Emoticons as Performance Feedback for College Students:

A Large-Classroom Field Experiment

By DARSHAK PATEL AND JUSTIN ROUSH

Online Appendix

A1 Sample Attrition:

At the start of the semester, students enrolled in both classes totaled 743: the control class had 283 students and the treatment had 460. After removing IRB non-respondents, students who chose to opt out of the study, and students who immediately dropped the course, we had 674 students enrolled in the study.

Table A1 summarizes sample attrition after the study began. Attrition occurred across the timeline of the study: 1 student did not take the first exam, 9 students took the first but not the second exam, and 70 students did not pick up their graded first exam. We eliminated students who did not pick up their exam from the study because they did not receive feedback (RPF in control or RPF plus emoticon in treatment). We were left with 594 students in the study: 230 in the control group and 364 in the treatment group.

Table A1: Sample Attrition by Treatment

	Control Class	Treatment Class	Total
No Exam 1	1	0	1
No Exam 2	7	2	9
Did Not Retrieve Exam 1	39	31	70
Total	47	33	80

There was less sample attrition in the treatment group, but our study is robust to sample selection bias through the difference-in-differences design. We did not compare average class outcomes across treatment and control groups, but rather, differences in outcomes for the same student. In other words, we employed a within-subjects and not between-subjects design.

A2 Sample Demographics:

Table A2: Demographics by Treatment

	(1) Full Sample	(2) Control	(3) Treatment
Male	0.4314	0.4183	0.4391
	(0.4957)	(0.4945)	(0.4970)
Freshman	0.09441	0.1916***	0.03631***
	(0.2926)	(0.3945)	(0.1873)
Sophomore	0.7238	0.5234***	0.8436***
	(0.4475)	(0.5006)	(0.3638)
Junior	0.1241	0.2009***	0.07821***
	(0.3300)	(0.4016)	(0.2689)
Senior	0.04729	0.06542*	0.03641*
	(0.2124)	(0.2478)	(0.1876)
EconMajor	0.08084	0.1132**	0.06162**
	(0.2728)	(0.3176)	(0.2408)
FirstEcon	0.9495	0.9398	0.9553
	(0.2192)	(0.2384)	(0.2069)
White	0.8339	0.8599	0.8187
	(0.3725)	(0.3479)	(0.3858)
Black	0.07308	0.06763	0.07627
	(0.2605)	(0.2517)	(0.2658)
Hispanic	0.04286	0.02427**	0.05367**
	(0.2027)	(0.1543)	(0.2257)
Asian	0.04100	0.04831	0.03672
	(0.1985)	(0.2149)	(0.1883)
Other Race	0.02143	0.009709*	0.02825*
	(0.1449)	(0.09829)	(0.1659)
HoursTaking	15.087	15.042	15.115
	(2.2874)	(2.5069)	(2.1487)
Job	0.4396	0.5023***	0.4017***
	(0.4968)	(0.5012)	(0.4909)
GPA Categorical	4.1429	4.000***	4.2244***
	(0.8844)	(0.9539)	(0.8325)
Observations	574	216	358

Given our sample of 594 subjects, an additional 20 did not fill out a demographic survey. The treatment and control groups differ across a number of dimensions (statistical significance is as follows: * 10%, ** 5%, *** 1%). These differences are controlled for by individual fixed effects in our difference in differences estimation.

