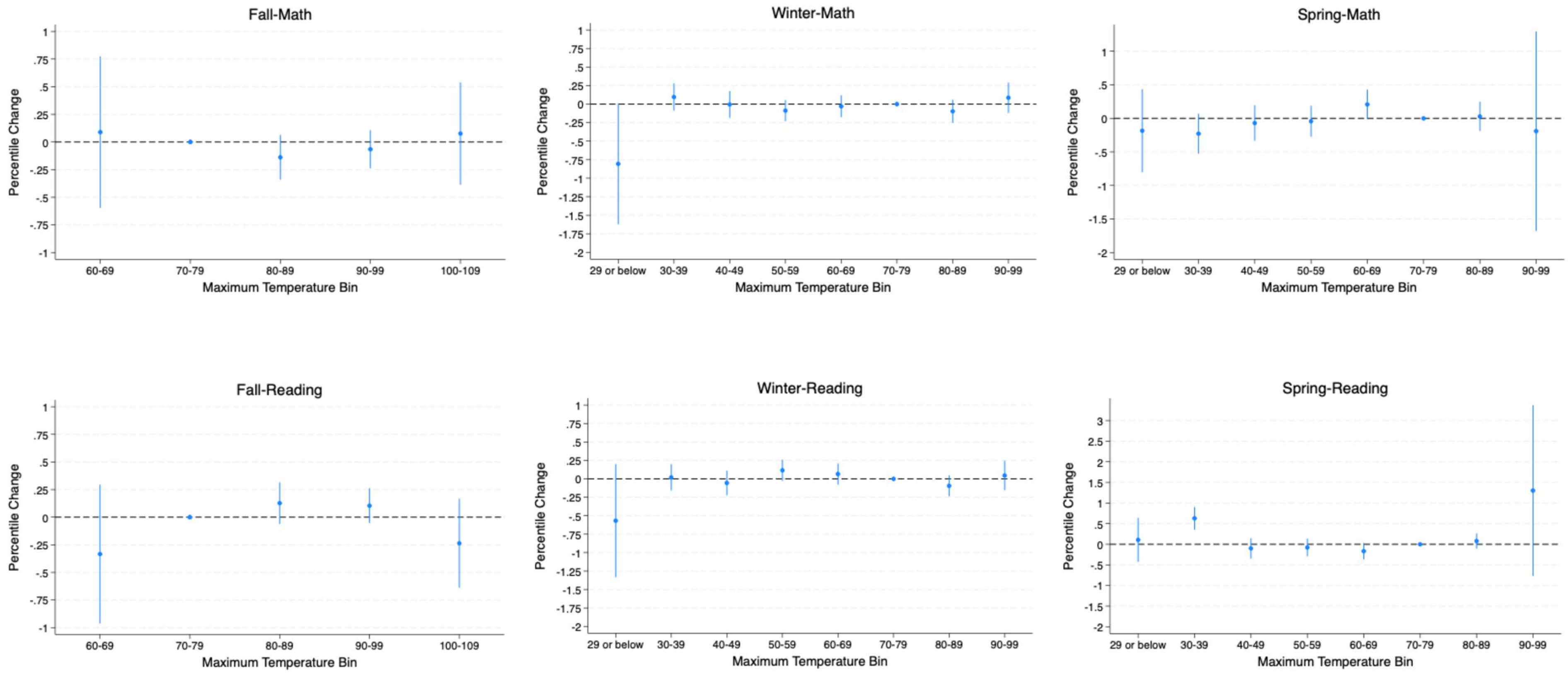
Figure 4. Effect of Three-month Temperature on Student MAP Percentile, by Temperature Bin and Subject





