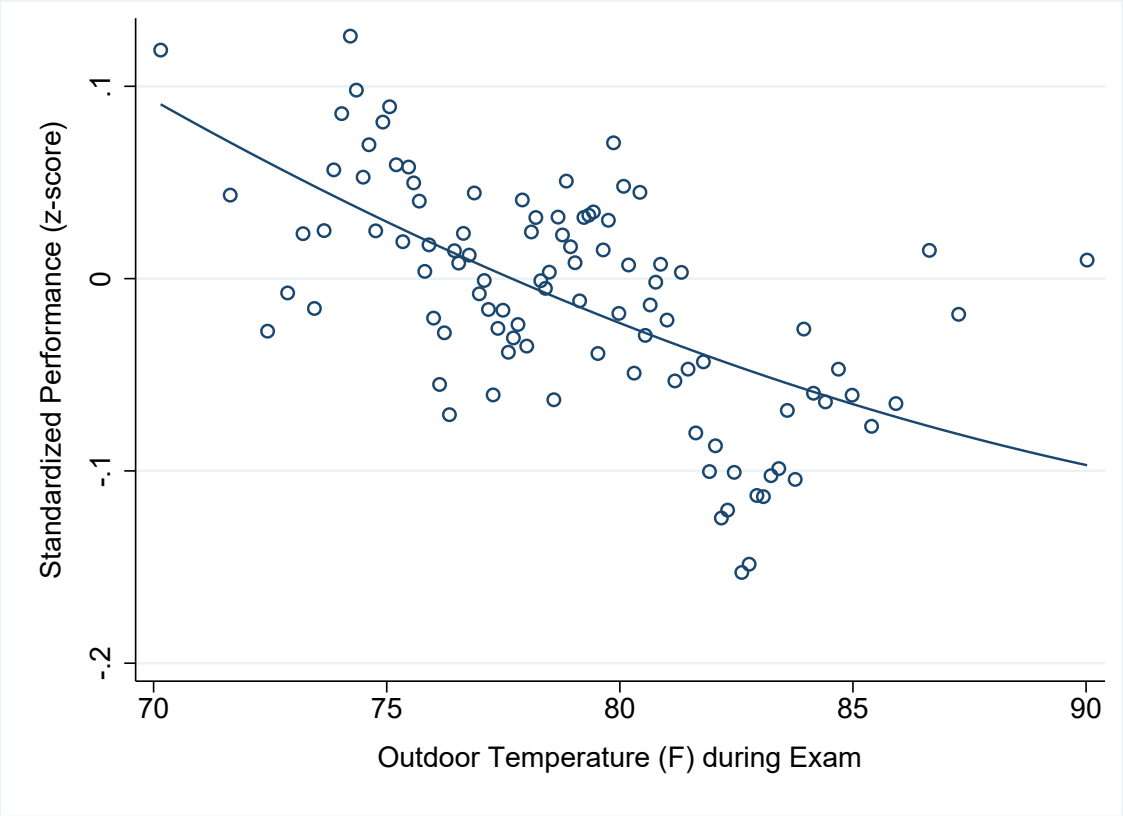
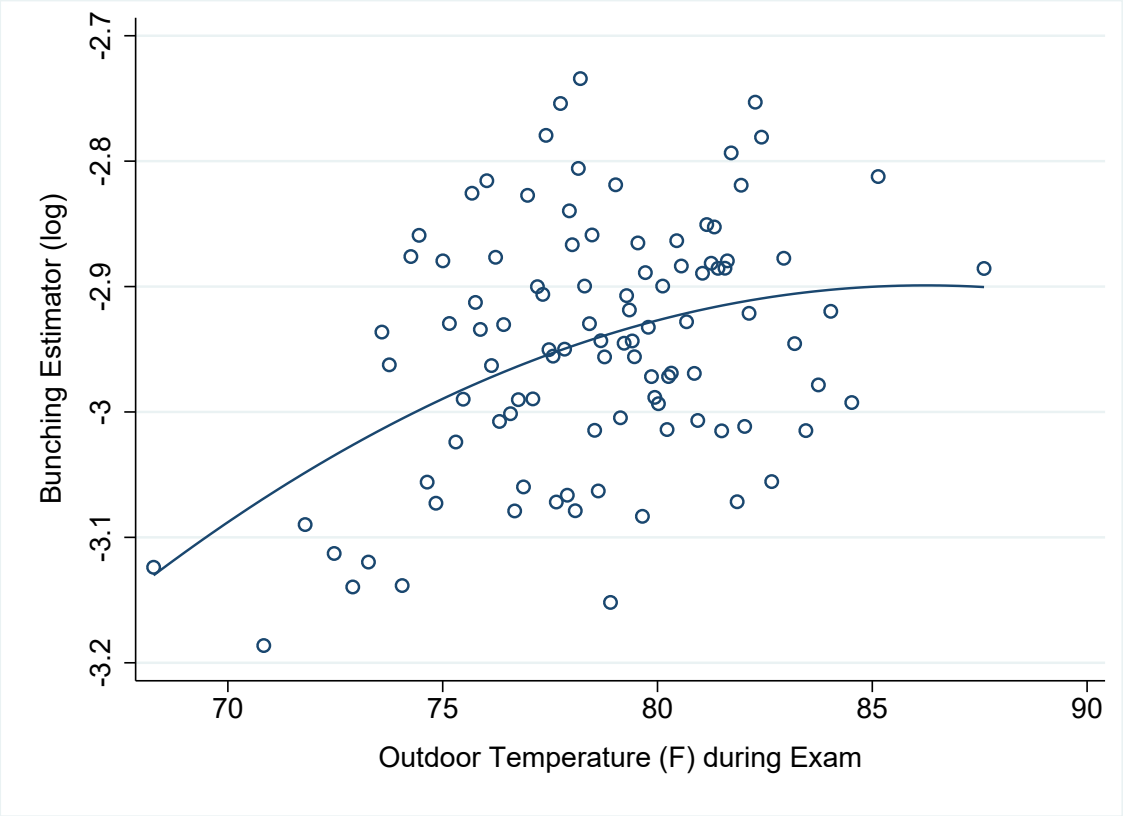
**Hot Temperature and High Stakes Performance: Online Appendix**
**R. Jisung Park**

