

Online Appendix for

Testing, Stress, and Performance: How Students Respond
Physiologically to High-Stakes Testing

Jennifer A. Heissel (jaheisse@nps.edu), Emma K. Adam, Jennifer L. Doleac,
David N. Figlio, and Jonathan Meer

Table A.1. Descriptive statistics for homeroom cortisol samples

	School 1	School 2	School 3
Hour of the day (hrs since midnight)	7.433 (0.240)	8.557 (0.193)	8.460 (0.174)
Waketime (hrs since midnight)	5.767 (0.591)	6.688 (0.736)	6.687 (0.810)
CAR sample (fraction during CAR)	0.030 (0.170)	0.046 (0.209)	0.073 (0.262)
Cortisol (in $\mu\text{g}/\text{dl}$)	0.280 (0.172)	0.274 (0.186)	0.238 (0.214)
Observations	202	241	55

Notes: Means by school weighted by number of observations of cortisol obtained during homeroom; standard deviations in parentheses.

Table A.2. Changes in level of before-testing homeroom period cortisol by week

	(1) No controls	(2) Add controls	(3) Add FE	(4) Post-double selection LASSO
Low-stakes testing	0.148 (0.090)	0.085 (0.085)	0.101 (0.087)	0.118 (0.088)
High-stakes testing	0.234** (0.081)	0.164* (0.080)	0.176* (0.076)	0.201* (0.081)
<i>Controls:</i>				
Time of day		-0.491 (0.580)	0.034 (0.644)	-0.300 (0.193)
Time of day-squared		-0.062 (0.583)	0.349 (0.682)	
Wake time		0.103 (0.074)	-0.183 (0.139)	
CAR timeframe		-0.221 (0.160)	-0.101 (0.175)	
Age in the fall		0.007 (0.040)		
Female		0.239** (0.082)		
Exceptionality		0.091 (0.120)		
Section 504 plan		-0.041 (0.093)		
McKinney-Vento Act		-0.096 (0.101)		
Constant	-1.640*** (0.072)	-2.582*** (0.650)	-0.529 (0.898)	-2.392*** (0.074)
Observations	489	489	489	489
Participants	93	93	93	93

