

Online Appendix: Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not

By SABRIN BEG, ADRIENNE M. LUCAS, WAQAS HALIM, AND UMAR SAIF

ONLINE APPENDIX

A1. *Additional Program Details*

We first provide additional details common to the two programs, then additional details specific to each program.

eLearn

eLearn was developed and implemented by the Punjab provincial government, Information Technology University Lahore, and the Punjab Information Technology Board, an autonomous department under the Planning and Development Department of the Punjab government. Videos were organized on the tablet by unit and the user could select which, if any, videos within a unit to watch. Some videos were primarily Urdu and others primarily English, as is typical in Pakistani middle schools where instruction occurs in a mix of these two languages. The presenters were experienced teachers, former teachers, university professors, and government officials working in the education sector.

eLearn Classrooms

The total content was 29 hours and 192 videos. The math content was a total of 12.4 hours divided among 77 videos with an average length of 10 minutes per video. For science, the total video length was 16.5 hours, spread across 115 videos with an average length of 8 minutes. The videos were designed to cover all topics of the curriculum. A single presenter, in all cases a government employee, appeared in all videos related to a particular unit. Men were the subject experts for 21 of the 22 units. The only female presenter appeared in the environment unit of the science curriculum. This technology was likely novel to some of the students in the sample. At our baseline, 40 percent of students reported having a computer at home and 30 percent reported using some sort of technology as a study aid at home.

As designed, at the conclusion of each unit, students should have received an SMS on their households' mobile phones that an Intelligent Tutoring System (ITS) module was available for use. This ITS system would have allowed students to use text messages to receive and respond to review questions. While almost all study students reported having at least one mobile phone in their household and 80 percent reported having more than one, due to delays this system was barely implemented during our study. Only one third of the treatment schools received

at least one module, consistent with the 25 percent of students in the treatment group who reported receiving at least one module. Among those students who received at least one module, only 10 percent, or 2.5 percent of the overall treatment group, received more than 3 of the 22 intended modules. Finally, the mobile carrier who broadcast the SMSs incorrectly charged the students to respond to these texts, leading to low take-up even among those who were reached. An interactive voice response system (IVR) was to call parents when a child missed school to inform them that their child was absent and allow parents to respond with the reason for the absence. The IVR system was not operational during our period of study. Therefore, while at home engagement was designed to be a component of the intervention, this piece was at most minimal.

eLearn Tablets

The total content was 16 hours divided among 234 videos. The science content was 13 hours divided among 202 videos with an average length of 3.8 minutes. The math content was 3 hours divided among 32 videos with an average length of 3 minutes. Gender balance in presenters was more equal in the Tablets intervention and the same presenter did not appear in all videos related to the same unit. Half of the units of the math curriculum had only male presenters and only one unit in science had only male presenters.

A2. Additional LASSO Details

For eLearn Classrooms, the 298 item set of potential controls for the LASSO specification are mean PEC scores, enrollment, number of sections, total present; all relevant teachers' genders, tenure, ages, qualification and experience; school fees, indicators of school facilities, school problems and learning hurdles identified by teachers and head teaches, trainings received by teachers; teachers' time used for class, non-classroom tasks, private tuitions and preparation; teachers' contract status, student demographics (age, parents education and qualification, siblings), parent-teacher meetings, student transportation, schooling expectation, and access to books and resources. All categorical variables are included as individual dummies for each category and squares of all continuous variables are also included. The LASSO method selected teacher employment rank, time spent on non-classroom duties and extra classes, mothers occupations, and parents relationship status. For eLearn Tablets the set of controls is approximately the same but because some questions had more possible response the total number possible increased to 353. The LASSO method selected teacher trainings, number of classes per week taught by teachers, teacher employment status and rank, teacher and student meetings out of class, language used by teacher, parent engagement, enrollment observed in class, and student time for weekly homework.

A3. Additional Testing Details

Our project-specific exams were designed by subject experts, not particularly involved with the design and implementation of the program, to cover the standard curriculum and be conceptual and less prone to rote memorization, criticisms of the PEC exam and other similar provincial exams in Pakistan (??). Each test included 80 grade level questions. The baseline and follow-up exams had the same questions, but in a different order. During baseline administration, enumerators invigilated the exams without teachers present and were careful not to leave any materials behind for teachers to see nor let teachers know the content of the exams. Students in both the control and treatment groups would have had the same level of familiarity with the exam at the follow-up. Appendix Figure A1 displays the baseline test score distributions for the project exams.