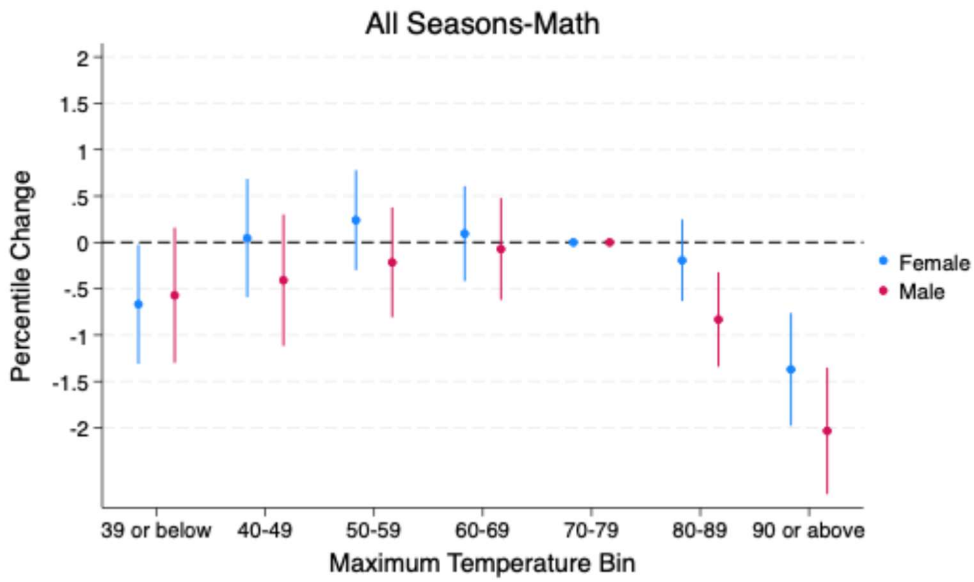
NOTE: Results in each panel of the figure come from a single regression estimated via OLS predicting student MAP percentile. The treatment specification is a series of variables counting the number of days over the prior three months with a high temperature in the ten-degree temperature bin (as specified along the x-axis). In addition to the treatment specification, all regressions contain covariates measuring test-day temperature, as well as the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring average pollution and humidity levels over the prior three months, as well as total precipitation over this period—the coefficients on these measures are allowed to vary by season. Finally, the regressions contain fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

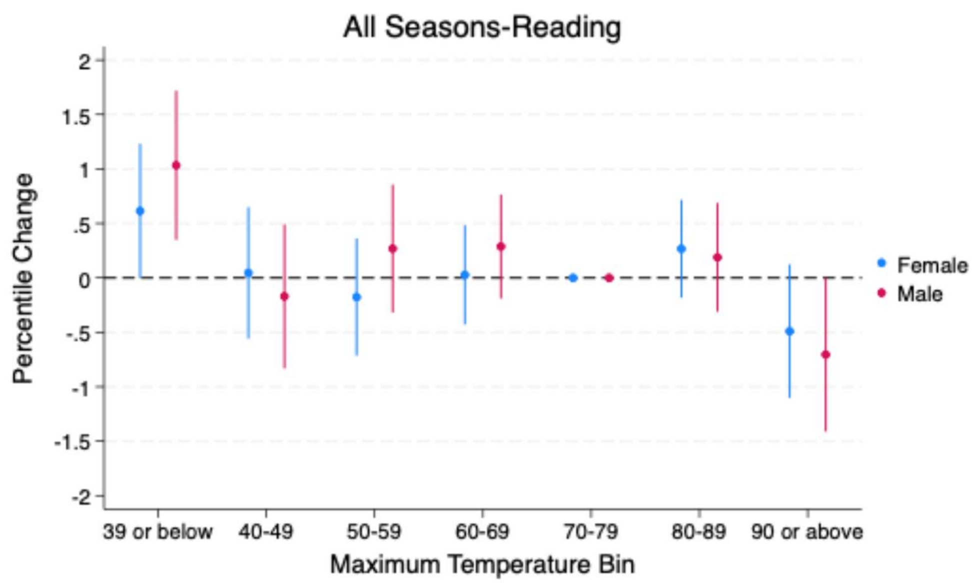
Figure 5. Effect of Three-month Temperature on Student MAP Percentile, by Temperature Bin, Season, and Subject



NOTE: Results in each panel of the figure come from a single regression estimated via OLS predicting student MAP percentile. The top panel displays math results while the bottom panel displays reading results. Each plot presents coefficient estimates and 95% confidence intervals for a series of variables counting the number of days over the prior three months with a high temperature in the ten-degree temperature bin (as specified along the x-axis). In addition to this treatment specification, all regressions contain covariates measuring test-day temperature, as well as the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring average pollution and humidity levels over the prior three months, as well as total precipitation over this period—the coefficients on these measures are allowed to vary by season. Finally, the regressions contain fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