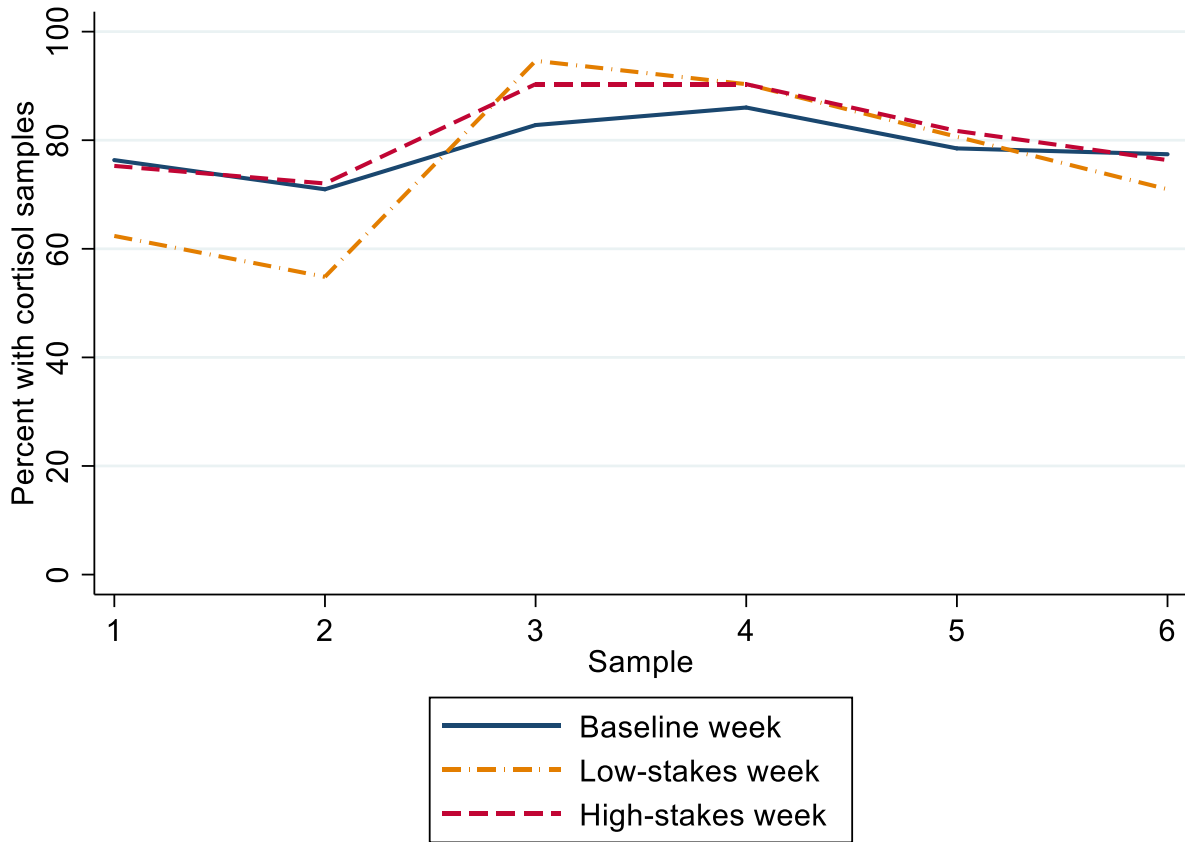
Notes: Robust standard errors clustered by student ID. Analysis conducted at the student-day level. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Outcome is the natural log of cortisol. Data comes from saliva collected in homeroom. Each column represents a different regression estimate. Model 1 includes no additional controls. Model 2 adds observable characteristics. Model 3 limits the comparison to within individuals, accounting for any constant observed and unobserved characteristics. Model 4 uses post-double-selection LASSO methods (Belloni, Chernozhukov, and Hansen 2014) to select a set of controls to avoid over-fitting but minimize omitted variable bias. Model 3 is the preferred overall model presented in the main paper.

Table A.3. Changes in test scores by cortisol responsivity to the test

	(1) Main model	(2) Main model	(4) Post-double selection LASSO	(5) Post-double selection LASSO
+/-10% from baseline	-0.443** (0.164)		-0.392** (0.143)	
10% above baseline		-0.437* (0.172)		-0.384** (0.149)
10% below baseline		-0.458* (0.192)		-0.406* (0.167)
<i>Controls:</i>				
Concurrent cortisol	-0.013 (1.453)	-0.103 (1.610)		-0.078 (0.292)
Concurrent cortisol-squared	0.123 (1.662)	0.195 (1.760)		
Math grade	0.013* (0.005)	0.013* (0.005)	0.014** (0.005)	0.014** (0.005)
ELA grade	0.017+ (0.009)	0.017+ (0.009)	0.013+ (0.007)	0.013+ (0.007)
Science grade	0.003 (0.008)	0.003 (0.009)	0.015* (0.006)	0.015* (0.006)
Social grade	0.025* (0.012)	0.025* (0.012)	0.012** (0.005)	0.013** (0.005)
Grades missing (0/1)	-0.932** (0.278)	-0.938** (0.284)		
Time relative to homeroom start	2.146 (2.797)	2.100 (2.820)		
Time-squared	1.709 (2.444)	1.658 (2.472)		
CAR timeframe	-0.809* (0.323)	-0.817* (0.329)	-0.569*** (0.169)	-0.575*** (0.169)
Average wake time	0.084 (0.104)	0.086 (0.103)		
Age	-0.052 (0.087)	-0.052 (0.088)		
Female	-0.089 (0.104)	-0.084 (0.110)		
Exceptionality	0.188 (0.258)	0.187 (0.261)		
Section 504 plan	0.120 (0.125)	0.118 (0.128)		
McKinney-Vento Act	0.240 (0.201)	0.240 (0.203)		
Constant	-1.586 (1.336)	-1.589 (1.344)	-2.374*** (0.196)	-2.351*** (0.206)
Observations	67	67	67	67
School FE?	Y	Y	N	N

