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

Beg: Department of Economics, University of Delaware. Lucas: Department of Economics, University of Delaware and NBER. Halim and Saif: IT University Lahore. 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 University of Virginia, 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 AEARCTR-0003536. The University of Delaware IRB determined 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
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 achievement 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 efforts 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 2014). 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 Pakistan specifically, these learning deficiencies start in primary school (Banerjee et al. 2013) 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. Middle school education suffers from at least two challenges beyond those experienced by primary school. First, teachers themselves need to be knowledgeable and trained in more complex subject matter. Second, subject-specific textbooks can quickly become outdated. 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). In developed countries, the growing field of education technology (ed-tech) is lauded for its potential to deliver high quality education without substantially retraining existing teachers (Escueta et al. 2017). In lower-resource settings, the model of a classroom full of computers or tablets is likely unsustainable. Our intervention leverages technology through a government implemented, scalable, relatively inexpensive program to address learning deficiencies. 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, 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 track student learning. 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 only 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 as part of the province-wide "Education Road Map." 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 an 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 this test that students take at the end of grade 8. Further, our intervention did not shift the focus away from other subjects as we find no changes to the other subject scores on the provincial standardized test. In contrast with other interventions that targeted grade level material and found larger, or the only gains, among students with high scores at baseline, students across the baseline test score distribution gained equally and it increased the likelihood that students passed the grade 8 test by 7 percentage points (Glewwe et al. 2004; Glewwe,

[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.

Kremer, and Moulin 2009). Therefore, grade level content implemented at the middle school level does not necessarily exacerbate pre-existing achievement heterogeneity.[2] When we test for heterogeneous effects by gender, we find some evidence that girls' scores increased more than boys on the provincially standardized exams.[3]

To understand the complementarity vs. substitutability of inputs in the education production function, we estimate the treatment effect on both student and teacher effort. We find that both students and teachers are more likely to be present in school.[4] Teachers in the treatment schools spend more time preparing for lessons and are more likely to report that students directly asked them for help outside of class time. We do not find any statistically significant changes in the self-reported likelihood that teachers hold private tutoring outside of school hours. We find almost no evidence of any heterogeneous changes in effort by the gender of pupils in the school.

Effective, inexpensive methods to improve achievement at the post-primary level or beyond foundational literacy and numeracy 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 (2016) found positive achievement effects for middle school students from an after school program that combined computer assisted learning (CAL) 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.

Most of the research on the effects of increasing school-based inputs in developing countries has focused on foundational literacy and numeracy skills in primary school. Research

[2] 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.

[3] As is common 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.

[4] This difference in attendance is not differential by both treatment status and achievement. Our estimates of Lee (2009) bounds confirm the achievement effects.

on primary 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 2012 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 and Kremer 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 CAL software. 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 CAL 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 content specific achievement beyond foundational literacy and numeracy. 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 sizes. The marginal cost per school was \$9/student. Even with the inclusion of the substantial content fixed costs, the cost would be \$15/student at

[5] For example, Banerjee et al. (2007), Linden (2008), He, Linden, and MacLeod (2008), Barerra-Osorio and Linden (2009), Carrillo et al. (2010), Lai et al. (2013), Mo et al. (2014), Lai et al. (2015), Bai et al. (2016), and Lai et al. (2016)

[6] The extent to which the SMS part of the intervention was implemented, it builds on Aker et al. (2013) that found that mobile technology was a complement to rather than a substitute for highly educated teachers.

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, as part of of the Education Road Map, 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 two-month break in June and July, 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, all students take provincial standardized exams. A student's score on this test signals completion of middle school and partially determines 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 Punjab, our region of study, Andrabi et al. (2007) found that government schools lacked basic facilities and teaching

[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. Similarly, Andrabi, D, Kremer, Moulinas, and Khwaja (2013) found that unavailability of qualified teachers in Pakistan constrained educational provision. 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 using a "chalk and talk" method to teach a classroom of students (Glewwe and Muralidharan 2015). 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 all students adequately follow the subject material. Additional instructional tools and training teachers on incorporating them in lectures could have a substantial impact.