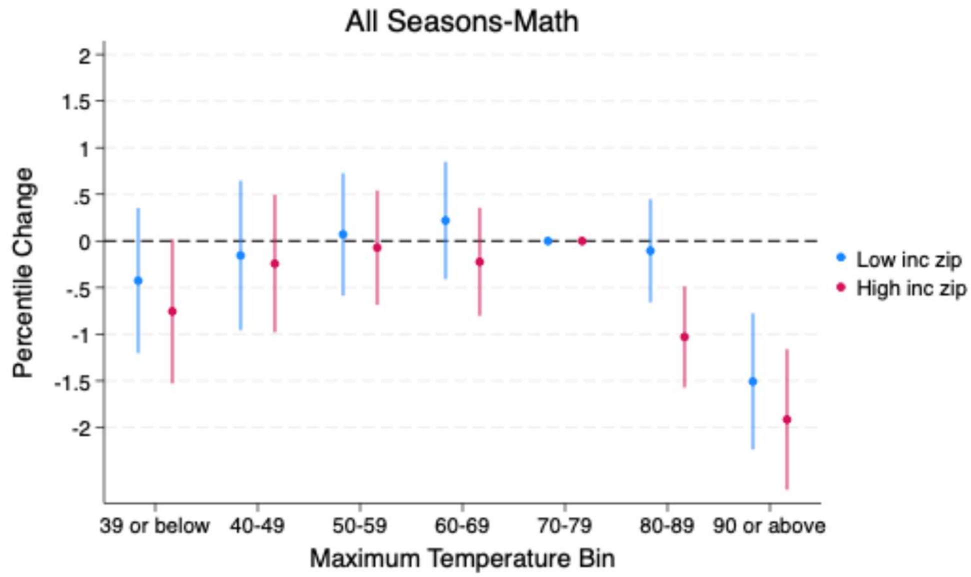
Figure 6. Effect of Test-Day Temperature on Student MAP Percentile, by Temperature Bin, Sex, and Subject

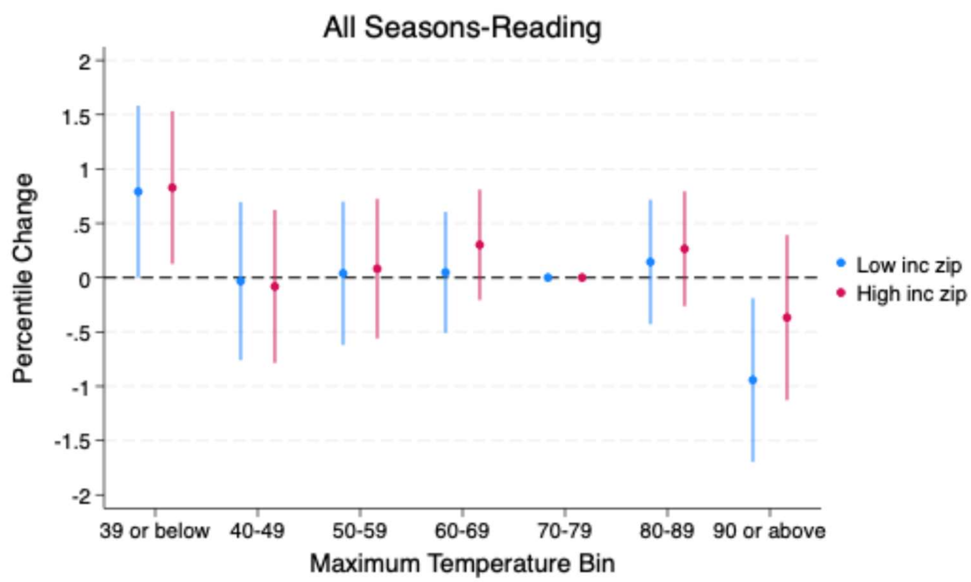




NOTE: Separately by student sex (as indicated in administrative records), the results in each panel of the figure come from a single regression estimated via OLS predicting student MAP percentile. The figure presents coefficient estimates and 95% confidence intervals for a series of variables indicating that the test-day temperature fell within the 10-degree temperature bin specified along the x-axis. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring test-day pollution, humidity, and precipitation—the coefficients on these measures are allowed to vary by season—as well as fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