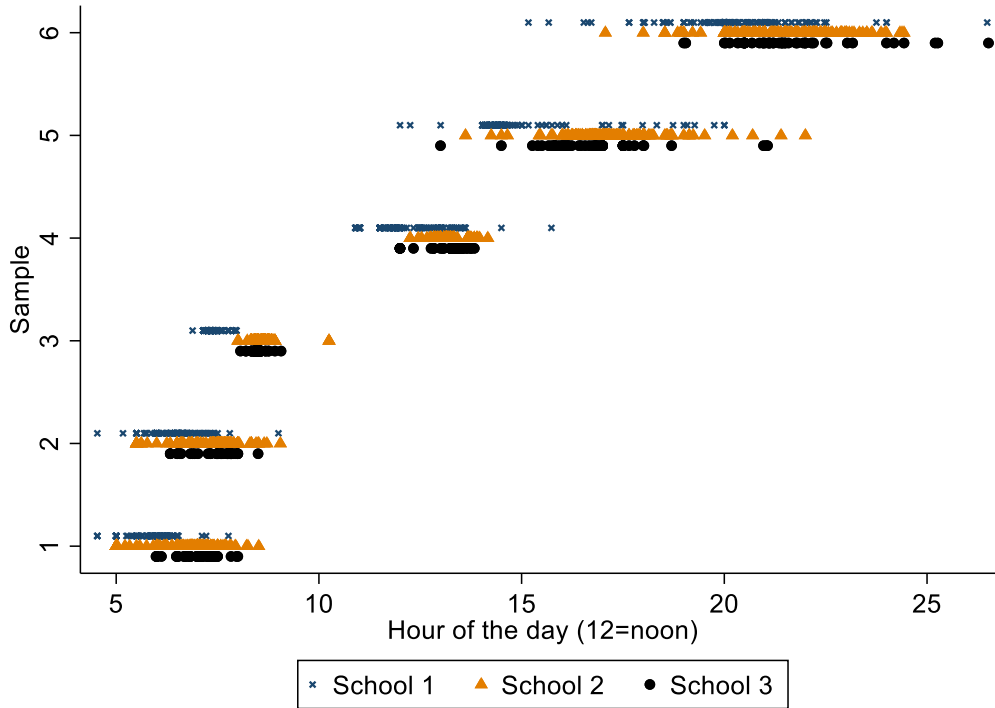
Notes: Robust standard errors. Analysis conducted at the student level. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Outcome is the z-score of the indicated test. Cortisol data comes from saliva collected in homeroom. Each column represents a different regression estimate. See Table 1 for control variable details. Model 1 is the main model to show the coefficients on the control variables. Model 2 separates the change by increases and decreases. Models 3-4 repeat Models 1-2, but use post-double-selection LASSO methods (Belloni, Chernozhukov, and Hansen 2014) to select a set of controls to avoid over-fitting. School fixed effects removed if the LASSO method indicated they were not needed. Model 1 is the preferred overall model presented in the main paper.

Figure A.1. Percent of participants with at least one sample by sample number and week.