A3 Robustness Check: Propensity Score Matching

Table A3: Propensity Score Matching

	Dependent Variables:				
	Attend	**Quiz Score**	**Homework Attempt**	**Homework Score**	**Log(Test)**
ATE	0.0986***	0.262***	-0.0108	4.885***	0.00784
	(0.0121)	(0.0426)	(0.0130)	(0.661)	(0.0211)
N	3850	2939	2200	2069	550

Robust standard errors in parentheses. Significance Levels: * p<0.10, ** p<0.05, *** p<0.01

The above table presents the average treatment effects from a propensity score matching estimator. In the first-stage propensity score logit estimation, we utilize all summarized demographics from Table A2, a measure of total attendance for each student (not included in the Attend specification), as well as dummy indicators for which emoticon range the student's exam 1 score fell into: Emoji1=1 for scores 68 or below (receiving a ☹ in treatment), Emoji2=1 for between 68 and the class average (receiving a 😐 in treatment, but scoring below the mean), Emoji3=1 if above the average and under 90 (receiving a 😐 in treatment, but scoring above the mean), and Emoji4=1 if scoring at or above 90 (receiving a ☺ in treatment). We match on total attendance, as well as sample demographics, because 8am versus 12:30pm classes may have different attendance dynamics, which may also influence performance. Statistical significance is in agreement between this estimator and our main difference-in-differences estimator with the exception of exam scores. All significant treatment impacts are larger.

A4 Robustness Check: High-Attendance Subsample Diff-in-Diff

Table A4: High and Perfect Subsample Difference-in-Differences

	High Attendance Subsample (2 or Less Absences)				Perfect Attendance Subsample			
Dependent Variables:	**Quiz Score**	**Homework Attempt**	**Homework Score**	**Log(Test)**	**Quiz Score**	**Homework Attempt**	**Homework Score**	**Log(Test)**
Post	0.0117	-0.0255***	-0.633	-0.159***	0.0352	-0.0250**	-0.633	-0.158***
	(0.0394)	(0.00868)	(0.641)	(0.0140)	(0.0501)	(0.0115)	(0.641)	(0.0197)
TreatmentXPost	0.116**	-0.00314	1.674**	0.0122	0.108*	0.000852	1.674**	0.0167
	(0.0457)	(0.0108)	(0.714)	(0.0160)	(0.0580)	(0.0140)	(0.714)	(0.0221)
N	6858	3536	3441	884	3936	1968	3441	492

Robust standard errors in parentheses. Significance Levels: * p<0.10, ** p<0.05, *** p<0.01

One concern is that attendance in the 8am courses would have trended away from attendance in the 12:30pm course naturally, even in the absence of treatment. If attendance is tied to performance, then our treatment effect is biased. Instead of picking up the impact of E-RPF in treatment, we instead estimating the performance premium on attendance. To address this, we re-estimated our overall results (Table 2) on subsamples with “high attendance” (2 or less absences; 84% of treatment and 60% of control) and with perfect attendance (48% in treatment, 30% in control). If we assume improved quiz, homework, and test scores are only the result of natural trending in absences among the 8am students as well (that is, our treatment has no impact), then only comparing treated versus control students with high attendance should result in no remaining differences in performance.

Table A4 above plots the results from each subsample and they generally agree with our full-sample estimates. The improvements in quiz and homework scores are maintained within those who attended class and received treatment. Test score improvements become statistically insignificant, although impact magnitudes are similar.

A5 Outcome Trends by Treatment:

Difference in differences estimators assume the existence of parallel trends in the outcome variable in the absence of treatment. This is usually accounted for by demonstrating that prior to treatment the difference in outcome variables for treatment and control remain constant over time. In this section we present plots of average outcomes (and 95% confidence intervals) by sample over time. In general, we find no evidence for differential pre-treatment trending.

Figure A1 plots attendance trends. Figure A2 plots daily quiz scores. In our analysis of attendance and quiz scores, day 17 is dropped as it is an outlier. This was a Tuesday following Halloween and attendance dropped to around 77% for both classes and the instructor awarded 5 points on the quiz regardless of performance. In both figures it is not evident that control is trending differently than treatment prior to receiving exam feedback.