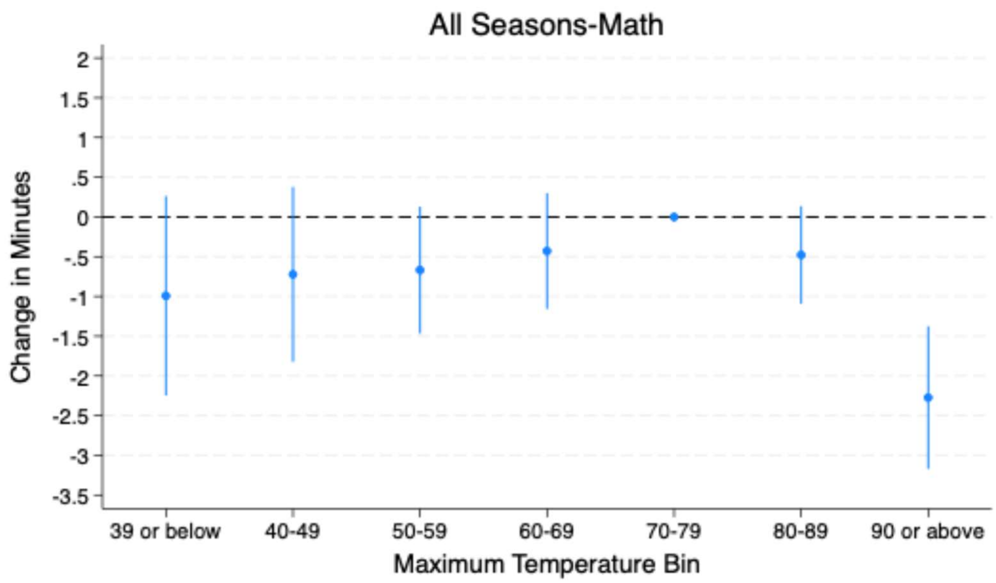
Figure 7. Effect of Test-Day Temperature on Student MAP Percentile, by Temperature Bin, Zip Code Income Level, and Subject

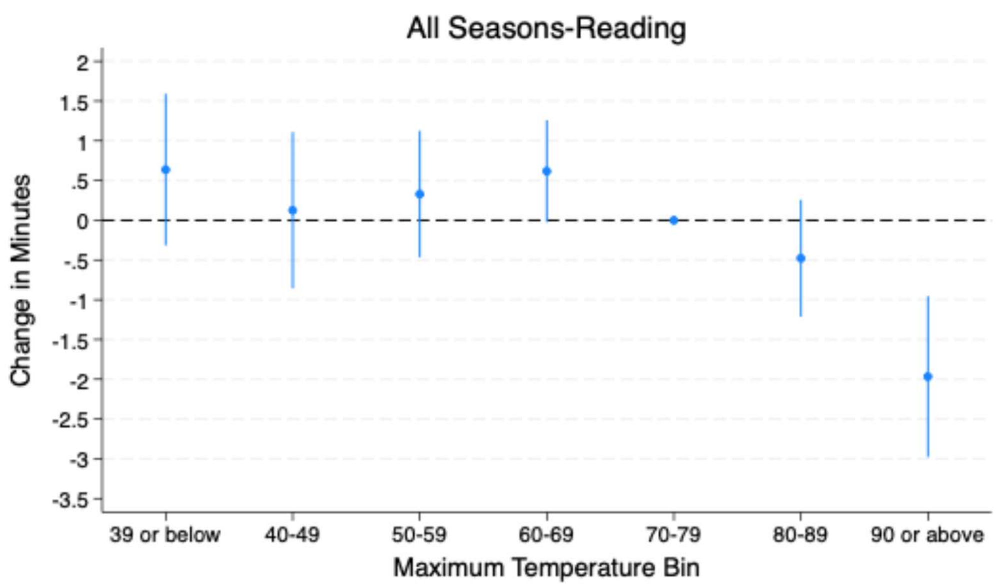




NOTE: Separately by student zip code income level, the results in each figure panel come from a single regression estimated via OLS predicting student MAP percentile. For this analysis, we split the student panel into two groups, where low-income and high-income were defined by whether a student's zip code was below (above) the panel's median Adjusted Gross Income from IRS tax filings. The figure presents coefficient estimates and 95% confidence intervals for a series of variables indicating that the test-day temperature fell within the 10-degree temperature bin specified on the x-axis. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring test-day pollution, humidity, and precipitation—the coefficients on these measures are allowed to vary by season—as well as fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

