NBER WORKING PAPER SERIES

# BEYOND THE BASICS: IMPROVING POST-PRIMARY CONTENT DELIVERY THROUGH CLASSROOM TECHNOLOGY

Sabrin A. Beg
Adrienne M. Lucas
Waqas Halim
Umar Saif

Working Paper 25704
http://www.nber.org/papers/w25704

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2019, Revised June 2019

For useful comments and suggestions we thank Jim Berry, Rebecca Dizon-Ross, Isaac Mbiti, Emily Oster, Maria Rosales-Rueda, Laura Schechter, L. Choon Wang, Matt White, and seminar participants at Innovations for Poverty Action Ghana, Pakistan Planning Commission, the University of Delaware, the University of Minnesota, the Barcelona Graduate School of Economics Summer Forum, and the Midwest International Economic Development Conference. We gratefully acknowledge funding from the Post-Primary Education Initiative of the Jameel Poverty Action Lab (J-PAL). AEA RCT Registry number AEARCTR-0003536. The University of Delaware IRB determined that this project was exempt from IRB review. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Beyond the Basics: Improving Post-Primary Content Delivery through Classroom Technology
Sabrin A. Beg, Adrienne M. Lucas, Waqas Halim, and Umar Saif
NBER Working Paper No. 25704
March 2019, Revised June 2019
JEL No. C93,I21,I25,I28,O15

**ABSTRACT**

Using an RCT in middle schools in Pakistan, we test the effect of a government-implemented in-class technology and brief teacher training intervention on student achievement in grade level mathematics and science. After only 4 months of exposure, student test scores increased by 0.2-0.3 standard deviations on math and science tests, 52 to 120 percent more than the control group, and students were more likely to pass the provincially standardized high-stakes exams. Increased effort by both students and teachers indicate a complementarity between technology and other inputs in education production. At a scale of 100 schools, this program is extremely cost-effective.

Sabrin A. Beg
University of Delaware
418 Purnell Hall
Newark, DE 19716
sabrin.beg@gmail.com

Adrienne M. Lucas
Lerner College of Business and Economics
University of Delaware
419 Purnell Hall
Newark, DE 19716
and NBER
alucas@udel.edu

Waqas Halim
IT University Lahore
waqashalim@gmail.com

Umar Saif
IT University Lahore
saif.umar@gmail.com

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/3536

# 1 Introduction

In Pakistan, as in much of the rest of the world, access to schooling has increased substantially since 2000 (World Bank 2017). The influx of new learners has strained existing school resources at the primary level and now post-primary level as these graduates progress to upper level education. As a result, in many countries, as in Pakistan, learning outcomes are inadequate even for those who are enrolled in school (Andrabi et al. 2007, Muralidharan 2013). In many low income countries, these learning deficiencies start in primary school and are exacerbated at the middle and secondary school levels as students progress to more advanced material, especially in technical subjects such as science and math (Banerjee et al. 2013). Despite the importance of subject-specific education for future labor market outcomes and poverty reduction, scant research exists on ways to increase student learning in developing countries beyond foundational literacy and numeracy (Banerjee et al. 2013). One challenge in many developing countries is teacher capacity–teachers themselves need to be knowledgeable and trained in the subject matter. As the subject matter becomes more complicated in higher levels of schooling, the demands for teacher capacity are also greater. Specifically in our middle-school context in Pakistan, 51 percent of school principals reported that insufficient teacher qualifications were a barrier to student learning. In developed countries, the growing field of education technology (ed-tech) is lauded for its potential to deliver high quality education given existing teacher capacity (Escueta et al. 2017). Our intervention leverages technology through a government implemented, scalable, relatively inexpensive program to address learning deficiencies and potentially circumvent limitations on teacher knowledge. Specifically, we use a randomized controlled trial (RCT) to test the impact of eLearn, a program that delivers expert math and science content through short videos to grade 8 students in Punjab, Pakistan. Since this change in available inputs could be either a complement to or substitute for other forms of effort, in addition to the reduced form effect on test scores, we test the effect of the program on additional intermediate outcomes.

In treatment schools the eLearn intervention consisted of short, multimedia video presentations that corresponded to concepts in the official science and math content from the 8th grade curriculum, a few multiple choice review questions after each lecture, a small tablet for teachers to use to serve this material to a larger screen and review it on their own, and an LED screen installed in each classroom to display this content to the class. The tablets could also be used to record student grades and attendance. Teachers were trained in how to use the multimedia content and incorporate it into a more effective teaching practice during a two day in-service training. Outside of the teacher training, all aspects of the intervention occurred during the school day.[1] Overall, this intervention contained 29 total hours of content to be spread over the entire school year. The Punjab Information Technology Board and Information Technology University Lahore, two government entities, created and implemented the program. All trainers and content providers were government employees. The intervention was designed to complement existing practices and teachers, not add additional employees or act as a substitute for existing personnel.

The intervention increased student achievement by 0.26 standard deviations in math, 0.27 standard deviations in science, and 0.32 standard deviations in a combined score despite students receiving only 4 months of exposure. Unlike many studies, we also test whether the intervention increased scores on a provincially standardized, high stakes test. Our score increase is not an artifact of our content being particularly relevant to our test. Students also scored 0.28 standard deviations higher on the combined math and science sections of the standardized test that students take at the end of grade 8. We find no statistically significant changes to the other subject scores on the provincial standardized test, suggesting that our intervention did not substantially augment or distract from other subjects. Further, the intervention increased the likelihood that students passed the standardized grade 8 test by 7 percentage points. Passing this examination determines what options are available for further study and acts as a proxy for longer run outcomes.

[1] The intervention was designed to have an additional at home component that would provide students with interactive SMS review questions on their households' mobile phones. Unfortunately, this component was barely implemented during the period under study. The estimations include this component, but it is likely at most a marginal contributor to the overall effect. See more details in Section 3.

To better understand the mechanisms behind these improvements change, we test for effects on both observed and self-reported effort. Both student and teacher attendance increased.[2] We find at most minimal increases in self-reported effort. In contrast with other interventions that targeted grade level material and found larger, or the only gains, among students with high scores at baseline (e.g. Glewwe et al. 2004; Glewwe, Kremer, and Moulin 2009), students across the baseline test score distribution gained equally. Therefore, grade level content implemented at the middle school level does not necessarily exacerbate pre-existing achievement heterogeneity.[3] When we test for heterogeneous effects by gender, we find some evidence that girls' scores increased more than boys on the provincially standardized exams.[4]

This work fits into three related literatures: improving education beyond foundational literacy and numeracy, the importance of schooling inputs in the education production function, and educational technology to improve student learning.

In the literature on methods to improve achievement at the post-primary level or beyond foundational literacy and numeracy, effective and inexpensive interventions are relatively unknown despite their rising importance as more children worldwide complete primary school. Existing post-primary research has compared secondary schools that varied on many dimensions of quality (e.g. Jackson 2010, Pop-Eleches and Urquiola 2013, and Lucas and Mbiti 2014) or focused on attributes of the school day (Bellei 2009). Muralidharan, Singh, and Ganimian (2019) found positive achievement effects for middle school students from an after school program that combined computer assisted learning software with personal tutoring. Despite their focus on students in middle school grades, the intervention was not primarily about middle school content since most of the study pupils were well behind grade level.

The literature on specific school-based inputs that improve educational outcomes mostly focuses on foundational literacy and numeracy skills in primary school. Research on primary

[2] Our test score findings are robust to attrition correction. See Section 7 for further details.

[3] As our students already completed seven years of school, the average student may be closer to grade level than the primary school age students in the other studies.

[4] As is common in middle schools in Punjab, all of our study schools are single gender, therefore these differences may be due to differences between schools and not student gender. See more details below.

schools has shown the importance, and effectiveness, of such interventions as instructional materials, teacher training, grouping students by ability, and providing additional contract teachers (e.g. Banerjee et al. 2007, Duflo, Dupas, and Kremer 2011 and 2015, Lucas et al. 2014, and Andrabi et al. 2015. See McEwan 2015 for a further summary). In contrast, other papers have improved school infrastructure or provided teaching materials without corresponding achievement effects (e.g Newman et al. 2002, Glewwe et al. 2004, Glewwe, Kremer, and Moulin 2009). The applicability of these methods to higher grades that require more specialized content knowledge is largely unknown.

Specific to leveraging technology in primary schools, the most common intervention in developing countries has been computers and software, a different version of education technology than the current study. The evidence of this model of technology introduction is mixed and often related to the amount of training received by teachers, whether the technology is integrated into the curriculum, or the existing knowledge of the students.[5] In developed countries, almost all computer assisted learning studies have found positive effects, especially in mathematics (Escueta et al. 2017). Other studies have provided classroom lessons to rural primary schools or preschools through satellites or audio CDs (e.g. Johnston and Ksoll 2017 and Naslund-Hadley, Parker, and Hernandez-Agramonte 2014), finding positive effects.[6]

Our research contributes to the understanding of improving achievement in six important ways. First, we leverage technology during the school day to improve math and science achievement in middle school. Second, the intervention is lighter touch than many programs that required multiple hours per week of student time either during or after school. Third, our intervention is substantially cheaper to scale than other interventions with similar effect

[5] For example, outside the school day, computer assisted learning (CAL) improved primary school student test scores in China (Mo et al. 2014; Lai et al. 2013; Lai et al. 2015; and Lai et al. 2016), Ecuador (Carrillo et al. 2010), and India (Banerjee et al. 2007; Linden 2008). The evidence during the school day is more mixed, with some positive effects (Banerjee et al. 2007 in India), some negative effects (Linden 2008 in India), and some heterogeneous effects by student baseline ability (He, Linden, and MacLeod 2008 in India) or integration with the broader curriculum (Bai et al. 2016 in China). When considering the hardware alone, Barerra-Osorio and Linden (2009) found that the installation of computers in Colombian public schools did not lead to increases in learning because they were not integrated into the learning process. Our program integrated technology into the existing pedagogy and curriculum.