3 Intervention

Our intervention, eLearn, was part of Punjab's broader "Education Road Map" to improve education throughout the province. 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. This small-scale implementation of the program was designed to inform the larger scale-up of the program that started in 2018.

To increase student learning, our eLearn intervention 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

[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

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 see and display these video lectures and multimedia content, teachers were given small, pre-loaded tablets. Teachers could use these tablets to watch the videos themselves when preparing for lectures and project them on installed LED television screens. These 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.

To display this content to the class, 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.

Teachers received a two day in-service training session that oriented them on the new technologies and introduced a blended learning approach that combined their own face-to-face teaching with technology-enabled multimedia content. The training encouraged them to use the technology to enhance their role in the classroom and trained teachers on how to use the videos, related assessment tools, and activities on their tablets for more effective teaching. 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

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.

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.[10] 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-July 2016 school holidays. The teacher trainings and hardware installation occurred in October. Our follow-up surveys and exams occurred in January. The PEC standardized exams occurred in February. Therefore, students and teachers were exposed to the intervention for at most 4 months between the baseline and follow-up testing.

[10] 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 Conceptual Approach and 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 60 school randomized controlled trial of our intervention. In this section we first provide a simplified framework to demonstrate how the intervention may enter the education production function and then provide our empirical strategy.

4.1 Conceptual Approach

Consider a simplified education production function in which a student's achievement (A) is a function of inputs (x), and effort (e), where e may respond to the other inputs. Thus we have

$$A = f(x, e(x)) \tag{1}$$

Our experimental design increased the teaching inputs available to a classroom. Since these items were not already available to schools, this shock should be an unequivocal increase in x.[11]

When we estimate the reduced form effect of our program, we are effectively taking the total derivative of Equation 1: $\frac{dA}{dx}$. Our first empirical specifications will do this.

In further exploration of the effects, we will consider the partial derivatives that make up the total derivative:

$$\frac{dA}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial e}\frac{\partial e}{\partial x}$$

If technology is a substitute for effort or other inputs, then $\frac{\partial e}{\partial x} < 0$, implying that schools, teachers, or students could react to this increase in technology by reducing their own effort. For example, despite their training on how to use this technology as a complement to their

[11] The increased technology could have been a substitute for other non-technology items in the classroom. Since we do not observe changes in these items, for simplicity, we remove this possible dependency.

own teaching, teachers might instead use it as a substitute reducing their own hours of preparation or attendance. Instead, if technology and effort or other inputs are complements, i.e. $\frac{\partial e}{\partial x} > 0$, then the increase in technology should increase the optimal level of effort. For example, students might be inspired by seeing someone other than the regular teacher explain a topic and put forth more effort. Similarly, teachers might feel more comfortable with the content and attend more. Therefore, we will provide estimates of $\frac{\partial e}{\partial x}$, the effect of the change in the technological inputs on a number of measures of teacher and student effort. These student and teacher responses could reinforce or reduce the overall effect of the program. Finally, we will hold any changes in effort constant and estimate the effort adjusted effect of the program.

4.2 Empirical Strategy

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

$$y_{is} = \alpha + \beta treatment_s + X_{is}^{'}\Gamma + \varepsilon_{is} \tag{2}$$

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 individual level controls, and ε_{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 baseline class size, principal qualifications, teacher qualifications, and school facilities.[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 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 is the total derivative, including any changes of 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 and better understand the partial vs. full derivative in the education production function, 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.

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

[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.

[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.

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 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.[14] 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 in order to securely install and power the LED screens. Further, schools had to be in areas with mobile phone coverage. 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.

[14] 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.

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 August 2016, near the beginning of the 2016-2017 academic year, after the June-July holiday but prior to the teacher training 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. After installation of the required equipment and providing the teacher trainings in October, the interventions proceeded in treatment schools. During the school year, treatment schools were visited by the implementing partner's technology support team to ensure equipment was secure and functioning as intended.[15] We administered follow-up surveys and exams in January 2017, near the end of the academic year. The same students were again surveyed and tested if possible. Head teachers and grade 8 subject teachers were again surveyed. All visits were unannounced. Participant schools were told at the baseline that we would be returning but not given a specific month or day.