Figure A1: Attendance Trends

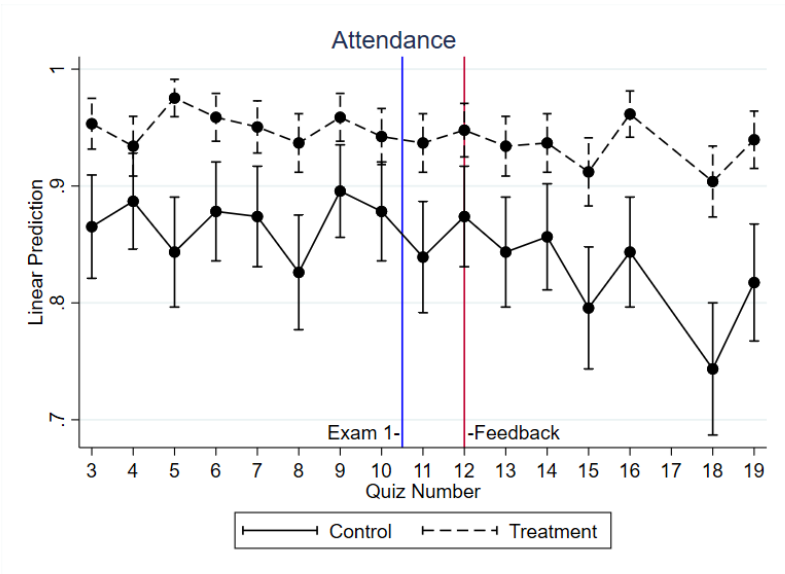
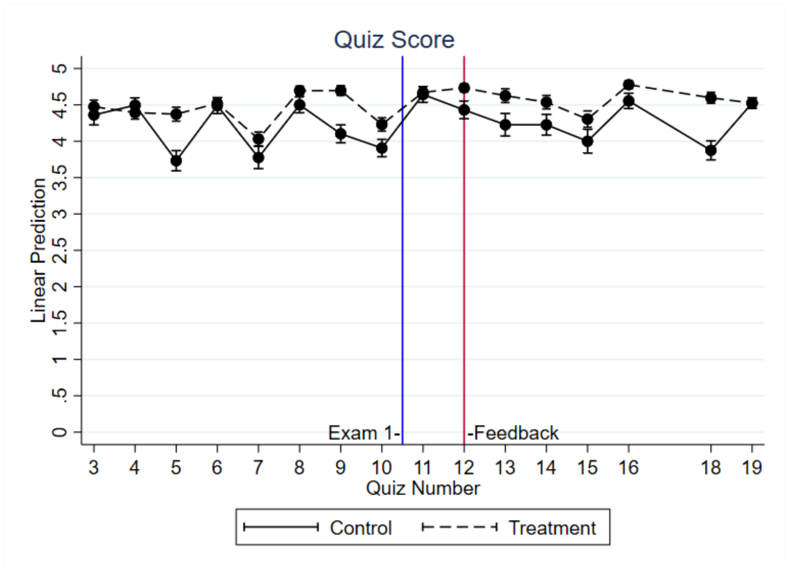


Figure A2: Quiz Score Trends



Figures A3 and A4 plot trends in homework attempt percentage and homework scores (conditional on attempting the homework) over time. In our analysis of homework scores, homework 8 is an outlier. Notably, performance on the homework converges between treatment and control. This is because the instructor created an in-class workday for the 8th homework set. Those who attended received 100 as the whole class worked together. Our results persist with and without this date included in the analysis.

The pre-trends for homework scores do not demonstrate ideal parallel trends. It appears that for the first three homeworks, control is trending away from treatment. But, this trending difference is broken by homework 4, in which the difference between treatment and control performance looks much more similar to homeworks 1 and 2.

With the exception of the aforementioned "class-completed" homework 8, the average difference between treatment and control homework performances widens after treatment, suggesting a treatment effect. If in fact the pre-trends are not parallel, this increase in relative performance may be biased upwards.