[6] The at most marginally implemented SMS component builds on Aker et al. (2013) that found that mobile technology was a complement to rather than a substitute for highly educated teachers.

sizes. The marginal cost per school was $9/student per year. Even with the inclusion of the substantial content fixed costs, the cost would be $15/student at the scale of 100 schools.[7] Fourth, our effect sizes are especially large for only 4 months of exposure. Fifth, our intervention was designed and implemented with the provincial government of Pakistan, not a non-governmental organization ensuring a program directly salient to issues the government found pressing and increasing the possibility of scale-up of the program. Sixth, we find significant positive effects on independent tests conducted by the government, indicating that our intervention not only assists student learning, it potentially benefits real, longer term student outcomes that may depend on performance on government tests.

## 2 Background on Schooling in Pakistan

The Pakistani school year begins in April, consists of a summer break from June to mid-August, and ends in March of the following year. In Pakistan, primary school, i.e. junior school, consists of grades 1 through 5. Middle school follows with grade 6 through 8. All of our study schools are single gender, as is typical of government middle schools in Pakistan.

At the conclusion of middle school, students take the provincial standardized exams. A student's score on this test signals completion of middle school and is required for admission to government secondary school. In Punjab the standardized exam is the Punjab Examination Commission (PEC) exam that covers 5 subjects: English, Islamic Studies (or Ethics for non-Muslim students), Mathematics, Science, and Urdu. The Islamic Studies, Mathematics, and Science portions of the test are available in both English and Urdu. Instruction at the middle school level occurs in a blend of English and Urdu. Secondary school and higher secondary school are grades 9-10 and grades 11-12, respectively. Government schools at all levels charge at most minimal tuition fees.

Student achievement in government schools is quite low, in part because of available

[7] These costs include the development of some content that was at most marginally implemented during our period of study and some content that was not implemented at all. See the Cost Effectiveness section for more details.

resources. In Punjab, our region of study, Andrabi et al. (2007) found that government schools lacked basic facilities and teaching resources. Similarly, Andrabi et al. (2013) found that unavailability of qualified teachers in Pakistan constrained educational provision. In our baseline data collection, 51 percent of school principals cited lack of teacher qualification as a constraint on student learning. Despite challenges faced in the education sector, many dedicated individuals are working in the sector under difficult circumstances, and this project focuses on supply side interventions that maximize and augment available inputs.

In most developing countries, including Pakistan, the "technology of instruction" typically consists of teachers standing at the front of the room lecturing to a classroom of students (Glewwe and Muralidharan 2016). From our experience in Punjab, this method seems to be ineffective in large classrooms if the teachers are not trained in the subject themselves, or if teachers lack continuous assessment and feedback to ensure that all students adequately follow the subject material. Additional instructional tools and training teachers on incorporating them into lessons could have a substantial impact.

## 3 Intervention

Our intervention, eLearn was designed to improve student learning by combining technology with existing teachers. This small-scale implementation and evaluation of the program was designed to inform the larger scale-up of the program that started in 2018. 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.

To increase student learning, eLearn increased the availability of high quality subject specific content through technology and a two day teacher in-service training.

The main component of the intervention was video lectures. Each video lecture was developed and presented by subject experts to explain a particular math or science concept. All videos directly mapped to the units of the official curriculum.[8] A single presenter appeared in all videos related to a particular unit. All presenters were government employees. Of the 22 units of content, men were the subject experts for 21 units.[9] The total content across all videos was about 29 hours. These lectures contained spoken Urdu with an occasional English word and all words written in English, as is typical in Pakistani middle schools where textbooks are often in English and instruction occurs in a mix of English and Urdu. Paired with some videos were an additional 3 to 5 minutes of multimedia content that the teacher could play to reinforce the content of the videos, e.g. an interactive animation of photosynthesis.

To view and display these video lectures and multimedia content, teachers were given small, pre-loaded tablets, and classrooms received LED television screens. Teachers could use these tablets to watch the videos themselves when preparing for lectures and project them on the installed screens. The 40 inch LED televisions were installed above the existing chalk or white board enabling teachers to continue to use the board in an interactive way with the videos. The tablets further contained 3 to 5 multiple choice assessment questions and their answers that teachers could use to engage the class after each video and suggestions for further in-class activities relevant to each topic. This classroom technology was designed to augment and complement the teachers' existing teaching techniques.[10]

Teachers received a two day in-service training session primarily focused on program implementation–orientation on the new technologies and how to combine their own face-to-face teaching with technology-enabled multimedia content. All treatment teachers in a district attended the same training regardless of their gender or the gender of their students.

An additional component of the intervention was designed to engage students and parents

[8]The math content was a total of 12 hours and 21 minutes broken into 77 videos of between 1.5 and 21 minutes in length with an average of 10 minutes per video. Each of the 10 units of the official curriculum had at least 35 minutes of content and 4 videos. For science, the total video length was 16.5 hours, spread across 115 videos of between 3 and 27 minutes in length with an average of 8 minutes. Each of the 12 units of the official science curriculum had at least 23 minutes of content and 2 videos.

[9]The only female presenter appeared in the environment unit of the science curriculum.

[10]This technology was likely novel to some but not all 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.

at home, but was at most marginally implemented. Students were to be engaged through an at home SMS-based Intelligent Tutoring System (ITS). As designed, at the conclusion of each of 22 chapters of the curriculum, students should have received an SMS blast on their household's mobile phone that the ITS system was available for their use.[11] This ITS system would have allowed students to use text messages to receive and respond to review questions. Unfortunately, due to delays this part of the intended intervention was barely implemented during our period under study. Project records indicate that 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 modules. Finally, the mobile carrier used for broadcasting 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 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 substantially less intense than intended.

Our intervention took place during the 2016-2017 school year. Teachers were surveyed and students were surveyed and given the baseline exams in August after the end of the June to mid-August 2016 school holidays. The teacher trainings and hardware installation were finished by the start of October. Our follow-up surveys and exams occurred in January 2017. The PEC standardized exams occurred in February 2017. Therefore, students and teachers were exposed to the intervention for at most 4 months between the baseline and follow-up testing. Figure 1 displays the study and academic year timeline. See Section 5.2 for additional data details.

[Figure 1 about here]

[11] Almost all study students reported having at least one mobile phone in their household and 80 percent reported having two or more in their household.

# 4 Empirical Strategy

The primary conceptual difficulty in assessing the effects of various inputs into the education production function are the non-random allocation of resources and their typical correlation with household and school attributes, leading to biased estimates. To alleviate this concern, we designed a randomized controlled trial of our intervention.

We randomly divided our study sample schools into treatment, i.e. eLearn schools, and control, i.e. "business as usual" schools.

From this randomization design, we compare outcomes between the treatment and control schools after the intervention. Formally we estimate

$$y_{is} = \alpha + \beta treatment_s + X_{is}^{'}\Gamma + \varepsilon_{is} \tag{1}$$

where $y_{is}$ is outcome $y$ for student $i$ in school $s$, $treatment_s$ is an indicator variable equal to one if the school was a treatment school, $X_{is}$ are a vector of school and individual level controls, and $\varepsilon_{is}$ is a cluster-robust error term assumed to be uncorrelated between schools but allowed to be correlated within a school. In all specifications in which the outcome of interest is a test score, we implement a lagged dependent variable model and include the test score from the baseline as a control in the $X_{is}$ vector. Additional controls in the $X_{is}$ vector are strata (school gender by district) dummy variables. In some specifications because of slight baseline imbalance and to improve precision we include the following school level controls: total baseline enrollment, total baseline attendance, number of grade 8 sections, dummy variables for the presence of school facilities (i.e. library, computer lab, and playground), experience and qualifications of the principal and math and science teachers, and contract status of the math and science teachers.[12]

Our first primary outcomes of interest are student test scores. A distinguishing feature of our study and research design is that we have two types of tests. We first test for the impact

[12] In Section 7 we use the LASSO approach to select co-variates. Our results are robust to this machine learning technique of co-variate selection.

on exams designed specifically for this project. These tests follow the established curriculum and test higher order conceptual and problem solving abilities than the official provincial tests that rely heavily on memorization. We are also able to link our study students to their official PEC exam scores and test the impact of the intervention on these scores as well.[13] We further test for heterogeneous effects by baseline test score and gender. When a student's test score is the dependent variable in Equation 1 the reduced form effect on achievement includes any changes to students' or teachers' effort and other inputs.

Our additional provision of technological inputs and teacher training could have been a substitute for other inputs, e.g. students spend less time studying in reaction to additional material being delivered at school, or complements, encouraging additional provisions of inputs, e.g. teachers could spend more time teaching and using the new technology.