When estimating the effects on test scores, we use item response theory to convert raw test responses to approximated latent student ability, and standardize based on the baseline mean and standard deviation. 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. These data were merged at the student level to the students in our sample. Because we do not have item level responses, these scores are simply scaled with a mean 0 with standard deviation of 1.

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

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

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 (School Education Department, 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 school checks. Monitoring and Evaluation Assistants conduct monthly, unannounced school visits and record teacher attendance. 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 44 different outcomes, some small differences are expected.

[Tables 1 and 2 about here]

6 Results

We first test for the effects of the program on achievement for both the project-specific and PEC exams. Then, we explore possible mechanisms behind the achievement results including student attendance and likelihood of taking the PEC exam, interesting outcomes themselves as well as potentially important to consider when interpreting the achievement results.

6.1 Achievement

To estimate the effect of the program on achievement we estimate Equation 2 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 combined score. Panel A includes only the strata and baseline test scores as 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. During this same period control 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. 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 reflect the official curriculum while 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 tests that all grade 8 students take in Punjab province. As with the previous table, Panel A includes limited controls and Panel B includes additional baseline controls including the average subject PEC score from each school from the prior year as a control. 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 0.28 standard deviation effect on the combined math and science test score, but the effect is not statistically significant for the math score alone.

[Table 4 about here]

One possible concern with the improvement in the math and science scores is that our interventions caused effort to shift away from other subjects towards the math and science curriculum. Column 4 of Table 4 provides estimates of the effect of the treatment on combined score of the subject tests that were not the focus of the intervention. The point estimate is positive and statistically insignificant. Therefore, we do not find evidence that the intervention led to a decrease 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 the intervention increased this likelihood by 7 percentage points.

6.2 Observed Attendance and Grade Completion

The achievement results in the Section 6.1 were the total derivative of the effect of the program 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 2, replacing the dependent variable each time with an intermediate 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 or if they took the end of year exam, a decent proxy for completing grade 8. In each case we use an indicator variable equal to 1 if the student was present at follow-up (took the exam) for the outcome, y_{is}, in Equation 2. As with all our other binary outcomes, we estimate the result with a linear probability model. The results appear in Table Columns 1 through 4 of Table 5.

[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 for the possibility of differential attrition by treatment status and ability by including an interaction of the treatment times the 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 Robustness Section we follow Lee (2009) and provide treatment bounds.

Columns 3 and 4 of Table 5 provide similar estimates for whether students completed the PEC exam in the baseline school, a key step to completing grade 8.[16] The point estimate of the treatment coefficient is both smaller in absolute magnitude than for the project exams, negative, and statistically insignificant. In column 4 we test for a differential effect by both treatment status and baseline achievement and find that for each additional one standard

[16] We are unable to locate students in PEC data if they transferred to a new school. Therefore, we cannot differentiate transfers from drop outs. Transferring in grade 8 is rare.

deviation in test score, treatment students were 1.5 percentage points more likely to take the PEC exam, a small increase over the control group mean of 93 percent.

As an objective measure of teacher effort we rely on Punjab Monitoring and Implementation Unit administrative data that records teacher attendance at the school level from a monthly unannounced visit. In columns 5 and 6 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 2 but with multiple observations per school. Overall teacher attendance is high: approximately 94 percent. Our intervention increased the teacher attendance rate by almost 1 percentage point (column 5). As this is a monthly measure, we can test the evolution over time in teacher attendance. In column 6 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 1 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. 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 appears to have been an important part of the intervention.

6.3 Self-Reported Outcomes

We additionally collected data on self-reported changes in take-up and inputs. As the intervention could have displaced other technology, we first test for whether the change in inputs increased the use of technology available to teachers (effectively $\frac{\partial f}{\partial x}$). Table 6 contains these results. Teachers were 35 percentage points more likely to report that they used technology to prepare for lessons (column 1) and 78 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."