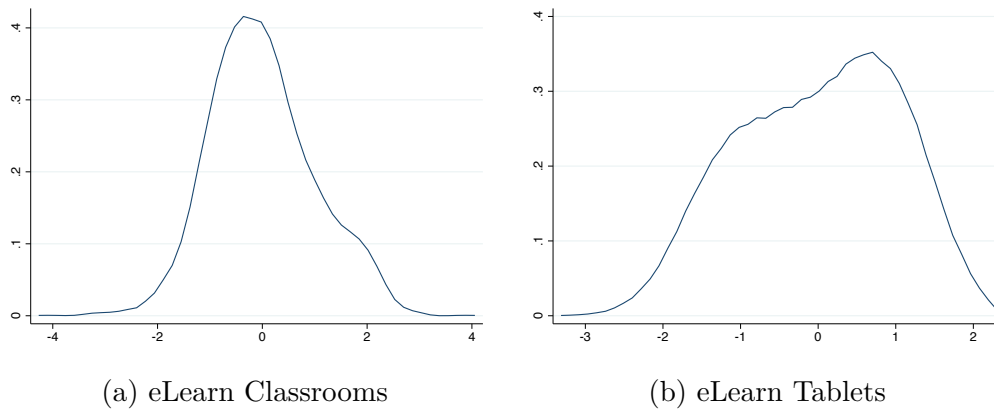


Figure A1. : Baseline Test Score Distributions

Note: Test scores standardized have a mean of 0 and standard deviation of 1. Horizontal axis is the standardized test score.

A4. Additional Summary Statistics, Figures, and Estimations

Appendix Table A1 contains additional summary statistics.

Appendix Table A2 uses alternative controls than those that appear in the main paper.

Appendix Table A3 provides the subject specific test score estimates.

Appendix Table A4 provides the Lee (2009) bounds for the achievement estimates.

Appendix Table A5 tests for changes in additional teacher inputs.

Appendix Table A6 presents tests for heterogeneity in effects by student baseline test score (Panel A), school baseline test score (Panel B), and school gender

(Panel C). The first two are discussed in the main paper. Appendix Figure A2 shows the non-parametric treatment effects by baseline test score. We examine heterogeneity non-parametrically as a function of baseline test scores. The test score lines for treatment and control lines are kernel-weighted locally-smoothed means of the endline test scores at each percentile of the baseline test-score distribution. The treatment effect for each baseline percentile is calculated as the difference between the treatment and control. The x-axis is the percentile of the residual of a regression of baseline test score on student and school characteristics (the same characteristics from the handpicked controls regression). The y-axis is the residual of a regression of the endline score on the same student and school characteristics.

When considering heterogeneity by gender, for the Classrooms intervention project test, the main effect is positive, and even though the interaction effect is negative, we reject at the 10 percent level that the sum is 0—the intervention increased test scores for both male and female students (Appendix Table A6, Panel C, column 1). In contrast, for the PEC score, the main effect is not statistically significant, the interaction is positive, and we reject (at the 1 percent level) that the coefficients sum to 0. The combined project and PEC score has both a positive (but statistically insignificant) main effect and (statistically insignificant) interaction effect. We reject at 1 percent that the overall effect on female students is 0. Therefore, while the intervention was not designed to favor students of a particular gender, instead providing expert content to assist all students, female students' scores increased on the PEC exam but males students' scores did not.¹

For the Tablets intervention, the main effect is negative and statistically significant, the interaction is negative and insignificant, and we reject that the overall effect on female students is 0. (Panel C, column 4).

As schools are single gender, the heterogeneous score gains by gender in the Classrooms intervention could be the result of both student and school characteristics that differ by gender. At the school level, female and male schools and teachers at those schools are statistically indistinguishable except female schools have a higher percentage of female teachers and higher average baseline test scores.²

At the student level, female students are statistically different than male students: they are more likely to expect to go to college (by 22 percentage points), younger (by 0.3 years), richer (households have 0.06 more cars), and less likely to work (by 4 percentage points). Some of these differences are likely due to selection. Nationwide, girls were 13 percentage points less likely to complete primary school and were only 38 percent of grade 8 students in 2016 (Government of Pakistan 2016). Therefore, female students who were still attending school in grade 8 in Pakistan are a more highly selected sample than male students.