To test for potential mechanisms, we further estimate the effect of the intervention on whether the student was present the day of the follow-up, reported using technology to study at home, time spent on homework, self-reported absenteeism in the last week, whether the student received private tutoring sessions, whether parents visited the school to meet with school faculty or staff, and whether parents expect a student to attend college.[14]

Additionally, based on data collected from teachers, we estimate a similar model, allowing $i$ to index the teacher instead of the student. The outcomes of interest for teachers are whether they used technology to prepare for classes, used technology to teach their classes, had been part of any training, held private tutoring sessions outside of school, performed other official duties, and were approached by students for help outside of class time. We also estimate the effect of the program on the teacher's average number of classes taught, how many hours they spent preparing for class, and how many extra classes they taught in a month during school hours to cover grade 8 syllabus. Finally, we use administrative data on teacher attendance collected by independent monitors to test for any effects of the program

[13] As students take the PEC exam only once, in the PEC exam specifications we include the school level previous year average PEC score and the students own baseline project-specific exam scores as the lagged dependent variable analog.

[14] A student being present is also our measure of attrition. Our findings are robust to attrition correction.

on objectively observed effort.

# 5 Sample Selection and Data

## 5.1 Sample Selection

Our study takes place within Lahore, Multan, and Rawalpindi districts of Punjab Province, Pakistan, the most populous province in Pakistan, home to over half of Pakistan's 208 million residents.[15] These districts contain 20 percent of the total population in the province. To be eligible for our study, schools had to appear in the Punjab School Census, include grades 1 through 10, and have a boundary wall, electricity, and physical classrooms—basic amenities in the Punjab context. These attributes were all necessary to securely install and power the LED screens. As is typical in Punjab, all schools were single gender in middle school. From eligible schools, we selected 60 schools, an equal number of boys' and girls' schools, for the sample. Randomization was stratified by district and gender. One control school dropped out by the endline stage, leaving us with 29 control schools and 30 treatment schools.

Overall, our sample schools are similar to the average school in Punjab based on infrastructure and test scores. First, while the conditions of a boundary wall and electricity might be binding or indicate particularly wealthy schools in other contexts, in Punjab 93 percent of schools have electricity and 97 percent have a boundary wall. Second, the average PEC score for our control schools was 53, the same as the provincial average for 2016.

## 5.2 Data

We use two sources of data: primary data collection and administrative data.

Our primary data were hand collected at each of the study schools. The baseline data collection occurred in late August 2016, two instructional months into the 2016-2017 academic year, after the June to mid-August (summer) holiday, but prior to the teacher training

[15]The study was limited to three districts to decrease the costs associated with on-site technology support of the screens and tablets. Two of these districts are in the north and one in the south of Punjab.

or availability of the new technology. The baseline surveys solicited information from head teachers, grade 8 math and science teachers, and randomly selected students in grade 8 present on the day of the baseline. All present students in grade 8 took mathematics and science tests. In the baseline we tested 2,999 students and conducted 1,690 student interviews across 59 schools. The installation of the required equipment and teacher trainings occurred in late September and was the technology was operational in all treatment schools by the beginning of October.[16] For the duration of the school year, treatment schools were visited by the implementing partner's technology support team to ensure equipment was secure and functioning as intended.[17] We administered follow-up surveys and exams in January 2017, near the end of the academic year. Enumerators told schools that we would be visiting them near the end of the school year, but they did not provide an exact date. The same students were again surveyed and tested, if present. Head teachers and grade 8 subject teachers were again surveyed.

When estimating the effects on test scores, we use item response theory (IRT) to convert raw test responses to approximated latent student ability, and standardize based on the baseline mean and standard deviation.[18] Our findings are similar using raw test scores.

The final point of student data are administrative student by subject level exam results from the Punjab Examination Commission. Students completed the PEC exams in mid-February. These data were merged to the students in our sample using students' and fathers' names.[19] Because we do not have item level responses, these scores are simply scaled with a mean 0 and standard deviation of 1.[20]

[16]The surveys and program implementation were originally designed to occur prior to the summer holiday but implementation funding was delayed.

[17]These teams were not designed nor equipped to support or improve teaching practices.

[18]We use a one parameter IRT logistic model.

[19]We match 93 percent of baseline students to their PEC record. Our match quality is not differential by treatment status or treatment status times baseline test score. Appendix Table A1 provides the regression results with a dummy variable for being matched as the dependent variable. The unmatched 7 percent includes both students who registered for the PEC but we were unable to match and those who did not register for the exam or changed schools. Of those not matched, and therefore more likely to have changed schools or dropped out, about a quarter of them were not present at our follow-up survey. When considering only those present at our follow-up, our match rate is 95 percent.

[20]The exact questions on PEC exams can vary across districts but not within them (Barrera-Osorio and Ganimian 2016). Our strata (i.e. district by gender) fixed effects will control for any district level differences

The two sets of exams—project-specific and administrative—were both designed to cover material from the same curriculum. Our project-specific exams were designed by subject experts, not particularly involved with the design and implementation of the program, to be more conceptual and less prone to rote memorization, an criticism of the PEC exam and other similar provincial exams in Pakistan (Government of Punjab 2013; Burdett 2017). No study teachers had access to the test and students were not allowed to keep any testing materials.

Administrative data on teacher attendance are from the Punjab Monitoring and Implementation Unit (PMIU) school checks, which are publicly available on the PMIU website. Monitoring and Evaluation Assistants conduct monthly, unannounced school visits and record teacher attendance.[21]

These data were available at the school level only. Therefore, they measure the percentage of teachers present during the visit.

Tables 1 and 2 display means and standard deviations of 44 different student (Table 1, Panel A), teacher (Table 1, Panel B), and school (Table 2) characteristics across the treatment and control schools. Almost all of the measures are statistically indistinguishable by treatment status with four exceptions: treatment students report being absent more often in the previous week by 0.3 days, the math teacher in treatment schools is 20 percentage points more likely to have exactly a college degree, teachers in treatment schools teach 3 more classes per month, and treatment schools are 10 percentage points less likely to have a computer lab (considering a sample size of 30 on each side, this reflects three treatment schools not having a computer lab). Given that we are testing 46 different outcomes, some small differences are expected. Nevertheless, to ensure we are not attributing baseline imbalance to the treatment effect, we provide estimates with school level controls.

between test scores.

[21]Even though they are government employees, these monitors were not affiliated with our program or the Punjab IT Board, the primary government implementing partner. They were not explicitly made aware of the program nor which schools were treatment or control. They might have observed LED screens in some grade 8 classrooms. We cannot reject that this might have influenced their overall assessment of teacher attendance in a school, but believe it to be unlikely.

[Tables 1 and 2 about here]

# 6 Results

We first test for the effects of the program on students' test scores for both the project-specific and PEC exams. Then, we explore possible mechanisms behind the achievement results including student attendance, an interesting outcome itself as well as our measure of attrition.[22]

## 6.1 Achievement

To estimate the effect of the program on achievement we estimate Equation 1 with a student's endline test score as the outcome of interest and include baseline test score as a control variable. The results of this estimation appear in Table 3. Columns 1 and 2 contain the subject specific math and science scores, and Column 3 combines these two scores into a single score. Panel A includes only the strata and baseline test scores as additional control variables. The treatment increased math achievement by 0.19 standard deviations (column 1), science achievement by 0.24 standard deviations (column 2), and the combined score by 0.26 standard deviations (column 3). Given our small sample size of 59 schools and slight imbalances from Tables 1 and 2, we include additional school and teacher control variables in Panel B to assist both with precision and ensure we are not attributing underlying differences between the groups to a treatment effect. These results are somewhat larger with stronger statistical significance. The treatment increased math achievement by 0.26 standard deviations (column 1), had a similarly sized effects on science achievement (column 2), and increased the overall math plus science achievement by 0.33 standard deviations (column 3), sizable effects for a 4 month treatment.[23] During this same period control

[22]We were unable to reach 13 percent of our baseline sample during our endline. In Section 6.2 we test for differential attrition by treatment status and provide Lee (2009) bounds in our Section 7.

[23]The increase in the point value between Panels A and B is mostly driven by the inclusion of the teacher qualifications and experience: teachers in treatment schools were more likely to have a college degree and

group students increased their test scores by 0.2 standard deviations in math, 0.5 standard deviations in science, and 0.4 standard deviations overall.[24] Therefore, this intervention more than doubled the achievement in math, increased achievement in science by 52 percent, and in the combined score by 74 percent relative to the gains in the control group.

[Table 3 about here]

Our exams were designed to test the content from the official curriculum, specifically including questions that required higher-order thinking and problem solving. Nevertheless, to alleviate concerns that the content of the tests was particularly well aligned to the intervention, leaving control students at an artificial disadvantage, in Table 4 we test the effect of treatment on the standardized PEC tests that all grade 8 students wishing to enter government secondary schools take.[25] As with the previous table, Panel A includes limited controls and Panel B includes additional baseline controls. Columns 1, 2, and 3 repeat the specification from Table 3 with the PEC math, science, and combined scores as the dependent variables. We do not find statistically significant effects on the limited controls version (Panel A). Once we include the additional controls in Panel B, we find that the intervention had about a 0.33 standard deviation effect on the science score and 0.28 standard deviation effect on the combined math and science test score. The effect is not statistically significant for the math score alone.

[Table 4 about here]

While not designed to change student achievement in other subjects, better math and science instruction could have freed student time to focus on other subjects or alternatively it could have caused students (or schools) to spend more time on the subjects with the

not a masters. This qualification is associated with lower student test scores. Not controlling for these qualifications introduced bias in Panel A.