Similarly, this could have displaced other trainings. Column 3 shows that treatment teachers attended 0.42 more in-serivce teacher training events during the school year.

[Table 6 about here]

In Table 7, we test for additional changes in teacher effort that might have occurred as a result of the intervention. We do not find any statistically significant effects on the likelihood of holding private tutoring sessions (column 2), the number of regularly scheduled classes taught per week (column 3), the number of extra classes per month during the school day (column 4), or the likelihood of being required to perform additional duties in addition to teaching (column 6). Two outcomes are statistically significant at the 10% level: teachers spent 11 more minutes per day planning lessons (column 1) and were more likely to report that students approached them outside of class during the school day for extra help in math or science (column 5). Therefore, teachers in treatment schools increased their use of technology and their observed effort (attendance) and at most marginally increased their self-reported effort.

[Table 7 about here]

In Table 8 we test 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, and whether they received out of school tutoring, their parents visited the school to meet with the teacher, or they expected to attend university.

[Table 8 about here]

In Table 9 we control for these behavioral changes at the teacher level (reported in Table 7 above) as an informal test of whether the intervention increased achievement directly net of any behavioral changes by teachers. The results are of similar magnitude of the original estimates, but no longer statistically significant for the science test score. Therefore, the intervention likely had a direct effect on students' achievement net of changes in teacher effort in response to the intervention.

[Table 9 about here]

7 Heterogeneity and Robustness

Because the intervention videos were at the level of the curriculum and some of the students could have been well behind grade level, the intervention could have differential effects by baseline test score. Panel A of Table 10 tests for this possibility by including an interaction between baseline test score and treatment as an additional regressor.[17] 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 statistically insignificant (columns 4 through 7). Therefore, the project appeared to help all learners equally, regardless of their baseline learning levels.

[Table 10 about here]

Panel B of Table 10 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 longer statistically significant. In all cases the coefficient on the interaction term is negative. Therefore, we test whether the sum of coefficient on the main effect plus the interaction effect is statistically different from 0. In all three columns, we fail to reject that the program had a 0 effect on female students. In contrast, in columns 4 through 7 for the PEC exam, we reject in all cases except math that the sum of the coefficients for females is 0. Therefore, the program improved PEC scores for females. Unlike the project exams, 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.

Given the higher average scores for girls on the exams (0.39 SD higher for the overall PEC score), the two panels of this table 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.

[17] 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.

We find similar patterns to Panel A within gender (results not presented). Therefore, the test score and gender differences appear to be separate phenomenon.

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. 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). 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 mechanisms that we tested in the previous section, with statistically significant differences by treatment 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 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 to be more likely to visit school.

In Table 11 we vary the specification and the sample, finding results that are similar to our preferred specifications above. Each Panel reports the coefficient of interest from a separate regression with the dependent variable as indicated. Column 1 repeats our estimates from Table 2. Column 2 implements the Belloni, Chernozhukhov, 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. Column 3 limits the sample to only those students for whom we collected surveys. The math and combined effects are similar, but smaller, and the science effect is smaller and no longer statistically significant. Finally in columns 4 and 5 we use Lee (2009) to adjust the attrition and find narrow bounds around our preferred specification.

8 Cost Effectiveness

One reason why technology is potentially promising in low resource settings is the ability to deliver content relatively cheaply. The marginal costs 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.[18]

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). Programs that provided information on earnings in Madagascar (Nguyen 2008) or linked school committees to local governments in Indonesia (Pradhan 2012) were 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 hypoth-

[18]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.

esis, 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 responses by students and teachers who were both more likely to be present in school. Further, students were more likely to approach teachers for help outside of class, 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.