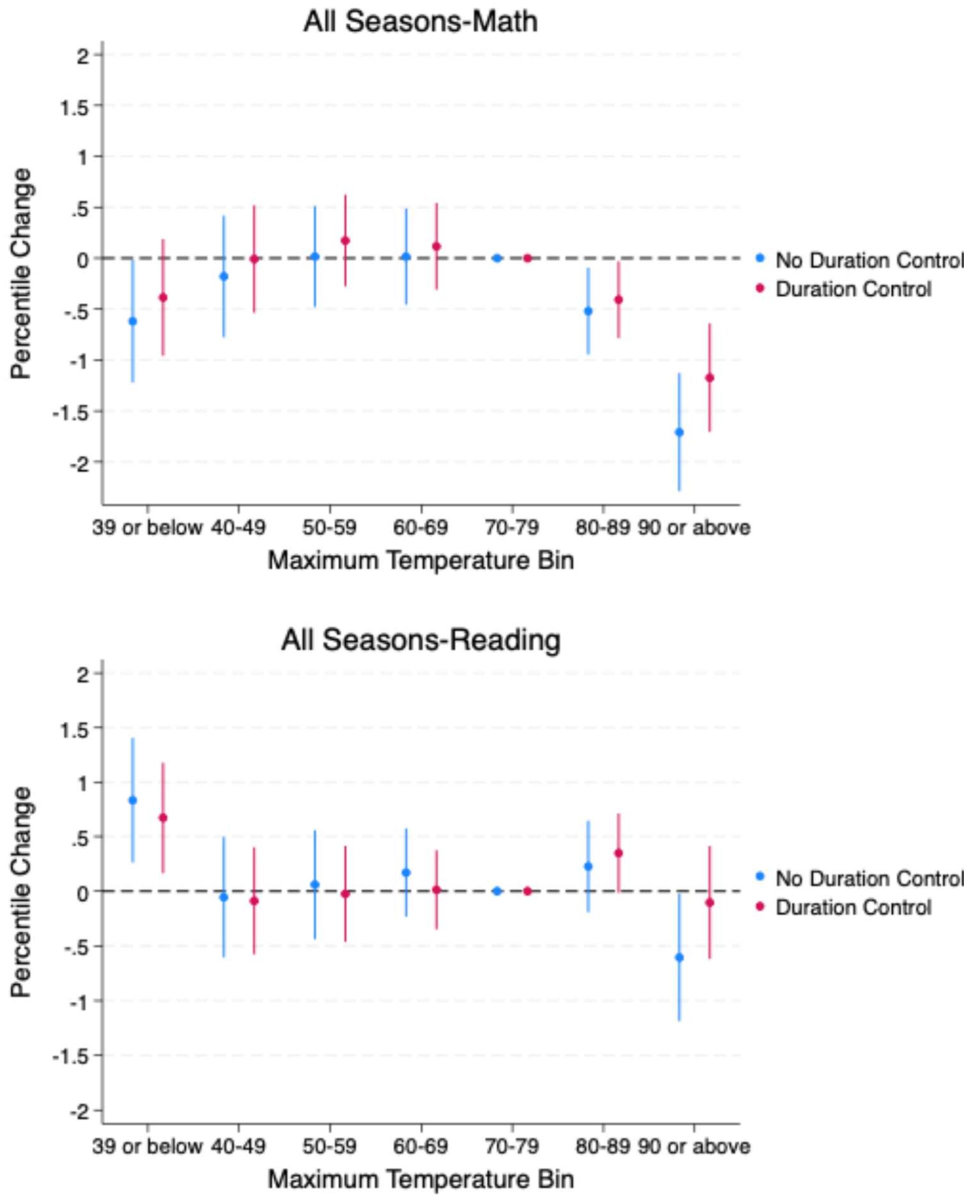
Figure 8. Effect of Test-Day Temperature on MAP Test Duration, by Temperature Bin and Subject





NOTE: Results in each panel of the figure come from a single regression estimated via OLS predicting the number of minutes a student spent taking the MAP assessment. The figure presents coefficient estimates and 95% confidence intervals for a series of variables indicating that the test-day temperature fell within the 10-degree temperature bin specified along the x-axis. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring test-day pollution, humidity, and precipitation—the coefficients on these measures are allowed to vary by season—as well as fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

Figure 9. Effect of Test-Day Temperature on Student MAP Percentile, by Temperature Bin, Control for MAP Duration, and Subject



NOTE: Separately for models with and without a covariate measuring the amount of time a student spent taking the MAP assessment, the results in each panel of the figure come from a single regression estimated via OLS predicting student MAP percentile. The figure presents coefficient estimates and 95% confidence intervals for a series of variables indicating that the test-day temperature fell within the 10-degree temperature bin specified along the x-axis. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain covariates measuring test-day pollution, humidity, and precipitation—the coefficients on these measures are allowed to vary by season—as well as fixed effects for student-by-school, season, and grade-by-year. All temperatures are provided in degrees Fahrenheit consistent with the main text.