[24]The control group increase in scores is the follow-up mean minus the baseline mean for the students who appeared in both rounds.

[25]Students who registered for the PEC but did not sit for the exam receive a score of 0 in the official score report.

new, exciting teaching methodology, reducing time on other subjects. Column 4 of Table 4 provides estimates of the effect of the treatment on the combined score of the subject tests that were not the focus of the intervention, a net effect of any countervailing forces. The point estimate is positive and statistically insignificant. Therefore, we do not find evidence that the intervention led to a change in non-intervention scores. Column 5 contains the estimate of the treatment on the overall average score. Despite increases in the combined science and math score, we do not find a statistically significant effect on the overall average of all 5 of the subject tests.

One goal of the program was to prepare students for future study. Passing the PEC exam is one measure of this readiness. In Column 6 we estimate the effect of the intervention on the likelihood that a student passed the PEC exam, and find that the intervention increased this likelihood by 7 percentage points (statistically significant at the 10 percent significance level).

## 6.2 Attrition and Attendance

The achievement results in the Section 6.1 were the overall treatment effect on student achievement. To understand potential mechanisms and whether students and teachers substituted the new content for other inputs into the educational production function, we re-estimate Equation 1, replacing the dependent variable each time with another input into the education production function.

As a first measure of observable effort, we separately estimate whether students who were present in the baseline were similarly present at the endline.[26] While an interesting outcome itself, it is also a measure of respondent attrition. We use an indicator variable equal to 1 if the student was present at follow-up as the outcome, $y_{is}$, in Equation 1. As with all our other binary outcomes, we estimate the result with a linear probability model. The results appear in Columns 1 and 2 of Table 5.

[26]We cannot use their presence in PEC data to measure grade 8 completion as that conflates our ability to match them and their completion.

[Table 5 about here]

Students in the treatment group were about 5 percentage points more likely to be present at follow-up (column 1). Relative to the control group mean of 85%, this is about a 6% increase in the likelihood of being present. While encouraging that our intervention increased attendance, one concern is that this differential attrition could be biasing our other outcomes of interest by inducing selection into the test. In column 2 we test if this differential attrition by treatment status is related to a student's baseline test score by including an interaction of treatment status times a student's standardized baseline test score as an additional regressor. We do not find evidence of differential attrition by baseline ability and treatment status with a statistically insignificant, small point value. Nevertheless, in our Section 7 we follow Lee (2009) and provide treatment bounds. Our test score findings are robust to this attrition adjustment.

Consistent with this improvement in attendance, teachers in treatment schools were 18 percentage points less likely to view student absenteeism as a significant problem in teaching (column 3). In the control group, 60 percent of teachers viewed this as a significant problem.

We also test for the treatment effect on teacher presence and effort. As an objective measure of teacher effort we rely on Punjab Monitoring and Implementation Unit (PMIU) administrative data that records teacher attendance at the school level from a monthly unannounced visit. In columns 4 and 5 we estimate the effect of the treatment on the overall portion of teachers present during these monthly unannounced school visits. In this specification, we include each monitoring visit as a separate observation, controlling for the portion of teachers present exactly one year prior, a model similar to Equation 1 but with multiple observations per school. Overall teacher attendance is high: approximately 94 percent. Our intervention increased the portion of teachers present in the school by almost 1 percentage point (column 4). Since we do not have PMIU data at the individual level, this is the effect on teacher attendance for the whole school. As this is a monthly measure, we can test the evolution over time in teacher attendance. In column 5 we test whether this response changes over time, and while the point estimate on the interaction between treatment and months of treatment is insignificant, it is negative. Therefore, the intervention appears to have increased teacher effort, but this effort might have diminished over time.

To measure whether teachers used the technology, the tablets recorded data on time of use and number of items used each month. The data collected by the tablets report that all schools used the technology, but some used it more than others. On average schools accessed 74 of 192 videos (39 percent), 11 of 50 simulations (22 percent), and 152 of 600 questions (25 percent). Almost all of this access occurred during school hours—81 percent of videos, 70 percent of simulations, and 90 percent of questions. Figure 2 displays the average monthly usage statistics for the videos, questions, and simulations. Across all three items, use peaked in November–the first full month of the intervention–and during school use (solid blue) exceeded use outside of school hours (red dashed) for almost all months and items. From the tablets we cannot know whether the content was displayed to the students. The students were asked how frequently their teachers displayed the content. These responses are highly correlated with the data from the tablets. Therefore, while this content might have increased the teachers' own subject knowledge, most of its use was targeted directly at students. The early peak in use in November and later decline is consistent with the evidence from teacher attendance that attendance initially increased, then reverted back to its pre-intervention levels. Even at its lowest point in February the average school was still accessing some content. Recall that students took the PEC exam in mid-February, therefore teaching time was both interrupted and structured differently in that month.[27] Implementors conducted two spot check visits to each school during the intervention. During these visits, 83 percent of schools were using at least one piece of technology. Based on a 27 question check-list on implementation, all schools received scores of at least 14 and 82 percent received scores of at least 22. Therefore, technology use in the classroom appears to have been an

[27]Teachers were encouraged to use all of the content but the actual use was left to their discretion. From the experimental design, we cannot know whether the ideal amount of use is closer to the November peak or the December and January levels. Further, the November peak could be due to learning how to navigate the software and selecting videos in error. Nevertheless, the long run effect of the program might not be the short run effect scaled for additional duration of exposure.

important part of the intervention.

[Figure 2 about here]

## 6.3 Self-Reported Outcomes

We additionally collected data on self-reported changes in take-up and inputs. We first test for whether the change in inputs increased the use of technology available to teachers. Table 6 contains these results. Teachers were 42 percentage points more likely to report that they used technology to prepare for lessons (column 1) and 77 percentage points more likely to report they used technology in the classroom (column 2). From survey responses, 95 percent of teachers reported using the screen and tablet at least twice a week and 70 percent of teachers and 80 percent of the students found the technology "very useful."

According to data collected during the training, all treatment teachers attended the training. During the school based survey administration, treatment teachers reported having attending more trainings. Column 3 shows that treatment teachers attended 0.44 more in-serivce teacher training events during the school year.

[Table 6 about here]

We tested for additional changes in teacher effort that might have occurred as a result of the intervention (see Appendix Table A2). At the ten percent significance level, treatment teachers reported spending 12 more minutes per day planning lessons and were 14 percentage points more likely to report holding private tutoring sessions.We do not find any statistically significant effects on the number of regularly scheduled classes taught per week, the number of extra classes per month during the school day, that the students approached them outside of class during the school day for extra help in math or science, or the likelihood of being required to perform additional duties in addition to teaching. Therefore, teachers in treatment schools increased their use of technology and their observed effort (attendance), but at most marginally changed their self-reported effort.

We further tested for changes in students' self reported effort. Our intervention did not change the likelihood that students used technology at home to study, the minutes per day spent studying, the self-reported number of days absent in the last week, whether they received out of school tutoring, whether their parents visited the school to meet with the teacher, or whether they expected to attend university (see Appendix Table A3).

Given the at most marginal changes in effort other than attendance, the effect on test scores is likely the direct effect of the intervention and not working solely through changes in effort or other inputs.

# 7 Heterogeneity and Robustness

Because the intervention videos were at the level of the curriculum and some of the students could have been behind grade level, the intervention could have differential effects by baseline test score. Panel A of Table 7 tests for this possibility by including an interaction between baseline test score and treatment as an additional regressor.[28] For the project-specific exams, all coefficient values on the interaction effects are small, negative, and statistically insignificant (columns 1 through 3). For the PEC exams, the coefficients are all positive, but also statistically insignificant (columns 4 through 7). Therefore, the project appeared to help all learners equally, regardless of their baseline learning levels.

[Table 7 about here]

Panel B of Table 7 tests for heterogeneity by school gender, replacing the interaction with one for treatment times female school. Recall that all schools are single gender, therefore, differential effects by school gender are testing the combined effect of the program on a student based on her gender as well as any differential effect of attending an all female school. For both the math (column 1) and combined score (column 3) the main effect is positive and statistically significant. For science (column 2) the coefficient is large, but no

[28]Since students only sit for the PEC exam once, we use the baseline project score in the interaction with treatment in those specifications as well.

longer statistically significant. In all cases the coefficient on the interaction term is negative. Therefore, we test whether the sum of coefficients on the main effect and the interaction effect is statistically different from 0. In all three columns, we fail to reject that the program had a 0 test score effect on female students. In contrast, in columns 4 through 7 for the PEC exam, we reject in all cases except math (p-value=0.12) that the sum of the coefficients for females is 0. The main effect is statistically insignificant. Therefore, the program improved PEC scores for females, and we fail to find a statistically significant improvement for males on PEC exams. Overall, the treatment improved female test scores on the PEC and male test scores on the project-specific tests.[29]

The intervention was not designed to favor students of a particular gender–instead providing expert content to assist all students. We are unable to explain exactly why female students appear to benefit on the PEC exams while male students do not. At the school level, female and male schools and teachers are statistically indistinguishable except female schools have a higher percentage of female teachers and higher average baseline test scores.[30]

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 (0.3 years), richer (households have 0.06 more cars), and less likely to work (4 percentage points). Some of these differences are likely due to selection. Nationwide, girls are 13 percentage points less likely to complete primary school and only 38 percent of students in grade 8 were girls in 2016 (Government of Pakistan 2016). Therefore, female students who make it to grade 8 in Pakistan are a more highly selected sample than male students.

