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 January 2020

For useful comments and suggestions we thank Natalie Bau, Jim Berry, Rebecca Dizon-Ross, Isaac Mbiti, Emily Oster, Maria Rosales-Rueda, Laura Schechter, L. ChoonWang, Matt White, and seminar participants at Innovations for Poverty Action Ghana, Pakistan Planning Commission, the University of Delaware, the University of Virginia, the Barcelona Graduate School of Economics Summer Forum, the Midwest International Economic Development Conference, and the North East Universities Development Consortium 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 reviewed this project and deemed it exempt. 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 January 2020
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's combined math and science score increased by 0.3 standard deviations on both project and government tests, 59 percent more than the control group. Students were also more likely to pass the provincial high-stakes exams. Increased attendance by both students and teachers indicate technology can increase other inputs. At the 200 school scale, 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

When schooling enrollment increases rapidly, as it has in poor countries since 2000, schooling systems often lack resources to adequately educate the influx of new learners (World Bank 2017). 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). Learning outcomes in many low income countries are constrained by gaps in teacher skill, content knowledge, and effort. When teaching advanced material in middle and high school, teachers themselves need to be knowledgeable and trained both in the subject matter and how to convey their knowledge to students.[1] This lack of capacity at higher levels of schooling exacerbates learning deficiencies that start in primary school and are especially acute in technical subjects such as science and math (Banerjee et al. 2013). Specifically in our middle-school context in Pakistan, 51 percent of school principals reported that insufficient teacher qualifications were a barrier to student learning. Even when teachers are academically qualified, they may be absent from school or the classroom, and/or unable to convey their knowledge to students.[2] Because the stock of teachers is relatively fixed, low income countries need innovative solutions to fill the teacher capacity and effort gaps, improving student achievement without substantial teacher re-training.

Our intervention leverages technology through a government implemented, scalable, relatively inexpensive program to address learning deficiencies and circumvent limitations on teacher knowledge or ability to explain complex ideas. Specifically in this study, we use a randomized controlled trial (RCT) to test the impact of eLearn, a program that delivers curriculum-based math and science content through short videos to grade 8 students in Punjab, Pakistan. Since this change in available inputs could be either encourage or discourage

[1] Averaging across 7 African countries, Bold et al. (2019) attribute 30 percent of students' knowledge gap relative to the curriculum to teachers' lack of content knowledge.

[2] In Punjab, Pakistan, the location of the current study, teachers self-reported being absent 3.2 days per month (Bau and Das forthcoming). In other low income settings, studies have observed absenteeism ranging from 19 to 35 percent with teachers engaged in teaching on average only half the time they were physically present in India and Ghana (Chaudhury et al. 2006; Duflo, Hanna, and Ryan 2012; Duflo, Kissel, and Lucas 2019).

other forms of effort, in addition to the reduced form effect on test scores, we test the effect of the program on additional inputs into the educational production function.

Fundamentally, our eLearn intervention was designed to teach students grade level material in a way that was both accessible to students and engaged the existing teachers, who might have had limited subject knowledge or ability to explain material clearly. The intervention was designed to enhance, not supplant, in-person teaching. In treatment schools the bulk of the eLearn intervention was 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. These components addressed potential deficiencies in teacher capacity, teaching both the teacher and the student the material, and exposed students to an alternative teacher in the video whose explanation might be clearer or more engaging. The tablets could also be used to record student attendance. Teachers received minimal training over two days: one day on how to use the multimedia content and one day on how to incorporate it into a more effective, blended teaching practice. The intervention did not provide scripted lessons nor were the video lessons designed to be total substitutes for the teachers. Outside of the teacher training, all aspects of the intervention occurred during the school day.[3] 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.