Table A1. Effect of Test-Day Temperature on Student MAP Percentile, by Temperature Bin, Season, and Subject

	Math			Reading		
Temperature Bin	**Fall**	**Winter**	**Spring**	**Fall**	**Winter**	**Spring**
Max Temp 39 or Below	.	-0.650*	.		0.876***	
		(0.329)			(0.320)	
Max Temp 40-49	.	-0.199	.		-0.014	
		(0.324)			(0.308)	
Max Temp 50-59	.	0.040	-1.631*		0.049	1.050
		(0.284)	(0.879)		(0.283)	(0.984)
		p test of equality: 0.073			*p* test of equality: 0.329	
Max Temp 60-69	-0.838	-0.107	0.189	-0.062	0.331	-0.026
	(1.476)	(0.306)	(0.371)	(0.941)	(0.282)	(0.313)
		p test of equality: 0.704			*p* test of equality: 0.680	
Max Temp 70-79	OMITTED	OMITTED	OMITTED	OMITTED	OMITTED	OMITTED
Max Temp 80-89	-0.082	.	-1.094***	0.559*	.	-0.153
	(0.290)		(0.334)	(0.305)		(0.310)
		p test of equality: 0.023			*p* test of equality: 0.103	
Max Temp 90 or Above	-1.264**	.	-8.816***	-0.259	.	-6.334***
	(0.345)		(2.669)	(0.358)		(2.720)
		p test of equality: 0.005			*p* test of equality: 0.027	
N		326,565			329,616	
N Student-School		62,666			63,390	
N School-Test Date		10,961			11,418	

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each panel from a single regression estimated via OLS predicting student MAP percentile. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year. In reading, a small number of days in the fall with highs in the 50s were combined with the 60-69 degree bin. In winter, a small number of winter days with high in the 80s were combined with the 70-79 degree bin.

Table A2. Effect of Three-Month Temperature on Student MAP Percentile, by Temperature Bin and Subject

	All Seasons	
	Math	**Reading**
Max Temp 10-19	-0.589*	-0.432
	(0.325)	(0.325)
Max Temp 20-29	-0.640***	-0.382**
	(0.239)	(0.194)
Max Temp 30-39	0.037	0.161***
	(0.062)	(0.054)
Max Temp 40-49	-0.023	0.078
	(0.068)	(0.066)
Max Temp 50-59	-0.057	0.100
	(0.054)	(0.050)
Max Temp 60-69	-0.010	0.034
	(0.059)	(0.054)
Max Temp 70-79	OMITTED	OMITTED
Max Temp 80-89	-0.073	0.021
	(0.055)	(0.048)
Max Temp 90-99	0.001	0.019
	(0.046)	(0.041)
Max Temp 100 or above	0.117	-0.520***
	(0.220)	(0.181)
N	326,565	329,616
N Student-School	62,666	63,390
N School-Test Date	10,961	11,418

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each column from a single regression estimated via OLS predicting student MAP percentile. Treatment specification is a series of variables counting the number of days over the prior three months with a high temperature in the ten-degree temperature bin. In addition to the treatment specification, all regressions contain covariates measuring test-day temperature, the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year.

Table A3. Effect of Three-Month Temperature on Student MAP Percentile, by Temperature Bin, Season, and Subject