We find minimal differences by school gender in the effort and implementation measures that we tested in the previous section, with statistically significant differences by treatment

[29] When comparing the performance of students by gender on low-stakes and high-stakes exams in China, Cai et al. (2019) found that female students underperformed on the high-stakes exams. In our setting, female students gain more from the intervention on the high-stakes PEC exam.

[30] Girls also score higher than boys in the PEC (0.39SD). Therefore, two panels Table 7 could be picking up the same pattern–girls have higher baseline test scores, therefore baseline score and the female school indicator variable are just proxies for each other. To test this, we re-estimate Panel A separately by gender. We find similar patterns to Panel A within gender (results not presented). Therefore, the test score and gender differences appear to be separate phenomenon.

status and gender only for the likelihood that parents have visited the school and university aspirations. Therefore, the differences 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 we find between genders are unlikely driving the gender heterogeneity in achievement, they are of note. The treatment increased the likelihood that male students reported that their parents have visited school by 23 percentage points, while the treatment effect is statistically insignificant for girls. Prior to the intervention this outcome was 4 percentage points higher for male versus female students (0.61 male vs. 0.57 female). 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 (0.50 male vs. 0.72 female). 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 male. 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.

In Table 8 we vary the specification and the sample, finding results that are similar to our preferred specifications in Section 6. Each Panel reports the coefficient of interest from a separate regression with the dependent variable as indicated in the Panel heading. Column 1 repeats our estimates from Table 2. Column 2 implements the Belloni, Chernozhukov, and Hansen (2014) post double Least Absolute Shrinkage and Selection Operator (LASSO) approach to specify the optimal controls to include along with the baseline test scores. In all cases the point estimates are similar to our preferred specification. Column 3 limits the sample to those students for whom we collected surveys. The math and combined effects are similar, but a larger point value for math and a smaller one for the combined score.

The science effect is smaller and no longer statistically significant. Column 4 uses the survey sample with additional individual level controls. The results are similar to column 3. Finally in columns 5 and 6 we use Lee (2009) to adjust the attrition and find narrow bounds around our preferred specification.

[Table 8 about here]

## 8 Cost Effectiveness

One reason why technology is potentially promising in low resource settings is its ability to deliver content relatively cheaply. The marginal costs, i.e. excluding the development of content, of our intervention are quite cheap. Because this intervention is at the classroom and not student level, adding an additional student to the classroom is costless, understanding that at some point a class would become too large for this method of instruction to be effective. The average classroom in our context had 60 students on the official roster. The marginal cost of adding an additional school, assuming schools the same size as our pilot, is US$9/student. Larger schools will have a smaller per student cost.

The content development fixed costs were the most expensive part of this intervention. The two largest fixed costs were related to the video lectures and the interactive content. The video lectures were fully implemented, while the interactive content was not. The interactive content costs included the development of the in-class simulations that were available for teachers to use and the SMS, ITS, and IVR systems that were at most only marginally included in the intervention during our period of study. In the interest of transparency, we include the combined costs of all aspects of the intended intervention even though some pieces were not full implemented during the period of our study. For this 30 school pilot, including the full development costs of all aspects of the program, the cost per student was US$83. Taking this intervention to a slightly larger scale increases the cost-effectiveness substantially. A 50 school intervention would have an average cost of $53/student and a 200 school intervention would have an average cost of \$20/student.[31]

Comparing the cost-effectiveness of this intervention to others is difficult because most studies do not report cost-effectiveness. Of those that do, one approach is to scale the effects to the expected return for \$100 (Kremer, Brannen, and Glennerster 2013). At the 200 school scale, for \$100 our effective size would be 1.6SD in the combined math and science score. This level exceeds the cost effectiveness of the other technology interventions reported in Kremer, Brannen, and Glennerster (2013). A program that linked school committees to local governments in Indonesia (Pradhan 2012) was more cost effective. None of the other available studies attempted to transform what was happening in a middle school classroom. A second measure to consider in cost effectiveness is student time. Most other effective technology interventions included out of school time, in some cases multiple hours per week. Our intervention does not include any out of school time for students.

# 9 Discussion and Conclusions

The delivery of content through technology has the potential to improve student achievement within the existing school and teacher pre-service training structure. To test this hypothesis, we partnered with the government of Punjab, Pakistan to implement a randomized controlled trial of an intervention that provided classrooms with LED screens, math and science multimedia content, teacher tablets, teacher in-service training, and minimal at home SMS review questions via mobile phones. We found that the intervention increased achievement on both the project-specific and provincially standardized math and science tests by about 0.25 standard deviations with under 4 months of exposure. Prior to this study very little was known about improving student test scores in developing country middle schools.

In addition to providing relevant content, this program induced positive behavioral re-

[31]Removing the costs of the only partially implemented interactive content puts the costs at \$28/student at 30 schools, \$20/student at 50 schools, and \$15/student at 100 schools. In our setting, boundary walls and electricity were standard. Upgrading schools to include this infrastructure would increase the costs substantially, but could also confer additional benefits.

sponses by students and teachers who were both more likely to be present in school, demonstrating increased effort by both teachers and students.

The achievement effects are not heterogeneous by baseline test score showing that effective interventions targeting grade level content can improve test scores for all students despite varying levels of baseline achievement.

Finally, this program was cost-effective even at the 30 school scale and substantial fixed costs. At the mere 100 school scale the cost effectiveness is on par with some of the most cost-effective technology RCTs and beyond 100 schools the cost effectiveness exceeds them, not even taking into account the substantially smaller time investment by students.

Even though the exact implementation might vary across settings, we show that integrating a novel approach to teaching grade level material into the existing teaching practice increases effort by students and teachers and substantially increases middle school learning for students of all baseline learning levels, potentially overcoming existing teacher capacity constraints.

# References


Aker, J. C., C. Ksoll, and T. J. Lybbert (2012). Can mobile phones improve learning? evidence from a field experiment in niger. *American Economic Journal: Applied Economics 4*(4), 94–120.

Andrabi, T., J. Das, and A. I. Khwaja (2013). Students today, teachers tomorrow: Identifying constraints on the provision of education. *Journal of public Economics 100*, 1–14.

Andrabi, T., J. Das, A. I. Khwaja, N. Singh, and S. Ozyurt (2015). Upping the ante: The equilibrium effects of unconditional grants to private schools. *Unpublished paper*.

Andrabi, T., J. Das, A. I. Khwaja, T. Vishwanath, and T. Zajonc (2007). Learning and educational achievements in punjab schools (leaps): Insights to inform the education policy debate. *World Bank, Washington, DC*.

Bai, Y., D. Mo, L. Zhang, M. Boswell, and S. Rozelle (2016). The impact of integrating ict with teaching: Evidence from a randomized controlled trial in rural schools in china. *Computers and Education 96*, 1–14.

Banerjee, A., S. Cole, E. Duflo, and L. Linden (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*.

Banerjee, A., P. Glewwe, S. Powers, and M. Wasserman (2013). Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence. *Cambridge, MA: Massashusetts Institute of Technology*.

Bank, W. (2017). *World Development Report: Learning to Realize Education's Promise. 2018*. World Bank.

Barrera-Osorio, F. and A. J. Ganimian (2016). The barking dog that bites: Test score volatility and school rankings in punjab, pakistan. *International Journal of Educational Development 49*, 31 – 54.

Barrera-Osorio, F. and L. L. Linden (2009). The use and misuse of computers in education: evidence from a randomized experiment in colombia.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? results from a natural experiment in chile. *Economics of Education Review 28*(5), 629–640.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*(2), 29–50.

Burdett, N. (2017, December). Review of high stakes examination instruments in primary and secondary school in developing countries. Technical Report RISE-WP-17/018, RISE Working Paper.

Cai, X., Y. Lu, J. Pan, and S. Zhong (2019). Gender gap under pressure: Evidence from china's national college entrance examination. *Review of Economics and Statistics 101*(2), 249–263.

Carrillo, P., M. Onofa, and J. Ponce (2010). Information technology and student achievement: Evidence from a randomized experiment in ecuador. Inter-American Development Bank Working Paper IDB-WP-223.

Duflo, E., P. Dupas, and M. Kremer (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from kenyan primary schools. *Journal of Public Economics 123*, 92–110.

Duflo, E., P. Dupas, and M. Kremera (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *The American Economic Review 101*(5), 1739–1774.

Escueta, M., V. Quan, A. J. Nickow, and P. Oreopoulos (2017). Education technology: An evidence-based review. National Bureau of Economic Research Working Paper 23744.

Glewwe, P., M. Kremer, and S. Moulin (2009). Many children left behind? textbooks and test scores in kenya. *American Economic Journal: Applied Economics 1*(1), 112–35.

Glewwe, P., M. Kremer, S. Moulin, and E. Zitzewitz (2004). Retrospective vs. prospective analyses of school inputs: the case of flipcharts in kenya. *Journal of Development Economics 74*, 251–268.

Glewwe, P. and K. Muralidharan (2016). Improving school education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of Economics of Education*, Volume 5, pp. 653–743. North Holland.

He, F., L. Linden, and M. MacLeod (2008). How to teach english in india: Testing the relative productivity of instruction methods within the pratham english language education program. *New York, United States: Columbia University. Mimeographed document*.