¹When comparing the performance of students by gender on low-stakes and high-stakes exams in China, ? found that female students under performed on the high-stakes exams. In our setting, female students gain more from the intervention on the high-stakes PEC exam.

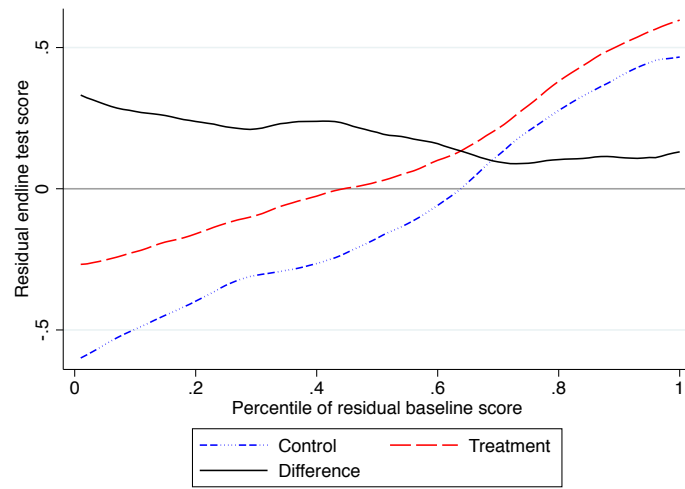
²The findings of heterogeneity by baseline test score holds within gender.

We find minimal differences by school gender in effort and implementation, with statistically significant differences by treatment status and gender only for the likelihood that parents have visited the school and university aspirations in the Classrooms intervention. Therefore, any differences by gender appear to be something about the interaction between the program and the students and not about the level of implementation or other effort changes.

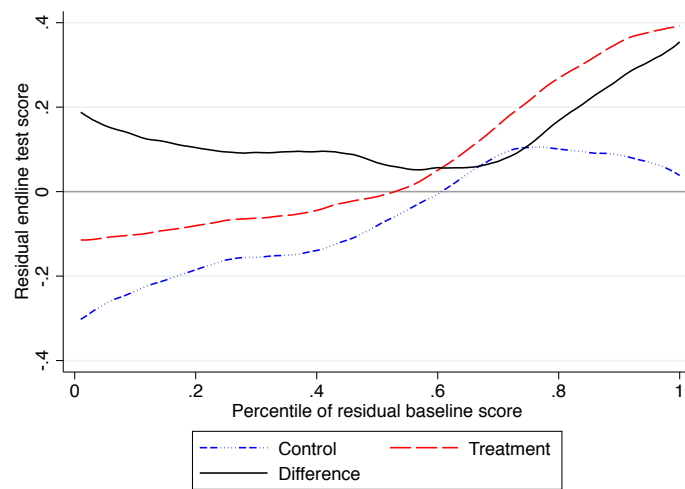
While the two differences in other outcomes between genders are likely not causing any gender heterogeneity in achievement, they are of note. The treatment increased the likelihood that male students reported that their parents had visited their school by 23 percentage points, while the treatment effect is statistically insignificant for females. Prior to the intervention this outcome was 4 percentage points higher for male versus female students (61 percent for males vs. 57 percent for females). Male students in treatment schools also increased their expectations regarding attending university by 17 percentage points with no statistically significant effect for girls. Prior to the intervention female students were 22 percentage points more likely to expect to attend college (50 percent for males vs. 72 percent for females). This program did not target either of these outcomes. Instead, an accidental side effect might have resulted from the gender of the experts on the videos. Of the 22 subject experts, 21 were men. Therefore, while we cannot directly test the mechanisms, these findings are consistent with the importance of a gender matching role model in future aspirations, which potentially led parents of boys to be more likely to visit school and boys to aspire to higher education.

Appendix Table A7 contains separate estimates the effect sizes for questions by quartile of the question difficulty for Classrooms. First, we used IRT to determine the difficulty of each question in the baseline. Then, we created a separate test score for each student based only on the questions in each quartile of difficulty, i.e. the easiest quartile, the second quartile of difficulty, the third quartile of difficulty, and the most difficult questions. We estimate the treatment effect separately for each newly created test score, finding statistically significant treatment effects in the questions of below median difficulty.