In our preferred specification, the intervention increased student achievement by 0.29

[3]The intervention was designed to have an additional at home component that would provide students with interactive short message service (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.

standard deviations (SD) on a combined math and science exam designed for the project. This gain occurred despite students receiving only 4 months of exposure and were 59 percent more than the control group score change over the same period. This improvement is approximately equivalent to increasing the value-added of a teacher by 1.9 SD in Punjab, Pakistan (Bau and Das forthcoming). 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 0.26 standard deviations higher on the combined math and science sections of the standardized test that students take at the end of grade 8. When we combine the two scores in a single measure of student achievement, our intervention improved test scores 0.29SD. Further, the intervention increased the likelihood that students passed the standardized grade 8 test by 5 percentage points. Passing this examination determines what options are available for further study and acts as a proxy for longer run outcomes.[4]

To better understand the mechanisms behind the test score improvements, we test for effects on both observed and self-reported effort. Both student and teacher attendance increased.[5] 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.[6] Therefore, grade level content presented in a novel way at the middle school level does not necessarily exacerbate pre-existing achievement heterogeneity. When we test for heterogeneous effects by gender, we find some evidence that girls' scores increased more than boys on the provincially

[4]Because of funding constraints our sample is not as large as we would have liked and our standard errors are larger than ideal. To maximize power in our estimates we use the machine learning Belloni, Chernozhukov, and Hansen (2014) post double Least Absolute Shrinkage and Selection Operator (LASSO) approach to specify the optimal controls in our preferred specification. In this specification, all point estimates mentioned in the previous paragraph are statistically significant at the 5 percent level. When we limit our controls to baseline test scores, our estimates are less precise. The impact of the program on test scores are statistically significant at the 5 or 10 percent level and the likelihood of passing the standardized test is no longer statistically significant.

[5]Our test score findings are robust to attrition correction. See Section 6 for further details.

[6]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.

standardized exams.[7]

Our findings that middle school grade-level achievement can be increased while working within the school day with existing teachers contributes to three related literatures: using educational technology to improve student learning, the importance of schooling inputs in education production, and improving education beyond foundational literacy and numeracy.

First, we build on the existing literature on educational technology by using a different type of technology and targeting the content knowledge of older students. 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). The most common technology intervention in developing countries has been computers and software and occurred almost exclusively in primary schools. 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.[8] Other technology models in developing country primary or pre-primary schools provided entire lessons that replaced classroom teachers as content providers, a much heavier touch intervention than eLearn and targeting younger students.[9] We show limited technology targeted at teacher knowledge capacity that is used during the school day can

[7]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 in Section 7.

[8]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.

[9]See Mayo et al. (1975) that compared outcomes between non-randomized groups of students who used 10 hours per week of television-based (*Telesecundaria*) versus traditional secondary schools in Mexico; Jamison et al. (1981) that evaluated 150 daily 30 minute radio broadcasts of lessons of first grade math content in Nicaragua and the accompanying teacher's guide and student worksheets; Naslund-Hadley, Parker, and Hernandez-Agramonte (2014) that proved 72 hours of audio CDs in Paraguay; Johnston and Ksoll 2017 that provided primary schools in Ghana with "solar panels, a satellite modem, a projector, a web cam, microphones, and a computer with interactive software" to interact with the two studios in Accra, Ghana that were set up to broadcast live math and English lessons; and Navarro-Sola 2019 that used the same *Telesecundaria* intervention to focus on school completion.

improve math and science achievement in middle school.[10] Our model is lighter touch than many programs that required multiple hours per week of student time either during or after school. It is also cheaper to scale than other interventions with similar effect 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 $20/student at the scale of 200 schools.[11]

Second, we expand the literature on specific school-based inputs to improve educational outcomes to include grade-level, subject-specific content. Research 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 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. Our model of providing expert content caused especially large achievement gains for only 4 months of exposure.