Figure A.1: Exam-Time Temperature and Performance: Robustness to Outliers



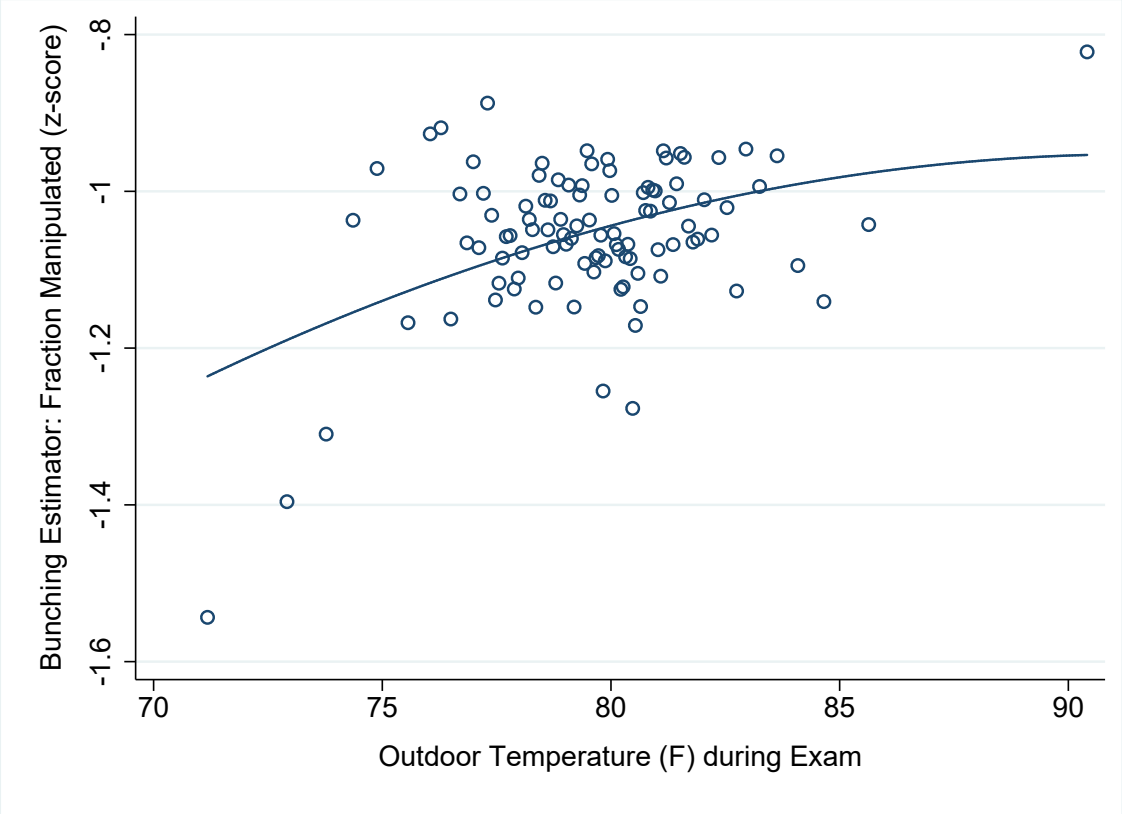

Notes: This figure presents a binned scatterplot of residualized exam performance by quantile of the temperature distribution controlling for school, subject, and year fixed effects, in addition to demographic observables, exam-day precipitation, air quality, and dewpoint. Each dot represents approximately 40,000 exam observations. Omitting observations with temperatures above 90°F.