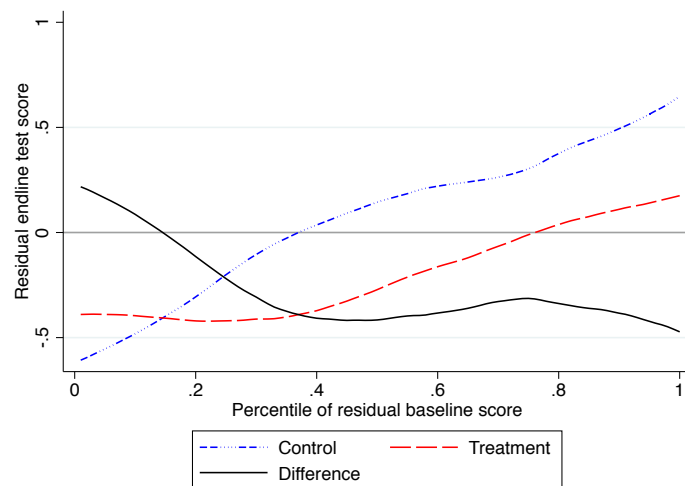
Appendix Table A8 tests for heterogeneity by whether one of the teachers in the school had below-median years of teaching experience (column 1) or whether at least one of the study teachers had a teacher peer who taught the same grade-level and subject (column 2). Students with inexperienced teachers and those with a teacher without a grade-level teacher peer gained relatively more.



(a) eLearn Classrooms Project Test



(b) eLearn Classrooms PEC Exam



(c) eLearn Tablets

Figure A2. : Nonparametric Treatment Effects

Table A1: Additional Summary Statistics

	eLearn Classrooms			eLearn Tablets		
	Treatment	Control	Difference T-C	Treatment	Control	Difference T-C
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Student Characteristics</i>						
Number of Siblings	3.95 (1.83)	3.80 (1.77)	0.15 (0.16)	4.56 (2.05)	4.60 (1.94)	-0.05 (0.07)
Receives Private Tutoring	0.46 (0.50)	0.47 (0.50)	-0.02 (0.05)	0.35 (0.48)	0.38 (0.49)	-0.03 (0.08)
Owns All Course Books	0.99 (0.08)	0.99 (0.12)	0.01 (0.00)	0.98 (0.16)	0.97 (0.18)	0.01 (0.10)
Works	0.03 (0.18)	0.02 (0.15)	0.01 (0.01)	0.01 (0.11)	0.01 (0.11)	0.00 (0.08)
Receives Homework	1.00 (0.07)	1.00 (0.04)	0.00 (0.00)	0.93 (0.25)	0.93 (0.25)	0.00 (0.13)
Minutes Spent on Homework	100.8 (55.2)	97.5 (53.2)	3.3 (6.0)	66.1 (46.0)	63.1 (44.2)	3.0 (2.9)
Parents Visit the School	0.57 (0.50)	0.60 (0.49)	-0.03 (0.06)	0.85 (0.35)	0.80 (0.40)	0.05 (0.06)
Expects to Attend University	0.62 (0.49)	0.63 (0.48)	-0.01 (0.06)	0.41 (0.49)	0.50 (0.50)	-0.09 (0.09)
<i>Panel B: Teacher Characteristics</i>						
Trainings This Year	0.93 (0.25)	0.96 (0.19)	-0.03 (0.04)	2.79 (1.37)	2.77 (2.10)	0.03 (0.33)
Contract Teacher	0.31 (0.46)	0.33 (0.47)	-0.02 (0.10)	0.21 (0.41)	0.26 (0.44)	-0.05 (0.09)
<i>Panel C: School Characteristics</i>						
Has a Library	0.70 (0.47)	0.55 (0.51)	0.15 (0.13)	0.50 (0.51)	0.63 (0.49)	-0.13 (0.28)
Has a Playground	0.63 (0.49)	0.52 (0.51)	0.12 (0.13)	0.60 (0.50)	0.57 (0.50)	0.03 (0.13)
Has a Security Guard	0.97 (0.18)	1.00 0.00	-0.03 (0.03)	0.70 (0.47)	0.72 (0.45)	-0.02 (0.05)
Has Latrine Facilities	0.97 (0.19)	0.96 (0.19)	0.00 (0.05)	0.85 (0.37)	0.96 (0.19)	-0.11 (0.18)
Has Running Water	0.97 (0.18)	1.00 0.00	-0.03 (0.03)	0.90 (0.31)	0.94 (0.23)	-0.04 (0.02)