Jackson, C. K. (2010). Do students benefit from attending better schools? evidence from rule-based student assignments in trinidad and tobago. *The Economic Journal 120*(549), 1399–1429.

Johnston, J. and C. Ksoll (2017). Effectiveness of interactive satellite-transmitted instruction: Experimental evidence from ghanaian primary schools. Center for Education Policy Analysis Working Paper 17-08.

Kremer, M., B. Conner, and R. Glennerster (2013). The challenge of education and learning in the developing world. *ScienceMag 340*.

Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle (2015). Does computer-assisted learning improve learning outcomes? evidence from a randomized experiment in migrant schools in beijing. *Economics of Education Review 47*, 34–48.

Lai, F., L. Zhang, Y. Bai, C. Liu, Y. Shi, F. Chang, and S. Rozelle (2016). More is not always better: evidence from a randomised experiment of computer-assisted learning in rural minority schools in qinghai. *Journal of Development Effectiveness 8*(4), 449–472.

Lai, F., L. Zhang, X. Hu, Q. Qu, Y. Shi, Y. Qiao, M. Boswell, and S. Rozelle (2013). Computer assisted learning as extracurricular tutor? evidence from a randomised experiment in rural boarding schools in shaanxi. *Journal of Development Effectiveness 5*(2), 208–231.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Linden, L. L. (2008). Complement or substitute?: The effect of technology on student achievement in india.

Lucas, A. M. and I. M. Mbiti (2014). Effects of school quality on student achievement: Discontinuity evidence from kenya. *American Economic Journal: Applied Economics 6*(3), 234–263.

Lucas, A. M., P. J. McEwan, M. Ngware, and M. Oketch (2014). Improving early-grade literacy in east africa: Experimental evidence from kenya and uganda. *Journal of Policy Analysis and Management 33*(4), 950–976.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries a meta-analysis of randomized experiments. *Review of Educational Research 85*(3), 353–394.

Mo, D., J. Swinnen, L. Zhang, H. Yi, Q. Qu, M. Boswell, and S. Rozelle (2013). Can one-to-one computing narrow the digital divide and the educational gap in china? the case of beijing migrant schools. *World Development 46*, 14–29.

Muralidharan, K. (2013). Priorities for primary education policy in indias 12th five-year plan. In *India Policy Forum*, Volume 9, pp. 1–61. National Council of Applied Economic Research.

Muralidharan, K., A. Singh, and A. J. Ganimian (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review 109*(4), 1426–1460.

Naslund-Hadley, E., S. W. Parker, and J. M. Hernandez-Agramonte (2014). Fostering early math comprehension: Experimental evidence from paraguay. *Global Education Review 1*(4).

Newman, J., M. Pradham, L. B. Rawlings, G. Ridder, R. Coa, and J. L. Evia (2002). An impact evaluation of education, health, and water supply investments by the bolivian social investment fund. *The World Bank Economic Review 16*(2), 241–274.

of Pakistan", G. (2014). National assessment report. Technical report.

Pop-Eleches, C. and M. Urquiola (2013). Going to a better school: Effects and behavioral responses. *The American Economic Review 103*(4), 1289–1324.

Pradhan, M., D. Suryadarma, A. Beatty, M. Wong, A. Alishjabana, A. Gaduh, and R. P. Artha (2011). *Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia*. The World Bank.

# 10 Appendix

In the following tables we provide a number of additional estimations.

[Appendix Table A1 about here]

[Appendix Table A2 about here]

Figure 1: Study and Academic Year Timeline

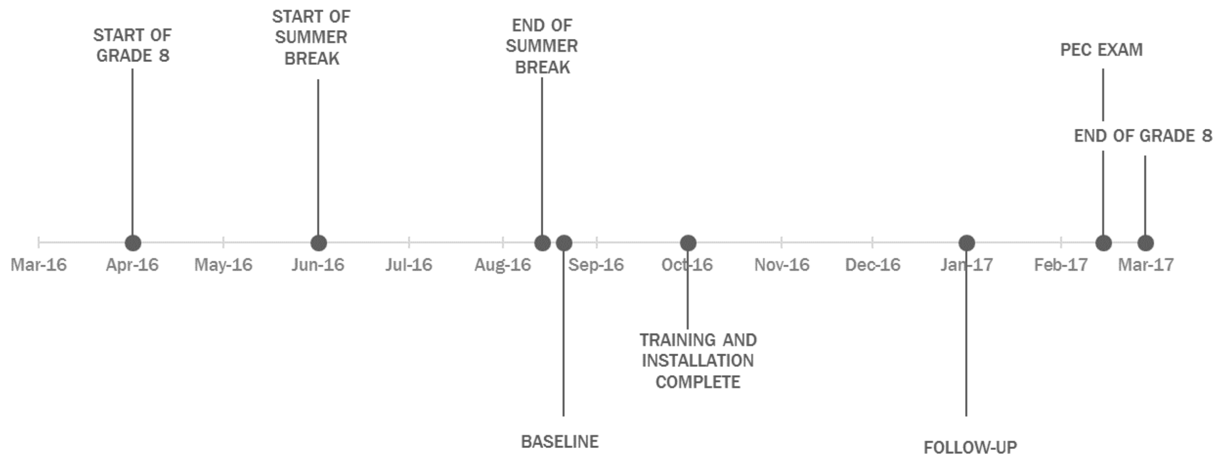
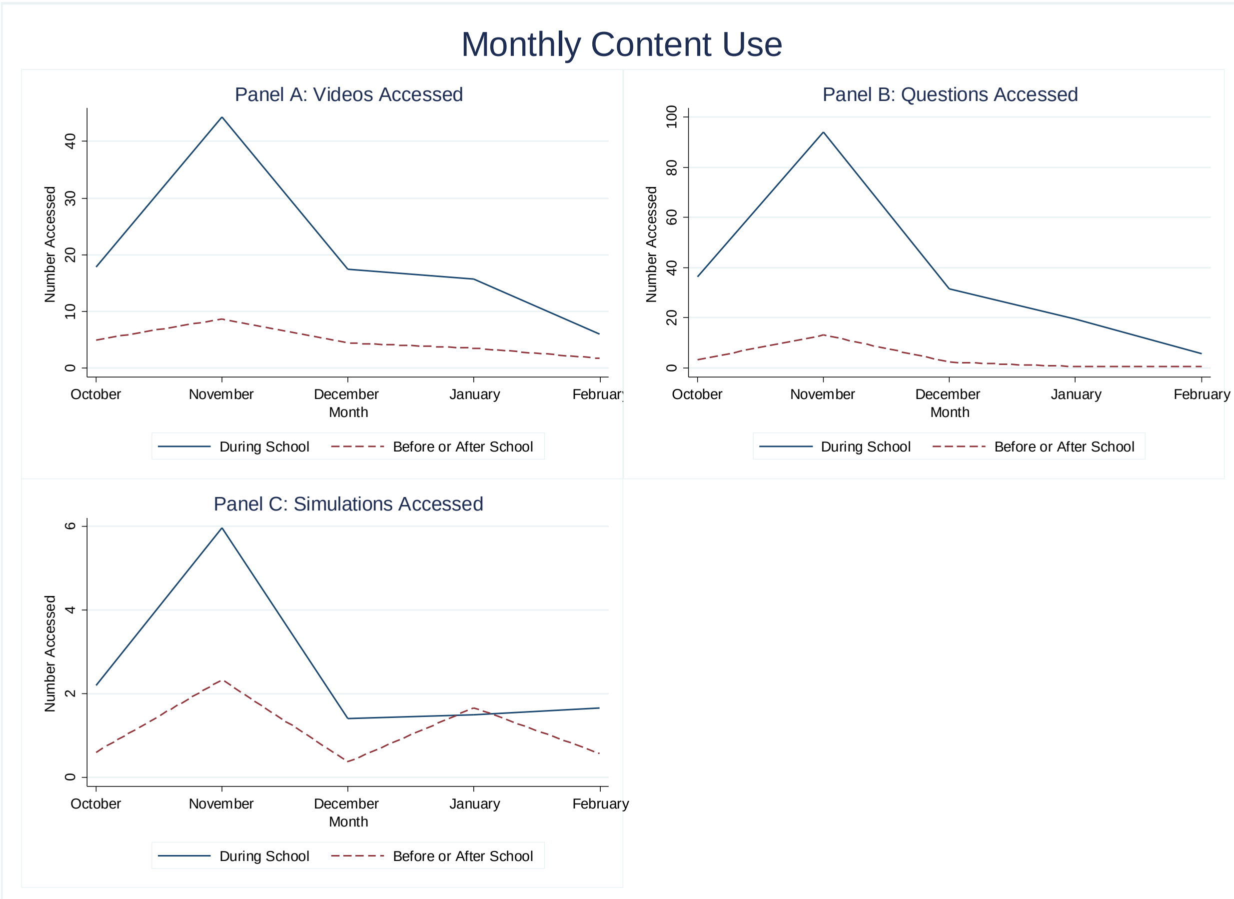


Figure 2: Monthly Content Use by Teachers




Notes: Based on data collected by tablets. Program was implemented in October.