Figure A.2: Exam-Time Temperature and Grade Manipulation: Robustness to Outliers



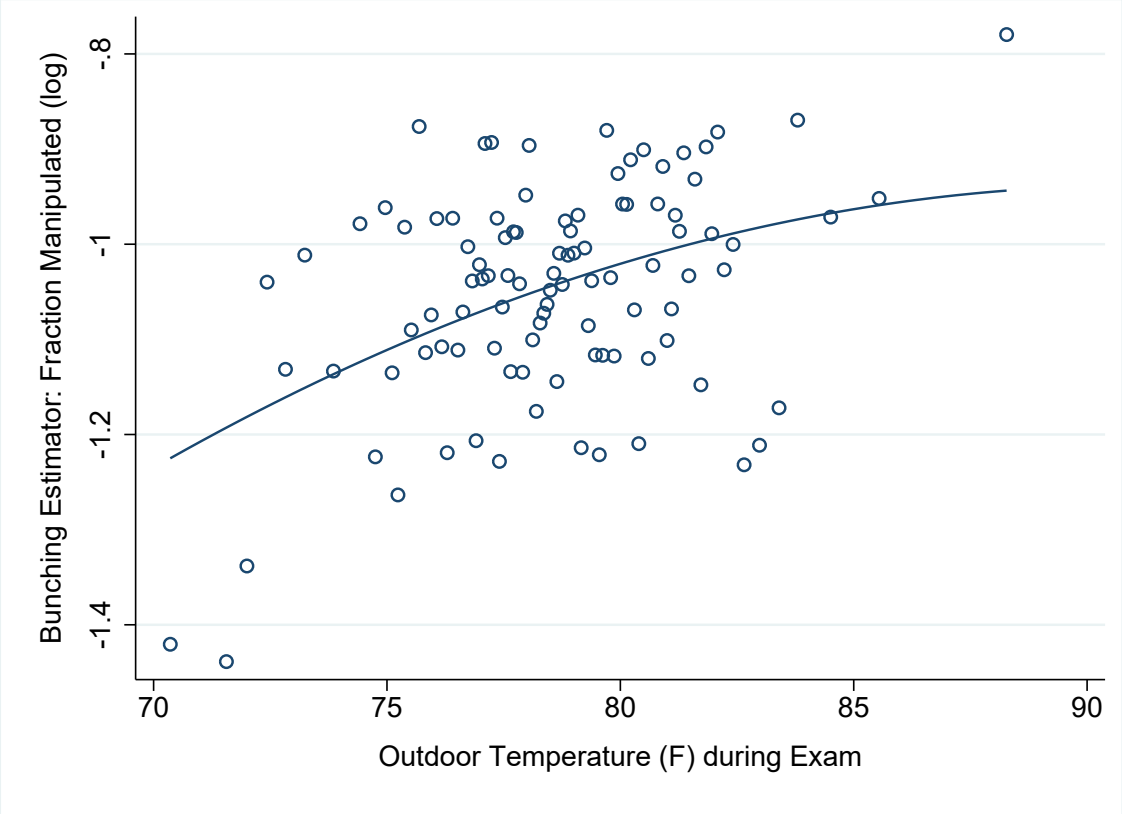

Notes: This figure presents a binned scatterplot of the natural log of the bunching estimator – the natural log of the estimated fraction of grades manipulated by school-subject-date – by quantile of the exam-time temperature distribution. This specification plots residuals after controlling for average differences in manipulation by subject, school, and year. Omitting observations with temperatures above 90°F.