	Math			Reading		
	Fall	**Winter**	**Spring**	**Fall**	**Winter**	**Spring**
Max Temp 29 or below	.	-0.806*	-0.185	.	-0.567	0.107
		(0.416)	(0.315)		(0.391)	(0.273)
	p test of equality: 0.245			*p* test of equality: 0.160		
Max Temp 30-39	.	0.095	-0.228	.	0.020	0.629***
		(0.094)	(0.151)		(0.091)	(0.141)
	p test of equality: 0.074			*p* test of equality: 0.001		
Max Temp 40-49	.	-0.005	-0.069	.	-0.055	-0.099
		(0.092)	(0.135)		(0.085)	(0.126)
	p test of equality: 0.696			*p* test of equality: 0.772		
Max Temp 50-59	.	-0.088	-0.043	.	0.116	-0.078
		(0.073)	(0.118)		(0.073)	(0.110)
	p test of equality: 0.744			*p* test of equality: 0.139		
Max Temp 60-69	0.089	-0.030	0.209*	-0.333	0.066	-0.168*
	(0.350)	(0.075)	(0.113)	(0.321)	(0.073)	(0.102)
	p test of equality: 0.214			*p* test of equality: 0.117		
Max Temp 70-79	OMITTED	OMITTED	OMITTED	OMITTED	OMITTED	OMITTED
Max Temp 80-89	-0.138	-0.097	0.029	0.127	-0.094	0.078
	(0.103)	(0.079)	(0.111)	(0.096)	(0.072)	(0.095)
	p test of equality: 0.517			*p* test of equality: 0.131		
Max Temp 90-99	-0.066	0.086	-0.190	0.104	0.046	1.302
	(0.088)	(0.105)	(0.758)	(0.080)	(0.102)	(1.057)
	p test of equality: 0.552			*p* test of equality: 0.479		
Max Temp 100-109	0.076	.	.	-0.236	.	.
	(0.236)			(0.206)		
N		326,565			329,616	
N Student-School		62,666			63,390	
N School-Test Date		10,961			11,418	

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each panel from a single regression estimated via OLS predicting student MAP percentile. Treatment specification is a series of variables counting the number of days over the prior three months with a high temperature in the ten-degree temperature bin. In addition to the treatment specification, all regressions contain covariates measuring test-day temperature, the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year.

Table A4. Effect of Test-Day Temperature on Student MAP Percentile, by Sex, Temperature Bin, and Subject

	Math		Reading	
	Female	**Male**	**Female**	**Male**
Max Temp 39 or Below	-0.667**	-0.571	0.615**	1.034***
	(0.328)	(0.372)	(0.314)	(0.349)
Max Temp 40-49	0.046	-0.407	0.046	-0.17
	(0.326)	(0.362)	(0.308)	(0.338)
Max Temp 50-59	0.240	-0.216	-0.177	0.269
	(0.275)	(0.302)	(0.275)	(0.300)
Max Temp 60-69	0.095	-0.072	0.027	0.288
	(0.261)	(0.280)	(0.232)	(0.243)
Max Temp 70-79	OMITTED	OMITTED	OMITTED	OMITTED
Max Temp 80-89	-0.193	-0.831***	0.267	0.189
	(0.225)	(0.260)	(0.229)	(0.255)
Max Temp 90 or Above	-1.369***	-2.033***	-0.489	-0.705*
	(0.310)	(0.348)	(0.313)	(0.360)
N	160,879	165,681	162,133	167,479
N Student-School	30,700	31,991	30,991	32,425
N School-Test Date	9,810	10,134	10,014	10,451

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each column from a single regression estimated via OLS predicting student MAP percentile. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year.

Table A5. Effect of Test-Day Temperature on Student MAP Percentile, by Zip Code Income Level, Temperature Bin, and Subject

	Math		Reading	
	Lower Income	**Higher Income**	**Lower Income**	**Higher Income**
Max Temp 39 or Below	-0.425	-0.755*	0.792*	0.829**
	(0.397)	(0.394)	(0.404)	(0.359)
Max Temp 40-49	-0.156	-0.243	-0.032	-0.081
	(0.409)	(0.376)	(0.371)	(0.359)
Max Temp 50-59	0.070	-0.072	0.039	0.082
	(0.335)	(0.313)	(0.337)	(0.328)
Max Temp 60-69	0.219	-0.223	0.048	0.302
	(0.321)	(0.296)	(0.284)	(0.259)
Max Temp 70-79	OMITTED	OMITTED	OMITTED	OMITTED
Max Temp 80-89	-0.105	-1.030***	0.144	0.265
	(0.282)	(0.277)	(0.293)	(0.270)
Max Temp 90 or Above	-1.507***	-1.915***	-0.942**	-0.368
	(0.373)	(0.384)	(0.384)	(0.388)
N	164,013	161,111	164,572	163,588
N Student-School	32,611	32,163	32,506	32,624
N School-Test Date	8,906	8,531	9,057	8,800

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each column from a single regression estimated via OLS predicting student MAP percentile. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year.