Figure A3: Pct. Homework Attempted Trends

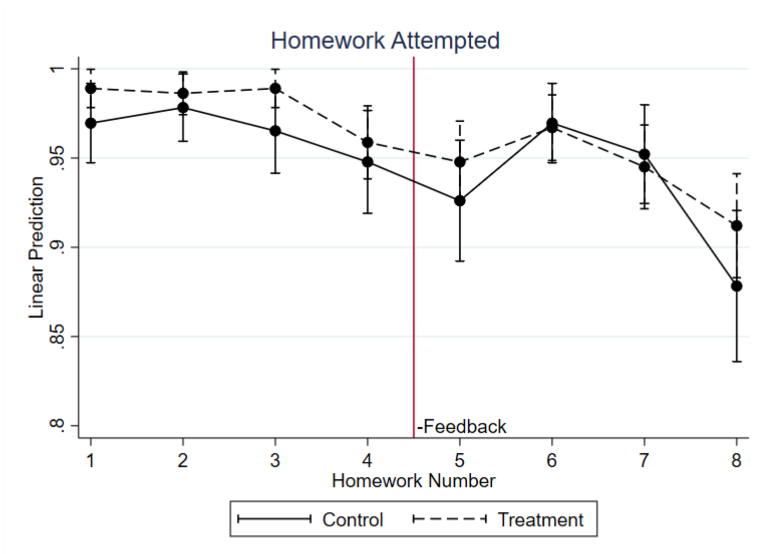