Figure A.3: Exam-Time Temperature and Grade Manipulation: Fraction of Manipulable Scores Manipulated



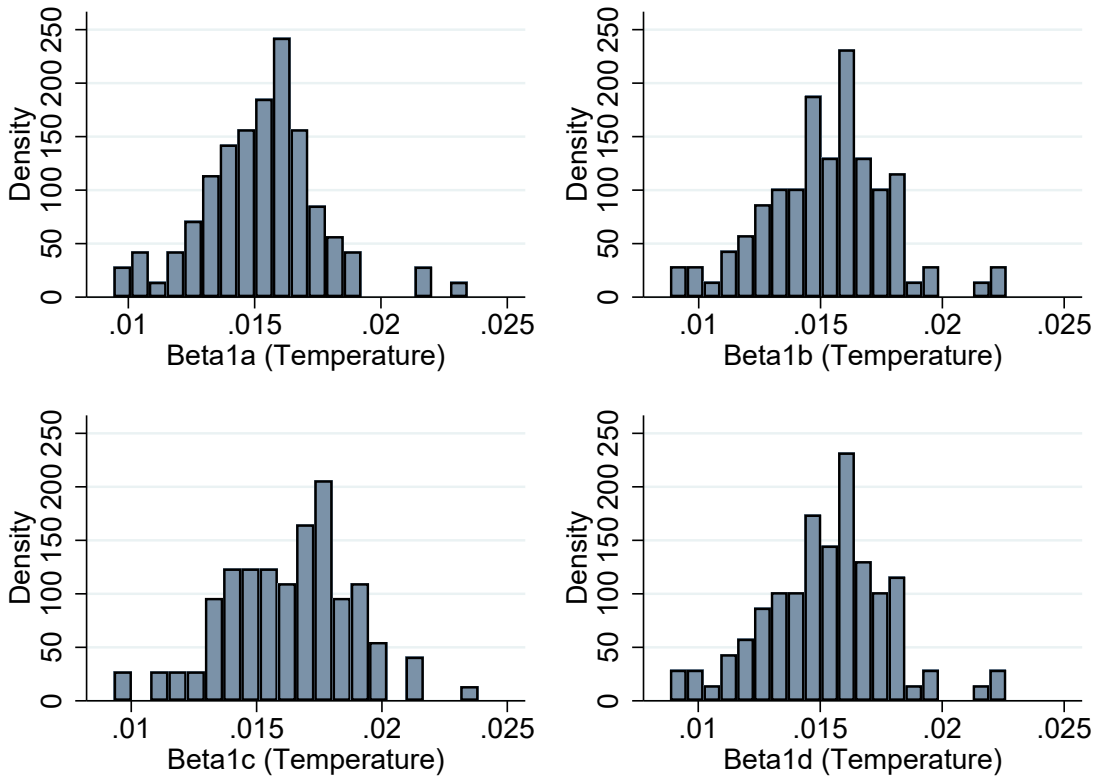

Notes: This figure presents a binned scatterplot of standardized values of the alternative measure of bunching – the natural log of the estimated fraction of manipulable grades manipulated by school-subject-date – by quantile of the exam-time temperature distribution. This specification plots residuals after controlling for average differences in manipulation by subject, school, and year, in addition to controls for the day of the week, daily precipitation and air quality, and observable demographic composition of each school each year (gender, ethnicity, subsidized school lunch status). Manipulation is estimated within school-subject-date cells using a cutoff rule described in section 6. Included in the analysis are all June Regents exams in core subjects between 1998 and 2011.