Third, the first two literatures are almost exclusively focused on primary or pre-primary education. Effective and inexpensive interventions at the post-primary level 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 (e.g. Jackson 2010; Pop-Eleches and Urquiola 2013; Lucas and Mbiti 2014; Navarro-Sola 2019) 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

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

[11] 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. Excluding these components, the cost would be $12/student at the 200 school scale. See the Cost Effectiveness section for more details.

about middle school content since most of the study pupils were well behind grade level. Particularly salient to students in higher grade levels, 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 students' performance on government tests.

Finally, we show that all of this was possible though an intervention that was designed by and implemented with the provincial government of Pakistan, not a non-governmental organization, ensuring a program directly addressing issues the government found pressing and increasing the possibility of scale-up of the program.

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, our focus, 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, potentially because of scant teaching resources and a lack of qualified teachers (Andrabi et al. 2007; Andrabi et al.

2013). 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 trained in how to teach the subject. Additional instructional tools and training teachers on incorporating them into lessons could have a substantial impact.

3 The eLearn Intervention

Our intervention, eLearn, was designed to improve student learning by combining technology with existing teachers. The small-scale implementation of the program that we evaluate was designed to inform the larger scale-up of the program that started in 2018.[12] 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 and were organized within

[12]This intervention was designed to mimic exactly the planned full roll-out. Teachers were not provided additional support nor did they have additional interactions with the development team. Some of the elements of the full intervention were not operational during this evaluation. We discuss those in detail below.

unit folders.[13] A single presenter, in all cases a government employee, appeared in all videos related to a particular unit.[14] Men were the subject experts for 21 of the 22 units.[15] The total content was about 29 hours. These lectures contained spoken Urdu with an occasional English word and all words were 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.[16] An additional component of the intervention was designed to engage students and parents at home, but was at most marginally implemented.[17]

[13] The math content was a total of 12.4 hours divided among 77 videos with an average length 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 with an average length of 8 minutes. Each of the 12 units of the official science curriculum had at least 23 minutes of content and 2 videos. The videos were designed to cover all topics of the curriculum. Videos were organized on the tablet by unit and teachers could select which, if any, videos to show on a particular unit.

[14] The presenters were experienced teachers, former teachers, university professors, and government officials working in the education sector.

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

[16] This technology was likely novel to some of the students in the sample. At our baseline, 40 percent of students reported having a computer at home and 30 percent reported using some sort of technology as a study aid at home.

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

Teachers received a two day in-service training session primarily focused on program implementation–one day on orientation to the new technologies and one day on how to combine classroom 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.

This model of technology implementation in which content is conveyed, review questions are provided, and a teacher or tutor is trained on how to implement the program is a bundle, one that is common with education technology (e.g. Banerjee et al. 2007; Muralidharan, Singh, and Ganimian 2019). Models that only provided technology without integrating it into the broader curriculum have been shown to be unsuccessful at increasing student achievement (e.g. Linden 2008, Barerra-Osorio and Linden 2009, and Bai et al. 2016 in China) as have models that provided other resources without corresponding teacher training (e.g. Glewwe et al. 2004, Glewwe, Kremer, and Moulin 2009, Banerjee et al. 2017). Further, inservice teacher training models that increased student achievement included material provision (e.g. Lucas et al. 2014; Kerwin and Thornton 2018). Therefore, the combination of materials, in this case technology, and training teachers how to use them is the meaningful bundle to evaluate to improve student achievement.

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

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

follow-up testing.[18] Figure 1 displays the study and academic year timeline. See Section 5.2 for additional details on data collection.