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Columns 1, 2, 4, 5: Standard deviations appear in parenthesis. Columns 3 and 6: Cluster-robust standard errors appear in parenthesis. Panel B Trainings: Columns 1-2 are whether the respondent attended any trainings. Columns 4-5 are the number of trainings attended.

Appendix Table A2: Achievement Effects--Alternative Controls

	eLearn Classrooms				eLearn Tablets
	Standardized Combined Math and Science Test Score				Standardized Combined Math and Science Test Score
	Project	PEC	Combined Project and PEC	Pass the PEC	
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Student Level Controls</i>					
Treatment	0.256* (0.134)	0.223* (0.128)	0.268** (0.119)	0.038 (0.0277)	-0.429*** (0.161)
Observations	2,622	2,766	2,463	2,766	3,058
R-Squared	0.14	0.27	0.21	0.07	0.37
<i>Panel B: Student, Teacher, and School Controls</i>					
Treatment	0.265** (0.125)	0.240** (0.113)	0.274** (0.111)	0.037 (0.0238)	-0.385** (0.175)
Observations	2,622	2,766	2,463	2,766	3,058
R-Squared	0.16	0.30	0.25	0.09	0.42
Average Control Group Change or Mean	0.49	0.00	0.00	0.92	0.45

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Includes all students who took a baseline test and the test at the top of the column. Panel A: Controls are strata, baseline student level test scores, student age, and mothers education. As students take the PEC only once, previous year's school level PEC is included in Columns 2 and 3. Panel B: Controls in Panel A plus school enrollment, facilities, indicator variables for math and science teachers' and head teacher's highest degree. Columns 1 and 5: Project exams. Control group change in the final row. Columns 2-4: Control group mean in the final row. Column 2: Punjab Examination Council high stakes test. Column 3: PCA of project exam and PEC score. Column 4: Linear probability model.

Appendix Table A3: Subject Specific Achievement Effects

	Standardized Test Score				
	Project Exams		PEC Exams		
	Math	Science	Math	Science	All Other Subjects
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: eLearn Classrooms</i>					
Treatment	0.198* (0.104)	0.281* (0.144)	0.181 (0.116)	0.187** (0.0950)	0.066 (0.100)
Observations	2,622	2,622	2,766	2,766	2,766
Average Control Group Change	0.31	0.51			
<i>Panel B: eLearn Tablets</i>					
Treatment	-0.534*** (0.169)	-0.103 (0.109)			
Observations	3,058	3,058			
Average Control Group Change	0.45	0.33			

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Includes all students who took the test at both baseline and endline. Controls include strata, baseline test scores, and those selected by LASSO method. Columns 1 and 2: Project exams. Columns 3-5: Provincially standardized exams. Column 5: Average of PEC scores other than Math and Science.

Appendix Table A4: Lee (2009) Bounds

	Standardized Combined Project Math and Science Test			
	eLearn Classrooms		eLearn Tablets	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)
Treatment	0.281** (0.130)	0.317** (0.132)	-0.426*** (0.150)	-0.425*** (0.150)
Observations	2,551	2,551	3,046	3,046

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Sample size adjusted for attrition following Lee (2009). Columns 1 and 2: Controls as in Table 2, Panel B, Column 1. Columns 3 and 4: Controls as in Table 2, Panel B, Column 5.

Appendix Table A5: Changes in Other Teacher Inputs--eLearn Classrooms

	Minutes Spent per Day Planning Lessons	Holds Private Tutoring Sessions	Number of Regular Classes Taught per Week	Number of Extra Classes per Month to Cover Syllabus	Students Approach Teacher for Help During the School Day
	(1)	(2)	(3)	(4)	(5)
Treatment	9.127* (5.334)	0.0722 (0.0540)	-0.876 (1.478)	0.972 (1.351)	0.158** (0.0665)
Observations	115	115	115	115	115
Control Group Mean	57.9	0.12	33.1	5.2	0.45

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Controls include the baseline value and those determined by LASSO. Includes all teachers surveyed at follow-up. Columns 2 and 5: Linear probability models.