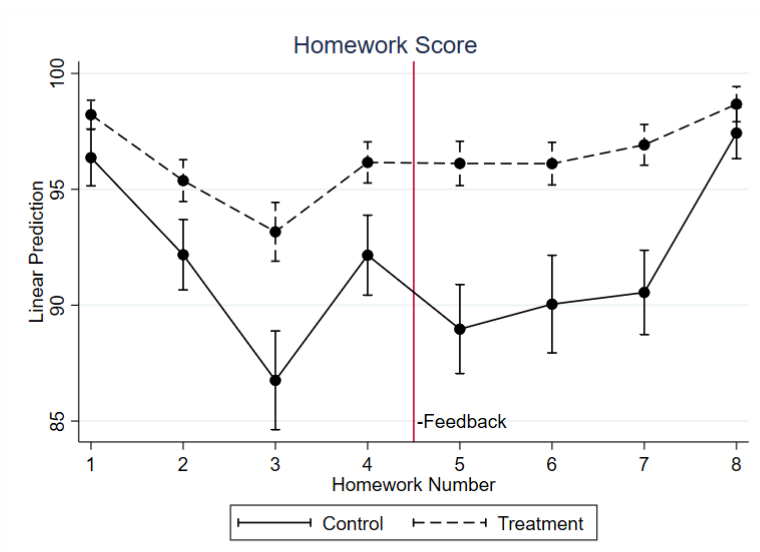
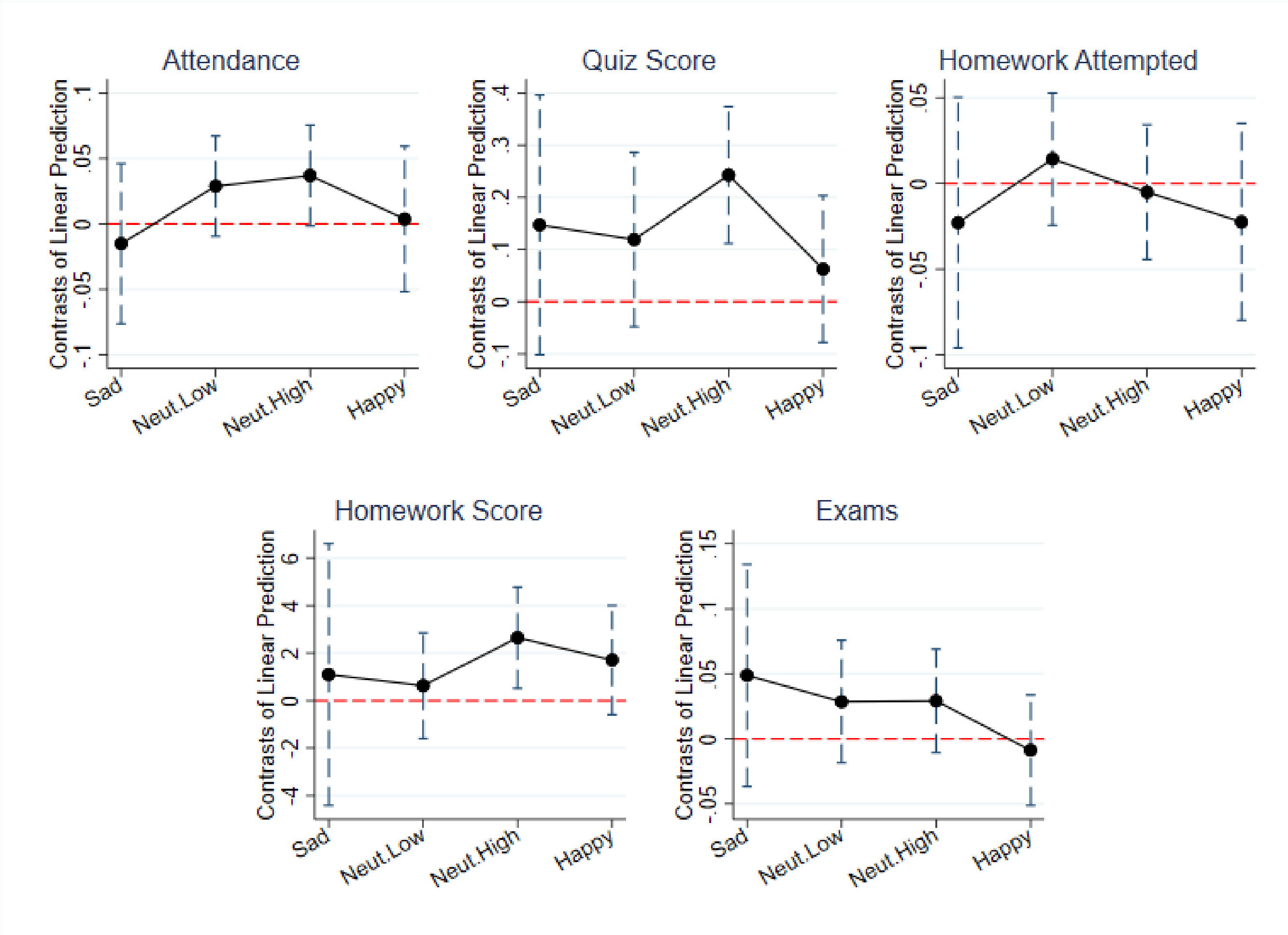