Table 1: Summary Statistics - Student and Teachers

| | Treatment (1) | Control (2) | Difference T-C (3) | | Treatment (4) | Control (5) | Difference T-C (6) |
|---|---|---|---|---|---|---|---|
| *Panel A: Student Characteristics* | | | | *Panel B: Teacher Characteristics* | | | |
| Math Score | -0.101<br>(0.939) | 0.107<br>(1.055) | -0.222<br>(0.182) | Use Technology to Prepare for Class | 0.581<br>(0.497) | 0.596<br>(0.495) | -0.0158<br>(0.106) |
| Science Score | 0.0230<br>(0.981) | -0.00802<br>(1.013) | 0.0466<br>(0.165) | Use Technology in Class | 0.155<br>(0.365) | 0.170<br>(0.379) | -0.0146<br>(0.0687) |
| Household Computer | 0.427<br>(0.495) | 0.407<br>(0.492) | 0.0195<br>(0.0484) | Years of Teaching Experience | 10.79<br>(8.480) | 10.93<br>(9.365) | -0.140<br>(1.796) |
| Household Cell Phone | 0.999<br>(0.0365) | 1.0000<br>(0.000) | -0.00133<br>(0.00131) | Has an Advanced Degree | 0.754<br>(0.434) | 0.789<br>(0.411) | -0.0354<br>(0.0819) |
| Use Technology to Study | 0.372<br>(0.484) | 0.292<br>(0.455) | 0.0805<br>(0.0552) | Part of Any In-Service Training | 0.919<br>(0.275) | 0.965<br>(0.186) | -0.0456<br>(0.0475) |
| Days Absent Last Week | 1.495<br>(2.409) | 1.167<br>(1.788) | 0.327*<br>(0.173) | Holds Private Tutoring Sessions | 0.113<br>(0.319) | 0.0702<br>(0.258) | 0.0427<br>(0.0542) |
| Take Tuitions Outside of School | 0.457<br>(0.499) | 0.473<br>(0.500) | -0.0151<br>(0.0542) | Performs other Official Duties | 0.613<br>(0.491) | 0.625<br>(0.489) | -0.0121<br>(0.118) |
| Hours Spent on Homework per | 1.612<br>(0.982) | 1.527<br>(0.969) | 0.0843<br>(0.112) | Number of Classes Taught per Month | 36.00<br>(7.071) | 32.70<br>(9.126) | 3.298*<br>(1.907) |
| Work | 0.0319<br>(0.176) | 0.0237<br>(0.152) | 0.00821<br>(0.0128) | Extra Classes Per Month | 3.871<br>(7.623) | 2.545<br>(5.350) | 1.326<br>(1.448) |
| Parents Visit School | 0.572<br>(0.495) | 0.603<br>(0.490) | -0.0312<br>(0.0647) | Student Approach for Help | 0.0492<br>(0.218) | 0.0526<br>(0.225) | -0.00345<br>(0.0473) |
| Expect to Attend College | 0.616<br>(0.487) | 0.625<br>(0.484) | -0.00949<br>(0.0567) | Minutes per Day Planning Lessons | 40.32<br>(33.69) | 33.51<br>(27.93) | 6.814<br>(5.548) |
| Number of Siblings | 3.949<br>(1.827) | 3.803<br>(1.765) | 0.147<br>(0.157) | | | | |
| Age | 13.90<br>(1.237) | 13.87<br>(1.231) | 0.0266<br>(0.102) | | | | |
| Meat Cooked (days per week) | 0.738<br>(0.440) | 0.696<br>(0.460) | 0.0417<br>(0.0393) | | | | |
| Number of Motorbikes | 1.036<br>(0.883) | 0.970<br>(0.864) | 0.0655<br>(0.0703) | | | | |
| Number of Cars | 0.168<br>(0.460) | 0.133<br>(0.411) | 0.0342<br>(0.0313) | | | | |

*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.

Table 2: Summary Statistics - School Characteristics

| | Treatment (1) | Control (2) | Difference T-C (3) |
|---|---|---|---|
| Math Teacher Has a College Degree | 0.333 | 0.138 | 0.195* |
| | (0.479) | (0.351) | (0.109) |
| Math Teacher Has Masters Degree | 0.500 | 0.655 | -0.155 |
| | (0.509) | (0.484) | (0.129) |
| Math Teacher Has a PhD | 0.00 | 0.000 | 0.00 |
| | (0) | (0) | (0) |
| Math Teacher Years of Experience | 12.14 | 14.62 | -2.479 |
| | (8.807) | (9.666) | (2.410) |
| Science Teacher Has a College Degree | 0.0667 | 0.138 | -0.0713 |
| | (0.254) | (0.351) | (0.0799) |
| Science Teacher Has Masters Degree | 0.900 | 0.759 | 0.141 |
| | (0.305) | (0.435) | (0.0982) |
| Science Teacher Has a PhD | 0.00 | 0.0345 | -0.0345 |
| | (0) | (0.186) | (0.0345) |
| Science Teacher Years of Experience | 3.533 | 3.948 | -0.415 |
| | (2.675) | (3.878) | (0.870) |
| Total Enrollment in Grade 8 | 63.10 | 63.21 | -0.107 |
| | (16.36) | (13.52) | (3.901) |
| Grade 8 Sections | 1.400 | 1.345 | 0.0552 |
| | (0.498) | (0.484) | (0.128) |
| Grade 8 Students Present | 49.10 | 50.31 | -1.210 |
| | (15.33) | (11.59) | (3.531) |
| Average 2016 Math PEC Score | -0.0768 | -0.0201 | -0.057 |
| | (0.398) | (0.596) | (0.132) |
| Average 2016 Science PEC Score | -0.0341 | 0.109 | -0.143 |
| | (0.524) | (0.412) | (0.123) |
| School Has a Library | 0.700 | 0.552 | 0.148 |
| | (0.466) | (0.506) | (0.127) |
| School Has a Playground | 0.633 | 0.517 | 0.116 |
| | (0.490) | (0.509) | (0.130) |
| School Has a Computer Lab | 0.900 | 1.000 | -0.100* |
| | (0.305) | (0) | (0.0557) |
| School Principal Has Masters Degree | 0.833 | 0.690 | 0.144 |
| | (0.379) | (0.471) | (0.111) |
| School Principal Has PhD | 0.133 | 0.310 | -0.177 |
| | (0.346) | (0.471) | (0.108) |
| School Principal Years of Experience | 3.533 | 3.948 | -0.415 |
| | (2.675) | (3.878) | (0.870) |

*Notes* : Columns 1 and 2: Standard deviations appear in parenthesis. Column 3: Cluster-robust standard errors appear in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Achievement Effects - Project-Specific Exams

| | Standardized Test Score | | |
|---|---|---|---|
| | Math | Science | Math + Science |
| | (1) | (2) | (3) |
| *Panel A: Limited Controls* | | | |
| Treatment | 0.185* | 0.239* | 0.256* |
| | (0.109) | (0.141) | (0.135) |
| Additional Controls | | | |
| Baseline Math Score | 0.190*** | 0.256*** | 0.270*** |
| | (0.0352) | (0.0510) | (0.0458) |
| Baseline Science Score | 0.144*** | 0.117*** | 0.162*** |
| | (0.0344) | (0.0371) | (0.0390) |
| Observations | 2,622 | 2,622 | 2,622 |
| R-Squared | 0.101 | 0.090 | 0.133 |
| *Panel B: School Level Controls* | | | |
| Treatment | 0.260** | 0.274* | 0.326** |
| | (0.113) | (0.141) | (0.132) |
| Additional Controls | | | |
| Baseline Math Score | 0.120*** | 0.145*** | 0.161*** |
| | (0.0409) | (0.0506) | (0.0511) |
| Baseline Science Score | 0.110*** | 0.0889** | 0.123*** |
| | (0.0286) | (0.0402) | (0.0340) |
| Observations | 2,622 | 2,622 | 2,622 |
| R-Squared | 0.167 | 0.172 | 0.220 |
| Average Control Group Change | 0.21 | 0.49 | 0.42 |

*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. Panel A: strata and baseline test scores only. Panel B: controls in Panel A and baseline total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library.

Table 4: Achievement Effects - Provincial Exam (PEC)

| | Standardized Test Score | | | | | |
|---|---|---|---|---|---|---|
| | Math | Science | Math + Science | Other Subject Tests | Overall Average | Pass |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A: Limited Controls* | | | | | | |
| Treatment | 0.155 | 0.187 | 0.191 | 0.052 | 0.122 | 0.038 |
| | (0.137) | (0.122) | (0.122) | (0.0922) | (0.104) | (0.0262) |
| Additional Controls | | | | | | |
| Baseline Math Score | 0.0941** | 0.128*** | 0.124*** | 0.119*** | 0.129*** | 0.0271*** |
| | (0.0404) | (0.0304) | (0.0356) | (0.0309) | (0.0328) | (0.00697) |
| Baseline Science Score | 0.0323 | 0.0526 | 0.0473 | 0.0676*** | 0.0626** | 0.00985* |
| | (0.0472) | (0.0315) | (0.0385) | (0.0250) | (0.0298) | (0.00584) |
| Observations | 2,798 | 2,798 | 2,798 | 2,798 | 2,798 | 2,798 |
| R-Squared | 0.17 | 0.22 | 0.23 | 0.21 | 0.23 | 0.06 |
| *Panel B: School Level Controls* | | | | | | |
| Treatment | 0.176 | 0.333** | 0.283* | 0.143 | 0.220 | 0.0686* |
| | (0.177) | (0.163) | (0.156) | (0.117) | (0.139) | (0.0359) |
| Additional Controls | | | | | | |
| Baseline Math Score | 0.114*** | 0.105*** | 0.122*** | 0.121*** | 0.130*** | 0.0198*** |
| | (0.0357) | (0.0322) | (0.0358) | (0.0258) | (0.0306) | (0.00715) |
| Baseline Science Score | 0.0499 | 0.0199 | 0.0392 | 0.0573** | 0.0527* | 0.00468 |
| | (0.0322) | (0.0285) | (0.0298) | (0.0239) | (0.0267) | (0.00637) |
| Observations | 2,798 | 2,798 | 2,798 | 2,798 | 2,798 | 2,798 |
| R-Squared | 0.24 | 0.27 | 0.28 | 0.25 | 0.27 | 0.12 |
| Control Group Mean | | | | | | 0.90 |

*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 and PEC exams. Panel A: includes strata and school level prior year PEC scores as controls. Panel B: controls from Panel A and baseline total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, library.