Table A6. Effect of Test-Day Temperature on Student MAP Duration, by Temperature Bin and Subject

	All Seasons	
	Math	**Reading**
Max Temp 39 or Below	-0.992	0.637
	(0.640)	(0.487)
Max Temp 40-49	-0.723	0.127
	(0.562)	(0.501)
Max Temp 50-59	-0.668	0.330
	(0.407)	(0.406)
Max Temp 60-69	-0.429	0.618*
	(0.372)	(0.327)
Max Temp 70-79	OMITTED	OMITTED
Max Temp 80-89	-0.478	-0.478
	(0.313)	(0.375)
Max Temp 90 of Above	-2.273***	-1.966***
	(0.458)	(0.516)
N	326,570	329,666
N Student-School	62,667	63,411
N School-Test Date	10,963	11,421

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each column from a single regression estimated via OLS predicting the number of minutes a student spent on the MAP. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year.

Table A7. Effect of Test-Day Temperature on Student MAP Percentile, by Duration Control, Temperature Bin, and Subject

	Math		Reading	
	No Duration Control	**Duration Control**	**No Duration Control**	**Duration Control**
Max Temp 39 or Below	-0.619**	-0.385	-0.835***	0.671***
	(0.299)	(0.292)	(0.292)	(0.259)
Max Temp 40-49	-0.179	-0.008	-0.055	-0.088
	(0.305)	(0.269)	(0.281)	(0.249)
Max Temp 50-59	0.016	0.174	0.059	-0.025
	(0.253)	(0.229)	(0.254)	(0.224)
Max Temp 60-69	0.016	0.117	0.170	0.013
	(0.241)	(0.217)	(0.206)	(0.184)
Max Temp 70-79	OMITTED	OMITTED	OMITTED	OMITTED
Max Temp 80-89	-0..520**	-0.407**	0.226	0.348*
	(0.217)	(0.192)	(0.214)	(0.185)
Max Temp 90 or Above	-1.708***	-1.173***	-0.605**	-0.103
	(0.296)	(0.272)	(0.297)	(0.263)
N	326,565	326,565	329,616	329,616
N Student-School	62,666	62,666	63,390	63,390
N School-Test Date	10,961	10,961	11,418	11,418

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each column from a single regression estimated via OLS predicting student MAP percentile. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year.

Table A8. Effect of Temperature 365 Days After MAP Test Date on Student MAP Percentile, by Temperature Bin and Subject

	All Seasons	
	Math	**Reading**
Max Temp 39 or Below	-0.595*	-0.253
	(0.332)	(0.302)
Max Temp 40-49	-0.239	0.182
	(0.284)	(0.251)
Max Temp 50-59	-0.309	0.444*
	(0.383)	(0.232)
Max Temp 60-69	-0.671***	-0.049
	(0.220)	(0.193)
Max Temp 70-79	OMITTED	OMITTED
Max Temp 80-89	-0.017	-0.397**
	(0.209)	(0.192)
Max Temp 90 or Above	-0.147	-0.278
	(0.259)	(0.242)
N	327,791	330,838
N Student-School	62,893	63,615
N School-Test Date	10,973	11,427

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each column from a single regression estimated via OLS predicting student MAP percentile. In addition to the placebo treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-school, season, and grade-by-year.

Table A9. Effect of Test-Day Temperature on Student MAP Percentile Using Within-Season, Cross-Subject Variation in Test-Day Temperature, by Temperature Bin and Subject

	All Seasons	
	Math	**Reading**
Max Temp 39 or Below	-0.331	0.348
	(0.291)	(0.275)
Max Temp 40-49	-0.173	-0.322
	(0.253)	(0.255)
Max Temp 50-59	0.441*	-0.223
	(0.225)	(0.222)
Max Temp 60-69	0.214	-0.450**
	(0.196)	(0.203)
Max Temp 70-79	OMITTED	OMITTED
Max Temp 80-89	-0.374***	0.414**
	(0.171)	(0.180)
Max Temp 90 or Above	-0.595**	0.466*
	(0.268)	(0.279)
N	646,894	
N Student-School	323,364	
N School-Test Date	11,609	

NOTE: *p<0.10, **p<0.05, ***p<0.01. Robust standard error clustered by school-by-test date in parentheses below coefficient estimate. Coefficients in each column from a single regression estimated via OLS predicting student MAP percentile. In addition to the treatment specification, all regressions contain covariates measuring the number of the days between the test administration and the beginning of the school year (for fall administration), winter break (winter administration), or the end of the school year (spring administration). Regressions also contain fixed effects for student-by-season-by-year, and student-by-subject.