[Figure 1 about here]

4 Empirical Strategy

The primary conceptual difficulty in assessing the effects of various inputs into the education production function is 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, α is the constant term, $treatment_s$ is an indicator variable equal to one if the school was an eLearn treatment school, X_{is} are a vector of school and 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 we include strata (school gender by district) dummy variables in the X_{is} vector. In specifications in which the outcome of interest is a test score, we implement a lagged dependent variable model and include the student's subject scores from the baseline as a control in the X_{is} vector. In addition to a parsimonious specification, because of slight baseline imbalance and to improve precision given our sample size and high intracluster correlation, we imple-

[18] As originally designed, students were to be tested prior to the June to August holidays and training and installation should have occurred during the holiday break. Once implementation delays were apparent, we delayed the baseline as well.

ment 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.[19]

Our 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 while testing 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.[20] One goal of the program was to prepare students for future study. Passing the PEC exam is one measure of this readiness, and we test whether the intervention increased this likelihood. 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 crowded out additional inputs, e.g. students spend less time studying in reaction to additional material being delivered at school, or encouraged 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 the student expects to attend college.[21]

Additionally, based on data collected from teachers, we estimate a similar model, allowing

[19]The variables we consider are listed in the Appendix.

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

[21]A student being present is also our measure of attrition. Our findings are robust to attrition correction. See Section 6. Self-reported absenteeism is an imperfect measure of actual attendance as students might mis-report and those present are a selected sample. This should be symmetric across treatment statuses.

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 on objectively observed effort.

5 Sample Selection and Data

5.1 Sample Selection and Randomization

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.[22] 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.

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.

From eligible schools, we selected 60 schools, an equal number of boys' and girls' schools,

[22] 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.

for the sample.[23] 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.

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 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 the technology was operational in all treatment schools by the beginning of October.[24] 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.[25] 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.[26] Head teachers and grade 8 subject

[23]We would have liked to have a larger sample but were only able to raise enough money for a 60 school sample. Implementation costs were covered through a grant directly to the government that did not cover evaluation costs. Within school randomization was deemed infeasible by our government partners. As enumeration costs were the binding constraint, we could not add additional control schools. When we returned to funding agencies to fund a larger follow-up, they determined that the evidence from this study rendered a larger study unnecessary and would not fund it.

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

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

[26]Students took the same exam at both baseline and follow-up. During baseline administration, enumerators invigilated the exams without teachers present and were careful not to leave any materials behind for teachers to see nor let teachers know the content of the exams. Students in both the control and treatment groups would have had the same level of familiarity with the exam at the follow-up.

teachers were again surveyed. Figure 1 above provides a school calendar and study timeline.

We first measure the effect of eLearn on student scores on the exam we administered during the follow-up visit. When estimating the effects on test scores, we use item response theory (IRT) to convert raw science and math test responses to approximated latent student ability, and standardize based on the baseline mean and standard deviation.[27] Our findings are similar using raw test scores.

The second measure of student achievement are administrative student by subject level exam results from the Punjab Examination Commission.[28] Students completed the PEC exams in mid-February. These data were merged to the students in our sample using students' and fathers' names.[29] Because we do not have item level responses, these scores are simply scaled with a mean 0 and standard deviation of 1.

As a third measure of achievement we use the first component from a principal component analysis of a student's project and PEC scores, standardized by the control group mean and standard deviation.

The two 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, criticisms 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.

As a final measure of achievement and prospects for future study, we estimate the effect on the likelihood that a student passes the PEC exam, a requirement for future study. This

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

[28]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 between test scores.

[29]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. See Section 6.2. 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.

outcome is from the Punjab Examination Council.

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 presence but not whether they were engaged with students.[30] These data were available at the school level only. Therefore, they measure the percentage of all teachers in the school present during the visit.

Table 1 displays means and standard deviations of student (Panel A), teacher (Panel B), and school (Panel C) characteristics across the treatment and control schools. Almost all of the measures are statistically indistinguishable by treatment status with three exceptions at the 10 percent level: treatment students report being absent more often in the previous month by 0.3 days, the math teachers in treatment schools are 20 percentage points more likely to have exactly a college degree, 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).[31] Given we are testing 23 outcomes, some small imbalances are expected. To ensure we are not attributing baseline imbalance to the treatment effect, our preferred specification uses LASSO to determine optimal controls.

[Table 1 about here]

[30]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.