Table 5: Attendance and Attrition

| | Students: | | Teachers: | | |
|---|---|---|---|---|---|
| | Present at Follow-up | | Student Absenteeism is a Signifcant Problem | Portion Present | |
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | 0.0530** | 0.0539** | -0.177* | 0.00988* | 0.0214** |
| | (0.0204) | (0.0204) | (0.103) | (0.00579) | (0.0101) |
| Treatment X Baseline Score | | 0.0141 | | | |
| | | (0.0173) | | | |
| Treatment X Months of Treatment | | | | | -0.00598 |
| | | | | | (0.00366) |
| Observations | 2,999 | 2,999 | 115 | 274 | 274 |
| R-Squared | 0.02 | 0.02 | 0.37 | 0.26 | 0.27 |
| Control Group Mean | 0.85 | | 0.60 | 0.94 | |

*Notes* : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Additional controls: strata and baseline test scores, total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library. Columns 1-2: Includes all students who took the baseline test. Column 3: All teachers who completed the follow-up survey. Columns 4-5: The portion of all teachers present in the school during an unannounced spot check. Measured monthly for each school.

Table 6: Changes in Inputs - Technology and Training

| | Teacher Uses Technology | | Number of In-service Trainings This Year |
|---|---|---|---|
| | To Prepare for Lessons | In the Classroom | |
| | (1) | (2) | (3) |
| Treatment | 0.419*** | 0.766*** | 0.443*** |
| | (0.0783) | (0.0727) | (0.127) |
| Observations | 115 | 115 | 115 |
| R-Squared | 0.35 | 0.75 | 0.39 |
| Control Group Mean | 0.60 | 0.17 | 3.62 |

*Notes* : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Additional controls: strata and baseline test scores, total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library. Includes all teachers surveyed at follow-up. Columns 1 and 2: Linear probability models.

Table 7: Achievement Effects - Heterogeneous Effects

| | Standardized Test Score | | | | | | |
|---|---|---|---|---|---|---|---|
| | Project Exams | | | PEC Exams | | | |
| | Math | Science | Math + Science | Math | Science | Math + Science | Overall Score |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A: By Baseline Test Score* | | | | | | | |
| Treatment | 0.254** | 0.269* | 0.319** | 0.182 | 0.339** | 0.290* | 0.227 |
| | (0.116) | (0.143) | (0.135) | (0.180) | (0.164) | (0.158) | (0.140) |
| Treatment X Baseline Score | -0.0586 | -0.0435 | -0.0645 | 0.0846 | 0.0916 | 0.0985 | 0.102 |
| | (0.0803) | (0.0897) | (0.0966) | (0.0871) | (0.0740) | (0.0801) | (0.0692) |
| Observations | 2,622 | 2,622 | 2,622 | 2,798 | 2,798 | 2,798 | 2,798 |
| R-Squared | 0.17 | 0.17 | 0.22 | 0.24 | 0.27 | 0.28 | 0.27 |
| *Panel B: By School Gender* | | | | | | | |
| Treatment | 0.398*** | 0.321 | 0.446*** | -0.0851 | 0.105 | 0.00993 | 0.0928 |
| | (0.137) | (0.195) | (0.165) | (0.233) | (0.194) | (0.207) | (0.201) |
| Treatment X Female School | -0.259 | -0.0875 | -0.227 | 0.498 | 0.435* | 0.522** | 0.242 |
| | (0.259) | (0.337) | (0.314) | (0.320) | (0.233) | (0.229) | (0.226) |
| Observations | 2,622 | 2,622 | 2,622 | 2,798 | 2,798 | 2,798 | 2,798 |
| R-Squared | 0.17 | 0.17 | 0.22 | 0.25 | 0.276 | 0.289 | 0.272 |
| F-test of coefficients on Treatment + Treament X Female School=0 | | | | | | | |
| p-value | 0.48 | 0.34 | 0.35 | 0.12 | 0.01 | 0.01 | 0.04 |
| Average Control Group Change | 0.21 | 0.49 | 0.42 | | | | |

*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. Additional controls: Additional controls: strata and baseline test scores, total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library.

Table 8: Robustness Checks

| | Preferred Specification | LASSO Controls | Survey Sample: School Level Controls | Survey Sample: Student and School Controls | Adjusted Attrition: Lower Bound | Adjusted Attrition: Upper Bound |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A: Math Score* | | | | | | |
| Treatment | 0.260** | 0.211** | 0.275** | 0.272** | 0.263** | 0.268** |
| | (0.113) | (0.105) | (0.126) | (0.124) | (0.115) | (0.107) |
| Observations | 2,622 | 2,622 | 1,262 | 1,262 | 2,553 | 2,552 |
| R-Squared | 0.167 | | 0.17 | 0.19 | 0.17 | 0.18 |
| *Panel B: Science Score* | | | | | | |
| Treatment | 0.274* | 0.272** | 0.110 | 0.105 | 0.255* | 0.277* |
| | (0.141) | (0.137) | (0.130) | (0.126) | (0.144) | (0.140) |
| Observations | 2,622 | 2,622 | 1,262 | 1,262 | 2,553 | 2,552 |
| R-Squared | 0.172 | | 0.20 | 0.21 | 0.17 | 0.18 |
| *Panel C: Math + Science Score* | | | | | | |
| Treatment | 0.326** | 0.291** | 0.242* | 0.238* | 0.318** | 0.333** |
| | (0.132) | (0.129) | (0.123) | (0.120) | (0.135) | (0.126) |
| Observations | 2,622 | 2,622 | 1,262 | 1,262 | 2,553 | 2,552 |
| R-Squared | 0.220 | | 0.24 | 0.26 | 0.22 | 0.23 |

*Notes* : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Column 1: From Table 3. Column 2: Covariates determined by LASSO. Columns 3-5: Additional controls: strata and baseline test scores, total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library. Columns 3 and 4: Sample limited to students who completed the baseline survey. Column 4: Additional student level controls. Columns 5 and 6: sample adjusted based on Lee (2009).

## Appendix Table A1: Matching Between Baseline and PEC

| | Matched to PEC | |
|---|---|---|
| | (1) | (2) |
| Treatment | -0.0192 | -0.0184 |
| | (0.0123) | (0.0119) |
| Treatment X Baseline Score | | 0.0130 |
| | | (0.0100) |
| Observations | 2,999 | 2,999 |
| R-Squared | 0.01 | 0.01 |
| Control Group Mean | 0.94 | |

*Notes* : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Additional controls: strata and baseline test scores, total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library. Includes all students who took the baseline test.

Appendix Table A2: Changes in Other Teacher Inputs

| | 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 | Performs Official Non-teaching Duties |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 12.43* | 0.136* | -0.790 | 0.631 | 0.144 | 0.0657 |
| | (6.560) | (0.0749) | (1.323) | (1.544) | (0.111) | (0.0991) |
| Observations | 115 | 115 | 115 | 115 | 115 | 115 |
| R-Squared | 0.23 | 0.20 | 0.48 | 0.27 | 0.20 | 0.31 |
| Control Group Mean | 34.15 | 0.076 | 33.19 | 2.75 | 0.45 | 0.63 |

*Notes* : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Additional controls: strata and baseline test scores, total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library. Includes all teachers surveyed at follow-up. Columns 2, 5, and 6: Linear probability models.

Appendix Table A3: Student and Household Inputs and Outcomes

| | Use Technological Study Aid at Home | Time Spent Studying (minutes per day) | Number of Days Absent in the Last Week | Receive Out of School Tutoring | Parents Visit School to Meet with Teacher | Expect to Attend University |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.0416 | 2.477 | -0.118 | -0.0206 | 0.0801 | 0.0427 |
| | (0.0344) | (4.226) | (0.0765) | (0.0307) | (0.0486) | (0.0448) |
| Observations | 1,270 | 1,249 | 1,270 | 1,270 | 1,270 | 1,270 |
| R-Squared | 0.107 | 0.172 | 0.081 | 0.373 | 0.234 | 0.230 |
| Baseline Mean | 0.29 | 97.49 | 1.17 | 0.47 | 0.60 | 0.63 |

*Notes* : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Additional controls: strata and baseline test scores, total enrollment, number of grade 8 sections, number of students present in grade 8, head teacher qualification and tenure, math and science teacher qualification, tenure and employment status; and dummy variables for computer lab, playground, and library. Students who completed the baseline and endline survey. Columns 1, 4-6: Linear probability models.