Figure A.4: Exam-Time Temperature and Grade Manipulation: Fraction of Manipulable Scores Manipulated (Robustness to outliers)



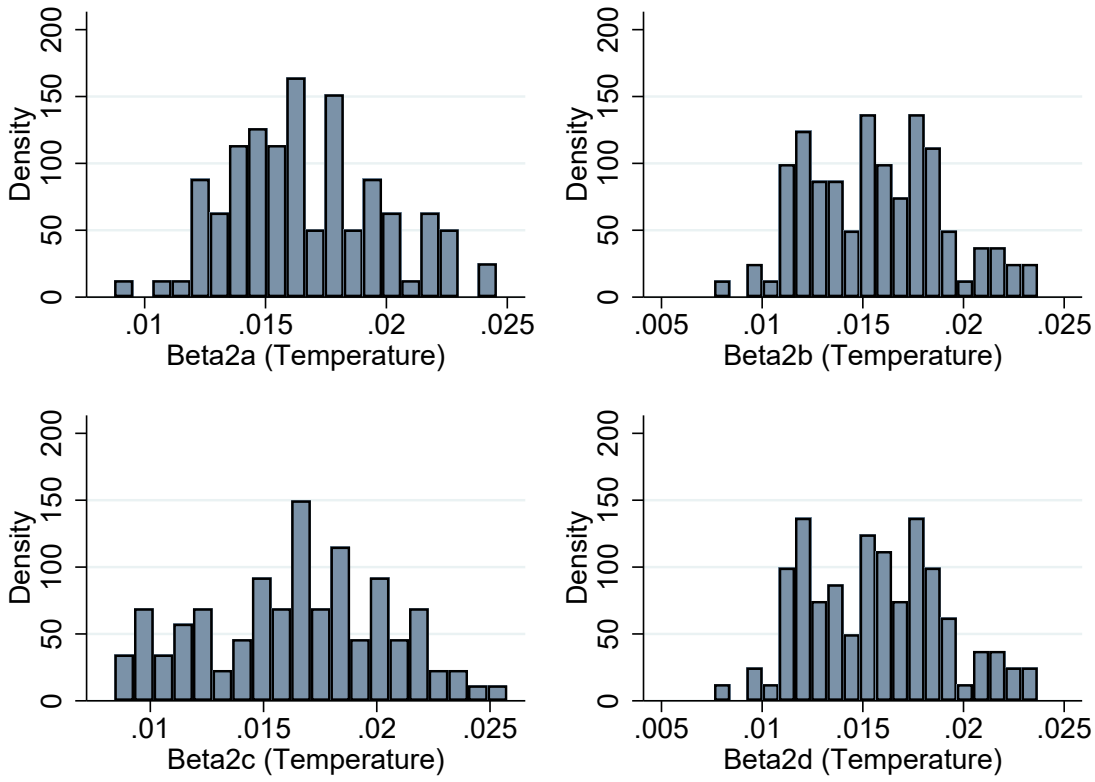

Notes: This figure presents a binned scatterplot of standardized values of the alternative measure of bunching – the natural log of the estimated fraction of manipulable grades manipulated by school-subject-date – by quantile of the exam-time temperature distribution. This specification plots residuals after controlling for average differences in manipulation by subject, school, and year, in addition to controls for the day of the week, daily precipitation and air quality, and observable demographic composition of each school each year (gender, ethnicity, subsidized school lunch status). Manipulation is estimated within school-subject-date cells using a cutoff rule described in section 6. Included in the analysis are all June Regents exams in core subjects between 1998 and 2011.

Figure A.5: Grade Manipulation: cluster-bootstrapped estimates (1)




Each of the above panels represents a histogram of beta-coefficient estimates from regressing bunching – the natural log of the estimated fraction of grades manipulated by school-subject-date – over 100 cluster resamples by school and date. Beta1a-Beta1d comes from regressing bunching on temperature and fixed effects and controls noted in the text for column 1-4 of table 5. Namely: fixed effects for subject, school, and year (Beta1a); subject, school, year, and day of week (Beta1b); subject, day of week, and school-by-year (Beta1c); subject, school, year, and day of week, with student-level demographic controls (Beta1d). Manipulation is estimated within school-subject-date cells using a cutoff rule described in section 6. Included in the analysis are all June Regents exams in core subjects between 1998 and 2011.

Figure A.6: Grade Manipulation: cluster-bootstrapped estimates (2)




Notes: Each of the above panels represents a histogram of beta-coefficient estimates from regressing the alternate measure of bunching – the natural log of the estimated fraction of manipulable grades manipulated by school-subject-date – over 100 cluster resamples by school and date. Beta2a-Beta2d come from regressing bunching on temperature and fixed effects and controls noted in the text for column 1-4 of table A.7. Namely: fixed effects for subject, school, and year (Beta2a); subject, school, year, and day of week (Beta2b); subject, day of week, and school-by-year (Beta2c); subject, school, year, and day of week, with student-level demographic controls (Beta2d). Manipulation is estimated within school-subject-date cells using a cutoff rule described in section 6. Included in the analysis are all June Regents exams in core subjects between 1998 and 2011.