[31]Even though most of the teachers have at least a college degree, this does not guarantee teacher content knowledge nor the ability to explain content. In Punjab, Bau and Das (forthcoming) found that a teacher having a college degree is associated with only a 0.2 standard deviation increase in a teacher's test score on a upper primary school exam. They estimate that including all available teacher characteristics (including training and experience) explains at most 9 percent of the variation in teacher content knowledge.

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.[32]

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 2. Panel A is a parsimonious specification that includes only the strata and baseline test scores as control variables. The treatment increased achievement by 0.26 standard deviations (SD) on the project test (column 1). Our exams were designed to test the content from 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 Column 2 we test the effect of treatment on the math and science portions of the standardized government PEC tests. The treatment increased this PEC score by 0.22 SD. In Column 3, we combine these tests into a single score measure. The treatment increased the combined test score by 0.27 SD. In Column 4 we estimate the effect of the intervention on the likelihood that a student passed the PEC exam, and find a positive, statistically insignificant effect.[33]

[Table 2 about here]

Given our small sample size of 59 schools and slight imbalances from Table 1, we include additional student, school, and teacher control variables in Panel B as determined by the

[32] 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 Section 7.

[33] The sample size changes across the columns as not all students took both tests. The results in Columns 1 and 2 are similar when the sample is limited to students who took both exams.

LASSO machine learning procedure 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 test scores on the project exam by 0.29 SD (column 1), had a 0.26 SD effect on the PEC score (column 2), and increased the combined score by 0.29 SD (column 3), sizable effects for a 4 month treatment. Further, the effect on the likelihood of passing the PEC remains positive and is now a statistically significant 5 percentage points.[34] [35]

During this same period control group students increased their project test scores by 0.49 standard deviations.[36] Therefore, based on Panel B, Column 1, this intervention increased the project test score by 59 percent relative to the gains in the control group (52 percent based on the estimate in Panel A).

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. While an interesting outcome itself, it is also a measure of respondent attrition. We use an indicator variable equal to 1

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

[35]Appendix Table A1 contains the subject-specific exam effects for both the project and PEC exams and the effect on the non-math and non-science subject scores of the PEC. 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 new, exciting teaching methodology, reducing time on other subjects. We find a smaller, statistically significant effect on the other test scores.

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

if the student was present at follow-up as the outcome, y_{is}, in Equation 1. As with all our other binary outcomes, we use a linear probability model. The results appear in Columns 1 and 2 of Table 3.

[Table 3 about here]

Students in the treatment group were about 4 percentage points more likely to be present at follow-up (column 1). Relative to the control group mean of 85 percent, this is about a 5 percent 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 Appendix Table A2 we follow Lee (2009) and provide treatment bounds. Our test score findings are robust to this attrition adjustment. Columns 3 and 4 perform a similar analysis to test for differential attrition for the PEC scores, finding no effect of treatment on the likelihood that we could match students into the PEC sample with a point estimate of about -0.01 percentage points.[37]

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 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 1 but with

[37] We are able to match about 93 percent of our baseline sample to a PEC score. The remaining 7 percent include students who did not take the PEC at their baseline school, whether because they changed schools or dropped out, and those who took the PEC but we could not match.

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 5). 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 6 we test whether this response changed over time. The point estimate on the interaction between treatment and months of treatment is negative. Therefore, the intervention appears to have increased teacher effort, but this effort might have diminished over time.[38]

6.3 Technology Use

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 reported 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). At an average video length of just over 9 minutes, this is implies just over 11 hours of content. 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, teachers and students likely viewed this content together. The early peak in use in November and later decline is consistent with the evidence from teacher attendance that attendance increased the most in November. Even at its lowest point in February the average school was still accessing some content. Recall that

[38] Based on the point estimates, the portion present would revert to the non-treatment level after 4.2 months.

students took the PEC exam in mid-February, therefore teaching time was both interrupted and structured differently in that month.[39] 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 was an important part of the intervention.