Figure A4: Homework Score Trends



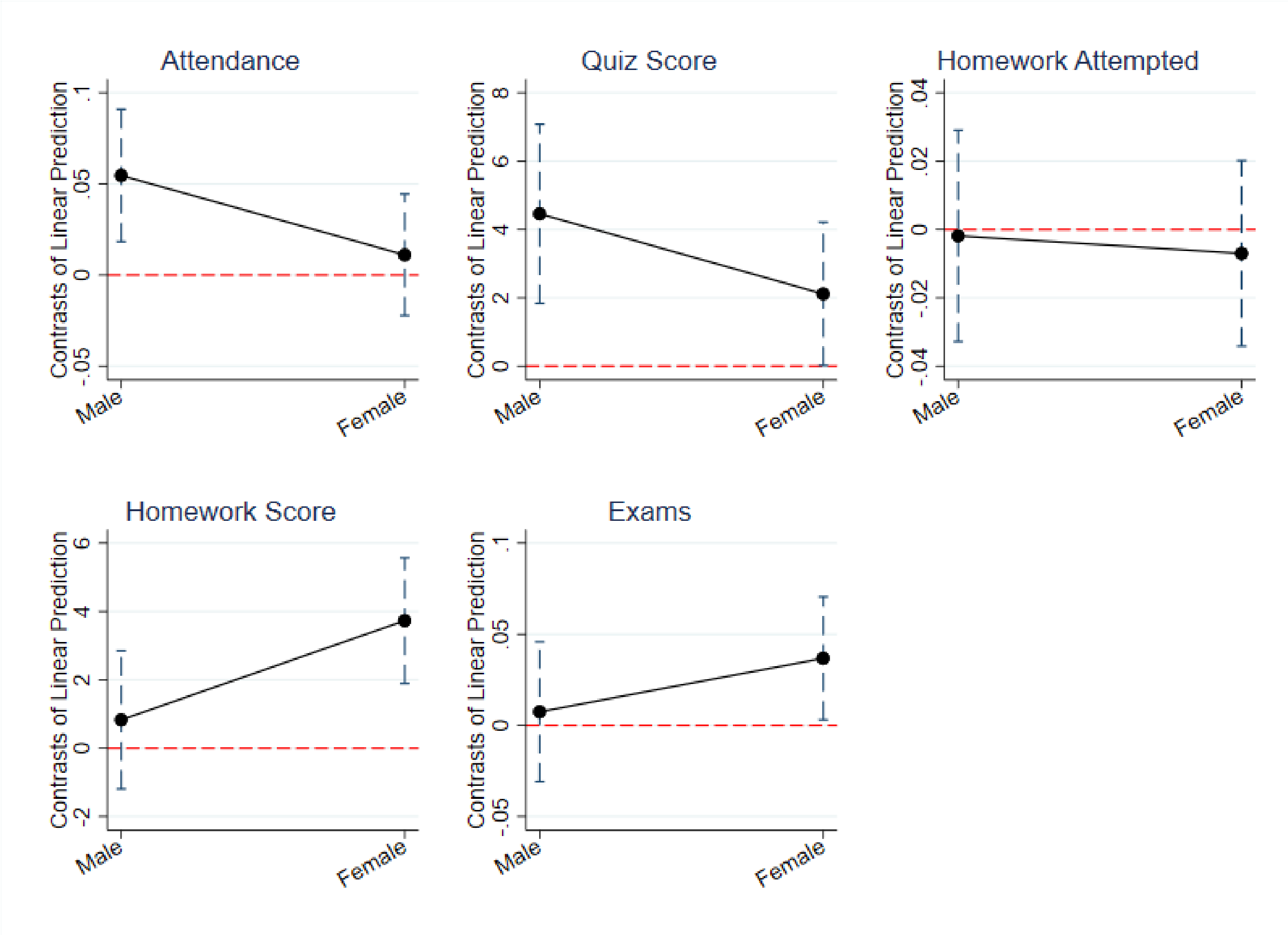
A6 Results by Treatment Level

Figure A5: Marginal Effects of Treatment by Type of Emoticon Received



A7 Results by Gender

Figure 6: Marginal Effects by Gender



Appendix References:

Dobkin, C., Gil, R. and Marion, J. 2010. Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment. *Economics of Education Review*, 29(4): 566–75.

Marburger, D. R. 2006. Does mandatory attendance improve student performance?. *Journal of Economic Education*, 37: 148–55.

Self, S. 2012. Studying Absenteeism in Principles of Macroeconomics: Do Attendance Policies Make a Difference?, *The Journal of Economic Education*, 43(3): 223-234.

Data Source:

Patel, Darshak, and Justin Roush. 2023. "Data for: Emoticons as Performance Feedback for College Students: A Large-Classroom Field Experiment" *American Economic Association Papers & Proceedings* [publisher], OICPSR [distributor]. https://doi.org/10.3886/E 183985 V1