Table A.1: Temperature and High Stakes Exam Performance: Robustness to Alternative Clustering

| | (1) Sensor and Date | (2) Sensor and Year | (3) School and Date | (4) School |
|---|---|---|---|---|
| Temperature (F) | -0.009* | -0.009** | -0.009*** | -0.009*** |
| | (0.004) | (0.003) | (0.003) | (0.001) |
| Afternoon | -0.030 | -0.030 | -0.030* | -0.030*** |
| | (0.021) | (0.021) | (0.018) | (0.006) |
| N | 3,581,934 | 3,581,934 | 3,581,934 | 3,581,934 |
| Fixed Effects | | | | |
| Student by Year | X | X | X | X |
| Subject | X | X | X | X |
| Day of week | X | X | X | X |

Notes: Heteroskedasticity robust standard errors are in parentheses (* p<.10 ** p<.05 *** p<.01). Clustering is by (1) sensor and date, (2) sensor and year, (3) school and date, and (4) school for columns (1)-(4) respectively. Coefficients in each column and panel come from a regression of Regents z-scores on the variables shown. The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.

Table A.2: Temperature and High Stakes Exam Performance: Subject-Specific Impacts

| | (1) Quant | (2) Verbal | (3) Quant | (4) Verbal |
|---|---|---|---|---|
| Temperature (F) | -0.013** | -0.009 | -0.012** | -0.008 |
| | (0.005) | (0.007) | (0.006) | (0.006) |
| N | 1,910,737 | 1,670,713 | 1,910,930 | 1,670,954 |
| Fixed Effects | | | | |
| School by Year | X | X | | |
| Day of week | X | X | X | X |
| Subject | X | X | X | X |
| School | | | X | X |
| Year | | | X | X |

Notes: Robust standard errors clustered by school and date in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of Regents z-scores on the variables shown, for (1)(3) Quant, (2)(4) Verbal subjects respectively. The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.

Table A.3: Temperature and High Stakes Exam Performance: Temperature Binning

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Temperature above 90F | -0.071 | -0.085* | -0.113* | -0.122* |
| | (0.046) | (0.049) | (0.061) | (0.066) |
| Temperature between 80F and 90F | -0.097** | -0.113*** | | |
| | (0.037) | (0.042) | | |
| Temperature between 70F and 80F | -0.048 | -0.057 | | |
| | (0.031) | (0.037) | | |
| Temperature between 85F and 90F | | | -0.137*** | -0.150*** |
| | | | (0.050) | (0.054) |
| Temperature between 80F and 85F | | | -0.128*** | -0.141*** |
| | | | (0.041) | (0.046) |
| Temperature between 75F and 80F | | | -0.081** | -0.087** |
| | | | (0.033) | (0.039) |
| Temperature between 70F and 75F | | | -0.045 | -0.054 |
| | | | (0.031) | (0.035) |
| N | 3,581,934 | 3,581,934 | 3,581,934 | 3,581,934 |
| Fixed Effects | | | | |
| Subject | X | X | X | X |
| Day of week | X | X | X | X |
| Year | X | | X | |
| School | X | | X | |
| School by Year | | X | | X |

Notes: Robust standard errors clustered by school and date in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of Regents z-scores on the variables shown. The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.

Table A.4: Temperature and High Stakes Exam Performance: Quadratic Specification

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Temperature (F) | -0.053* | -0.049** | -0.056** | -0.069** |
| | (0.028) | (0.022) | (0.026) | (0.034) |
| Temperature$^2$ | 0.000 | 0.000* | 0.000* | 0.000* |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Afternoon | -0.032* | -0.036* | -0.021 | -0.019 |
| | (0.018) | (0.018) | (0.022) | (0.020) |
| N | 3,581,934 | 3,581,934 | 3,581,934 | 3,581,934 |
| Fixed Effects | | | | |
| Student by Year | X | | | |
| Subject | X | X | X | X |
| Day of week | X | X | X | X |
| Student | | X | | |
| Year | | X | X | |
| School | | | X | |
| School by Year | | | | X |

Notes: Robust standard errors clustered by school and date in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of Regents z-scores on the variables shown. The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.