We are not able to specifically test the exact reason for the large effects or why the results are uniform across the baseline score distribution. Part of the overall effectiveness of the intervention could have been due to its relative novelty—about 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. Regarding the uniform nature of the findings, our students are in a higher grade level and potentially already a selected sample. Alternatively, something about the technology or integrating it into the curriculum could have been crucial. Recording 29 hours of video lecture and providing the hardware to display it to students may not be possible in all settings. Nevertheless, we show that integrating a novel approach to grade level material into the existing teaching practice can substantially increase middle school learning for students of all baseline learning levels.

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*.

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*.

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

Barrera-Osorio, F., L. L. Linden, and M. Urquiola (2007). The effects of user fee reductions on enrollment: Evidence from a quasi-experiment. *Manuscript. Department of Economics. The University of Texas at Austin*.

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.

Borkum, E. (2012). Can eliminating school fees in poor districts boost enrollment? evidence from south africa. *Economic Development and Cultural Change 60*(2), 359–398.

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.

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.

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.

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 (2016). Disrupting education? experimental evidence on technology-aided instruction in india. Technical report, National Bureau of Economic Research.

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.

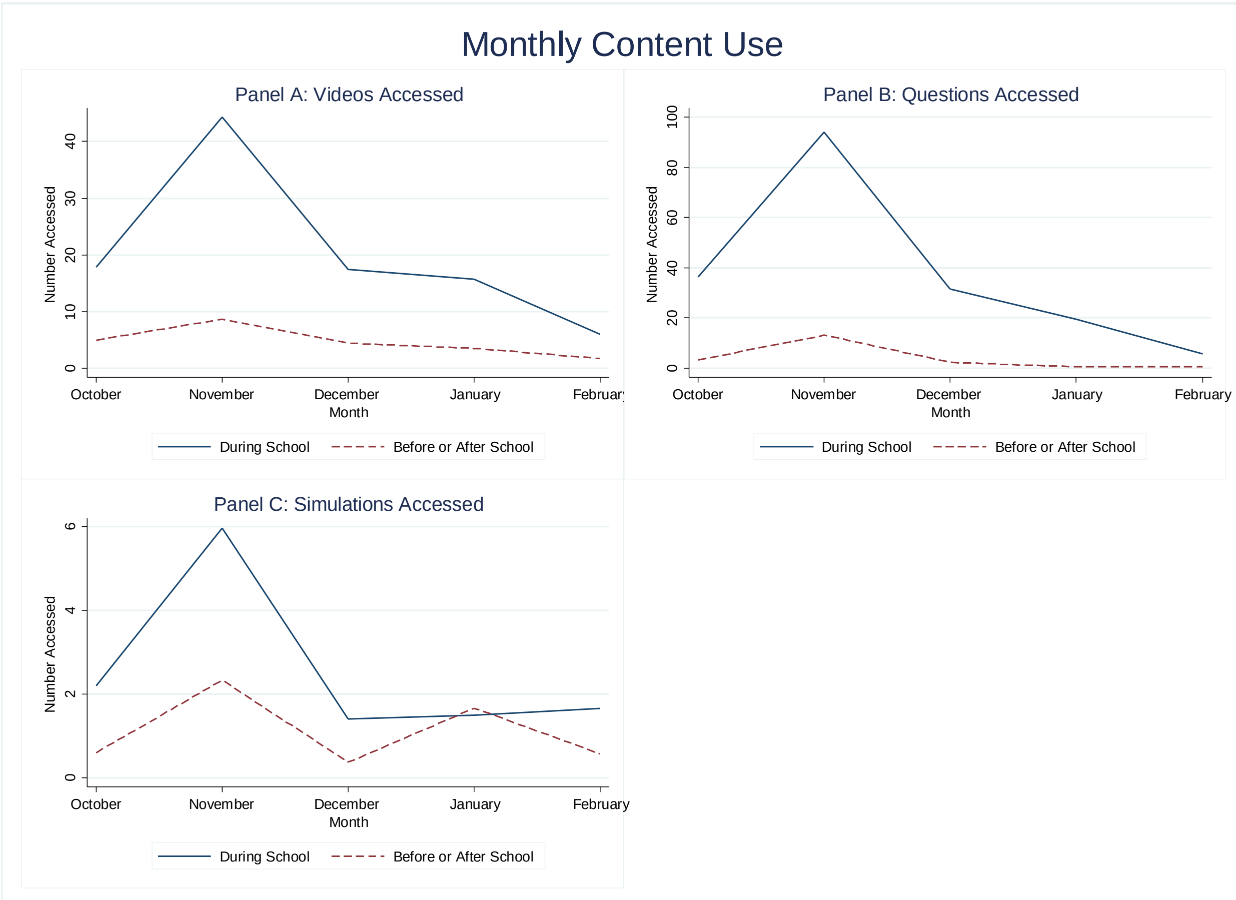
Pop-Eleches, C. and M. Urquiola (2013). Going to a better school: Effects and behavioral responses. *The American Economic Review 103*(4), 1289–1324.