Appendix Table A6: Heterogeneous Achievement Effects

	Standardized Combined Math and Science Test Score			
	eLearn Classrooms			eLearn Tablets
	Project	PEC	Combined Project and PEC	
	(1)	(2)	(3)	(4)
<i>Panel A: By Student Baseline Test Score</i>				
Treatment	0.346** (0.169)	0.241* (0.146)	0.241 (0.172)	-0.555*** (0.215)
Treatment X 2nd Quartile	-0.287** (0.139)	-0.0393 (0.0808)	-0.149 (0.114)	-0.222 (0.161)
Treatment X 3rd Quartile	-0.178 (0.151)	-0.00229 (0.0997)	-0.125 (0.130)	-0.238 (0.257)
Treatment X Top Quartile	-0.0595 (0.291)	0.109 (0.160)	-0.126 (0.208)	-0.123 (0.334)
Observations	2,622	2,766	2,463	3,058
p-values of F-tests of coefficients on Treatment + Treatment X Quartile sum to 0				
Quartile 2	0.70	0.20	0.54	0.00
Quartile 3	0.19	0.04	0.42	0.00
Top Quartile	0.24	0.00	0.46	0.01
<i>Panel B: By School Baseline Test Score</i>				
Treatment	0.406 (0.278)	0.761*** (0.227)	0.657** (0.288)	-0.469** (0.226)
Treatment X 2nd Quartile	-0.384 (0.401)	-1.214*** (0.323)	-0.898*** (0.329)	-0.124 (0.405)
Treatment X 3rd Quartile	0.192 (0.431)	-0.998*** (0.303)	-0.510 (0.324)	1.634*** (0.555)
Treatment X Top Quartile	-0.0654 (0.441)	-0.730** (0.326)	-0.265 (0.482)	0.780 (0.532)
Observations	2,622	2,766	2,463	3,058
p-values of F-tests of coefficients on Treatment + Treatment X Quartile sum to 0				
Quartile 2	0.92	0.02	0.24	0.05
Quartile 3	0.05	0.13	0.45	0.02
Top Quartile	0.34	0.85	0.22	0.48
<i>Panel C: By School Gender</i>				
Treatment	0.328* (0.197)	0.210 (0.197)	0.145 (0.178)	-0.407** (0.184)
Treatment X Female School	-0.0622 (0.247)	0.0934 (0.214)	0.248 (0.212)	-0.163 (0.261)
Observations	2,622	2,766	2,463	3,058
p-value of F-test of coefficients on Treatment + Treatment X Female School sum to 0				
p-value	0.099	0.00	0.01	0.00
Average Control Group Change	0.49			0.45

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Includes all students who took the baseline test and the test indicated at the top of the column. Controls include strata, baseline test scores, and those determined by LASSO.

Appendix Table A7: Achievement Effects by Question Difficulty--eLearn Classrooms

	Easiest Questions	2nd Quartile of Difficulty	3rd Quartile of Difficulty	Most Difficult Questions
	(1)	(2)	(3)	(4)
Treatment	0.238** (0.0932)	0.264** (0.126)	0.081 (0.111)	0.118 (0.0799)
Observations	2,622	2,622	2,622	2,622

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Includes all students who took the test at both baseline and endline. Controls include strata, student baseline test scores, and those selected by the LASSO method. After dividing questions based on IRT reported difficulty parameters, student test scores were calculated based on only the questions indicated at the top of the column.

Appendix Table A8: Achievement Effects by Teacher Experience and Peers - eLearn Classrooms

	Combined Project and PEC Score	
	(1)	(2)
Treatment	0.0716 (0.165)	0.284** (0.132)
Treatment X Inexperienced Teacher	0.312 (0.268)	
Treatment X Grade-level Peer		-0.330 (0.239)
p-value on F tests that coefficients sum to 0	0.09	
Observations	2,463	2,463

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Includes all students who took the test at both baseline and endline. Controls include strata, student baseline test scores, and those selected by the LASSO method. Column 1: Inexperienced teachers are those with less than the median level of experience. Column 2: Grade-level peer is whether another teacher taught the same grade level subject.