Table A.5: Temperature and Likelihood of Achieveing Proficient Status

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Temperature (F) | -0.004** | -0.003** | -0.006*** | -0.006*** |
| | (0.001) | (0.001) | (0.002) | (0.002) |
| Afternoon | -0.011 | -0.010 | -0.002 | -0.001 |
| | (0.007) | (0.006) | (0.009) | (0.009) |
| N | 3,581,934 | 3,581,934 | 3,581,934 | 3,581,934 |
| Fixed Effects | | | | |
| Student by Year | X | | | |
| Subject | X | X | X | X |
| Day of week | X | X | X | X |
| Student | | X | | |
| Year | | X | X | |
| School | | | X | |
| School by Year | | | | X |

Notes: Robust standard errors clustered by school and date in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column and panel come from a regression of a dummy variable for scoring above proficiency standards on an exam on the variables shown. The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.

Table A.6: Persistent Impacts of Temperature on Educational Attainment: Clustering

| | (1) School and Cohort | (2) School | (3) Sensor |
|---|---|---|---|
| Mean Temperature (F) | -0.008*** | -0.008*** | -0.008*** |
| | (0.001) | (0.001) | (0.002) |
| N | 515,199 | 515,199 | 515,199 |
| Fixed Effects | | | |
| School | X | X | X |
| Number of takes | X | X | X |
| Cohort | X | X | X |

Notes: Robust standard errors clustered by (1) school and cohort, (2) school, and (3) weather sensor in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column come from a regression of a dummy for graduation status on the variables shown. Temperature is measured with average exam-time temperature experienced by a student (at the school level) during exams preceding senior year of high school. All regressions include controls for observable demographic characteristics including ethnicity, subsidized school lunch status, and English language learner status.The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011. All regressions include controls for daily dewpoint, precip, ozone, and pm2.5.

Table A.7: Temperature and Teacher Grade Manipulation Robustness: Fraction of Manipulable Scores Manipulated

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Temperature (F) | 0.017 | 0.016 | 0.016 | 0.016 | 0.015 |
| Bootstrapped SD's | 0.00318 | 0.00343 | 0.00407 | 0.00342 | 0.00355 |
| Observations | 3,276,705 | 3,276,705 | 3,276,497 | 3,276,705 | 3,144,224 |
| Fixed Effects | | | | | |
| Subject | X | X | X | X | X |
| School | X | X | | X | X |
| Year | X | X | | X | X |
| Day of Week | | X | X | X | X |
| School by Year | | | X | | |
| Demographic controls | | | | X | X |
| Dropping outliers | | | | | X |

Notes: Standard deviations of the associated sampling distributions of beta are based on replications of the regressions across 100 cluster bootstrap resamples, clustered by school and date. Coefficients in each column come from a regression of the bunching estimator – which is the natural log of the excess mass above the passing threshold by school and exam take, as a proportion of the fraction of scores in the manipulable zones (50-54, 60-64) – on the variables shown. Temperature is measured at the school level by exam date and time. All regressions include controls for daily precipiation, dewpoint, ozone, and pm2.5. The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011. Column 5 reports results after dropping observations with temperature readings above 90F.

Table A.8: Temperature and Teacher Grade Manipulation: Robustness to Alternative Clustering

| | (1) Sensor and Date | (2) Sensor and Year | (3) School and Year | (4) School |
|---|---|---|---|---|
| Temperature (F) | 0.017* | 0.017* | 0.017* | 0.017*** |
| | (0.008) | (0.007) | (0.008) | (0.003) |
| N | 3,277,014 | 3,277,014 | 3,277,014 | 3,277,014 |
| Fixed Effects | | | | |
| Subject | X | X | X | X |
| School | X | X | X | X |
| Year | X | X | X | X |

Notes: Robust standard errors clustered by (1) weather sensor and date, (2) weather sensor and year, (3) school and year, and (4) school in parentheses (* p<.10 ** p<.05 *** p<.01). Coefficients in each column come from a regression of the bunching estimator – which is the natural log of the excess mass above the passing threshold by school and exam take, as a proportion of the fraction of scores in the manipulable zones (50-54, 60-64) – on the variables shown. Temperature is measured at the school level by exam date and time. All regressions include controls for daily precipiation, dewpoint, ozone, and pm2.5. The sample comprises all students in the New York City public high school system who took Regents exams during the years 1998-2011.