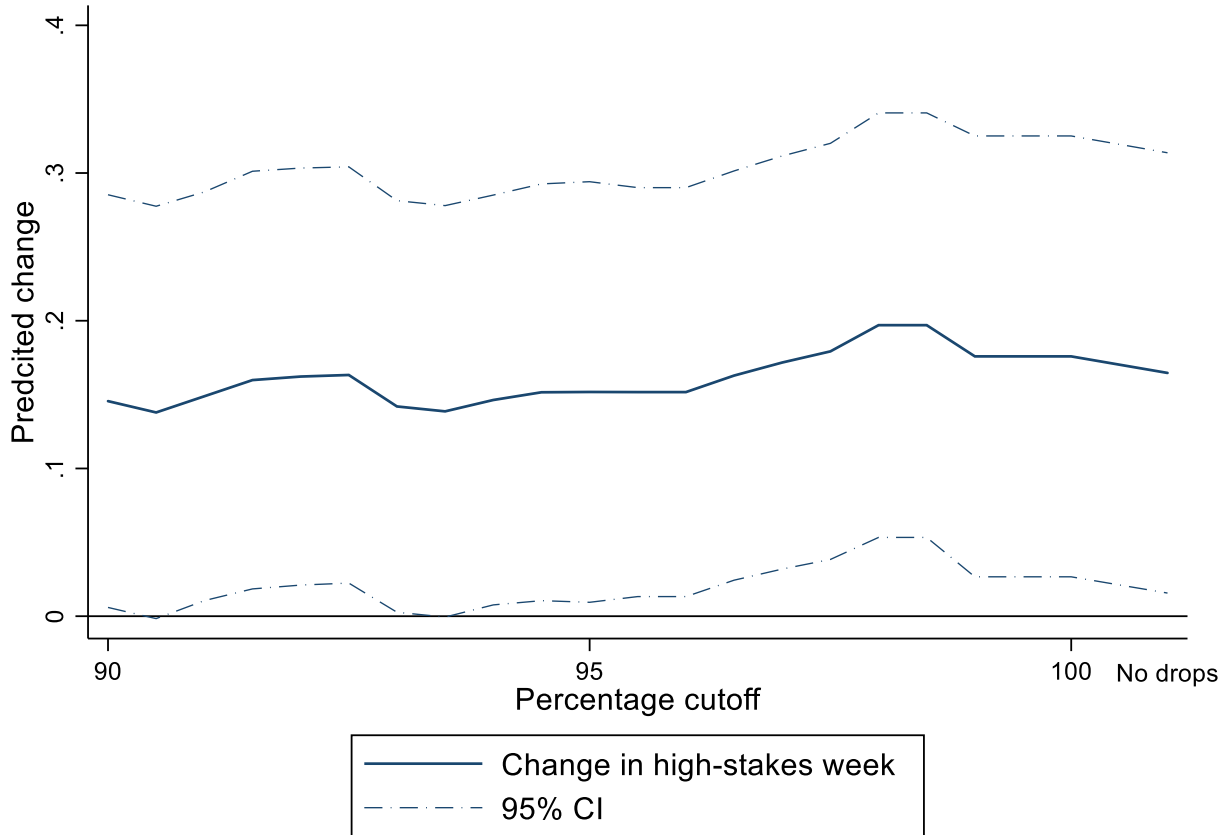
Notes: Study protocol specified Sample 1 as wake; Sample 2 as wake+30 minutes, Sample 3 as before-test (homeroom), Sample 4 as after-test (before-lunch), Sample 5 as after school, and Sample 6 as bedtime. Research team supervised and verified timing for collection in Sample 3 and Sample 4.

Figure A.2. Distribution of the timing of samples by collection period and school



Notes: Study protocol specified Sample 1 as wake; Sample 2 as wake+30 minutes, Sample 3 as before-test (homeroom), Sample 4 as after-test (before-lunch), Sample 5 as after school, and Sample 6 as bedtime. Research team supervised and verified timing for collection in Sample 3 and Sample 4; on-the-ground school needs meant that the timing of Sample 4 changed week-to-week. Sample 3 is the most consistently-timed sample. Homeroom began at 7am for School 1 and 8am for Schools 2 and 3. Times greater than 24 indicates a bedtime after midnight.

Figure A.3. The estimated change in cortisol from baseline to the high-stakes week for alternative methods of limiting potentially contaminated samples, including the estimate that includes all available cortisol (far right).



Notes: The 100th percentage cutoff excludes the two Week 1 ID's with consistently high cortisol indicating steroid medication; the limits for 90 to 99.5 drop any cortisol levels above those percentiles and also drop the two Week 1 ID's.

References

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* 81(2): 608-650.