Schultz, T. P. (2004). School subsidies for the poor: evaluating the mexican progresa poverty program. *Journal of development Economics 74*(1), 199–250.

Solis, A. (2011). Credit constraints for higher education.

Wydick, B., P. Glewwe, and L. Rutledge (2013). Does international child sponsorship work? a six-country study of impacts on adult life outcomes. *Journal of Political Economy 121*(2), 393–436.

Figure 1: 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	Control	Difference T-C		Treatment	Control	Difference T-C
	(1)	(2)	(3)		(4)	(5)	(6)
Panel A: Student Characteristics				*Panel B: Teacher Characteristics*			
Math Score	-0.101	0.107	-0.222	Math Teacher Has a College Degree	0.333	0.138	0.195*
	(0.939)	(1.055)	(0.182)		(0.479)	(0.351)	(0.109)
Science Score	0.0230	-0.00802	0.0466	Math Teacher Has Masters Degree	0.500	0.655	-0.155
	(0.981)	(1.013)	(0.165)		(0.509)	(0.484)	(0.129)
Household Computer	0.427	0.407	0.0195	Math Teacher Has a PhD	0.00	0.000	0.00
	(0.495)	(0.492)	(0.0484)		(0)	(0)	(0)
Household Cell Phone	0.999	1.0000	-0.00133	Math Teacher Years of Experience	12.14	14.62	-2.479
	(0.0365)	(0.000)	(0.00131)		(8.807)	(9.666)	(2.410)
Use Technology to Study	0.372	0.292	0.0805	Science Teacher Has a College Degree	0.0667	0.138	-0.0713
	(0.484)	(0.455)	(0.0552)		(0.254)	(0.351)	(0.0799)
Days Absent Last Week	1.495	1.167	0.327*	Science Teacher Has Masters Degree	0.900	0.759	0.141
	(2.409)	(1.788)	(0.173)		(0.305)	(0.435)	(0.0982)
Take Tuitions Outside of School	0.457	0.473	-0.0151	Science Teacher Has a PhD	0.00	0.0345	-0.0345
	(0.499)	(0.500)	(0.0542)		(0)	(0.186)	(0.0345)
Hours Spent on Homework per	1.612	1.527	0.0843	Science Teacher Years of Experience	3.533	3.948	-0.415
	(0.982)	(0.969)	(0.112)		(2.675)	(3.878)	(0.870)
Work	0.0319	0.0237	0.00821	Use Technology to Prepare for Class	0.581	0.596	-0.0158
	(0.176)	(0.152)	(0.0128)		(0.497)	(0.495)	(0.106)
Parents Visit School	0.572	0.603	-0.0312	Use Technology in Class	0.155	0.170	-0.0146
	(0.495)	(0.490)	(0.0647)		(0.365)	(0.379)	(0.0687)
Expect to Attend College	0.616	0.625	-0.00949	Years of Teaching Experience	10.79	10.93	-0.140
	(0.487)	(0.484)	(0.0567)		(8.480)	(9.365)	(1.796)
Number of Siblings	3.949	3.803	0.147	Has an Advanced Degree	0.754	0.789	-0.0354
	(1.827)	(1.765)	(0.157)		(0.434)	(0.411)	(0.0819)
Age	13.90	13.87	0.0266	Part of Any In-Service Training	0.919	0.965	-0.0456
	(1.237)	(1.231)	(0.102)		(0.275)	(0.186)	(0.0475)
Meat Cooked (days per week)	0.738	0.696	0.0417	Holds Private Tutoring Sessions	0.113	0.0702	0.0427
	(0.440)	(0.460)	(0.0393)		(0.319)	(0.258)	(0.0542)
Number of Motorbikes	1.036	0.970	0.0655	Performs other Official Duties	0.613	0.625	-0.0121
	(0.883)	(0.864)	(0.0703)		(0.491)	(0.489)	(0.118)
Number of Cars	0.168	0.133	0.0342	Number of Classes Taught per Month	36.00	32.70	3.298*
	(0.460)	(0.411)	(0.0313)		(7.071)	(9.126)	(1.907)
				Extra Classes Per Month	3.871	2.545	1.326
					(7.623)	(5.350)	(1.448)
				Student Approach for Help	0.0492	0.0526	-0.00345
					(0.218)	(0.225)	(0.0473)
				Minutes per Day Planning Lessons	40.32	33.51	6.814
					(33.69)	(27.93)	(5.548)

Notes : 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	Control	Difference T-C
	(1)	(2)	(3)
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.

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 : 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 : 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 Grade Completion

	Students				Teachers	
	Present at Follow-up		Completed End of Year Exam		Present	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0530**	0.0540**	-0.0181	-0.0169	0.00988*	0.0214**
	(0.0204)	(0.0203)	(0.0135)	(0.0130)	(0.00579)	(0.0101)
Treatment X Baseline Score		0.0115		0.0149**		
		(0.0103)		(0.00653)		
Treatment X Months of Treatment						-0.00598
						(0.00366)
Observations	2,999	2,999	2,999	2,999	274	274
R-Squared	0.02	0.02	0.01	0.01	0.26	0.27
Control Group Mean	0.85		0.93		0.94	

Notes : 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 1-4: Includes all students who took the baseline test. Columns 5 and 6: the percent of all teacher present in the school during an unannounced spot check. Measured monthly.

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	0.349***	0.782***	0.415***
	(0.0706)	(0.0677)	(0.118)
Observations	115	115	115
R-Squared	0.31	0.74	0.35
Control Group Mean	0.60	0.17	3.62

Notes : 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 baseline and follow-up. Linear probability models.

Table 7: Other Changes in 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	10.98*	0.0729	-0.828	1.268	0.174*	0.0270
	(5.926)	(0.0600)	(1.431)	(1.429)	(0.101)	(0.0930)
Observations	115	115	115	115	115	115
R-Squared	0.23	0.27	0.41	0.21	0.12	0.31
Control Group Mean	34.15	0.076	33.19	2.75	0.45	0.63

Notes : 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 baseline and follow-up. Columns 1, 2, 5, and 6: Linear probability models.

Table 8: 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 : Standard errors clustered at the school level appear in parenthesis. 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. Students who completed endline survey. Columns 1, 4-7: Linear probability models.

Table 9: Achievement Effects, Net of Other Teacher Effort Changes

	Standardized Test Score		
	Math	Science	Math + Science
	(1)	(2)	(3)
Treatment	0.273**	0.184	0.286*
	(0.134)	(0.148)	(0.152)
Additional Controls			
Baseline Math Score	0.0872**	0.0842*	0.105**
	(0.0352)	(0.0452)	(0.0445)
Baseline Science Score	0.109***	0.0871***	0.121***
	(0.0231)	(0.0325)	(0.0254)
Observations	2,622	2,622	2,622
R-Squared	0.22	0.25	0.30
Average Control Group Change	0.21	0.49	0.42

Notes : 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 10: 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 : 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 11: Robustness

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

Notes : Standard errors clustered at the school level appear in parenthesis. Column 2: Includes all students who took the test at both baseline and endline. Column3: Sample limited to students who completed the baseline survey. 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. Sample limited to students present at baseline.