[Figure 2 about here]

6.4 Self-Reported Outcomes

We additionally collected data on self-reported changes in take-up and inputs. We first test whether the change in inputs increased the use of technology available to teachers. Table 4 contains these results. Teachers were 31 percentage points more likely to report that they used technology to prepare for lessons (column 1) and 79 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 in treatment schools reported having attending more trainings relative to teachers on control schools. Column 3 shows that treatment teachers attended 0.29 more in-service teacher training events during the school year.

[Table 4 about here]

We tested for additional changes in teacher effort that might have occurred as a result of the intervention finding no statistically significant changes in the minutes per day spent

[39] 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.

planning lessons, the number of regularly scheduled classes taught per week, the number of extra classes per month during the school day, or whether students approached them outside of class during the school day for extra help in math or science. Teachers increased the likelihood they held private tutoring sessions (significant at 10 percent). Full point values appears in Appendix Table A3. Therefore, teachers in treatment schools increased their use of technology and their observed effort (attendance), but at most marginally 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 (results not presented).

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

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. We test for this possibility by including an interaction term between treatment and baseline test score as an additional regressor in Equation 1 using the project test, the PEC, and the combined project and PEC score as three separate outcomes.[40] Across all three outcomes, the coefficients on the interaction terms are statistically insignificant (see Panel A of Appendix Table A4).

We further test for heterogeneity by school gender, replacing the test score interaction

[40] 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.

with one for treatment times female school. Recall that all schools were 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 from attending an all female school. For the project test, the main effect is positive, but the interaction effect is negative and we fail to reject that the coefficients sum to 0 (Panel B of Appendix Table A4). In contrast, on both the PEC and the combined project and PEC score the interaction is positive, and we reject (at the 1 percent level for the PEC and 5 percent level for the combined) that the coefficients sum to 0. Therefore, while the intervention was not designed to favor students of a particular gender, instead providing expert content to assist all students, female students' scores increased on the PEC exam but males students' scores did not.[41]

As the heterogeneous score gains by gender could be the result of both student and school characteristics that differ by gender, we test for differences by gender between both schools and students in the baseline data. At the school level, female and male schools and teachers at those schools are statistically indistinguishable except female schools have a higher percentage of female teachers and higher average baseline test scores.[42]

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

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 status and gender only for the likelihood that parents have visited the school and university

[41] 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.

[42] The finding of no heterogeneity by baseline test score holds within gender.

aspirations. Therefore, any differences by gender appear to be something about the interaction between the program and the students and not about the level of implementation or other effort changes.

While the two differences in other outcomes we find between genders are likely not causing any gender heterogeneity in achievement, they are of note. The treatment increased the likelihood that male students reported that their parents 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.

8 Cost Effectiveness

One reason why technology is potentially promising in low resource settings is its ability to deliver content relatively cheaply. Once the fixed costs of development are paid, the marginal costs of an additional school are quite low. 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 school in our sample had 63 students on the official grade 8 roster. Using only the marginal costs, adding an additional school, assuming schools the same size as our study, would be US$9/student.[43] Larger schools would 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 our study. For this study, 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 would increase the cost-effectiveness substantially. A 100 school intervention would have an average cost of $31/student, a 200 school intervention would have an average cost of $20/student, and a 1000 school intervention would have an average cost of $11/student.[44]

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 modest 200 school scale, for $100 our effective size would be 1.4 SD in the combined math and science score, increasing to 2.6 SD at the 1000 school scale. The cost effectiveness at 200 schools exceeds the cost-effectiveness of the other technology interventions reported in Kremer, Brannen, and Glennerster (2013) and at 1000 schools it exceeds the cost-effectiveness of Muralidharan, Singh, and Ganimian (2019).[45] A program that linked school committees

[43]For ease of comparison across studies, we use the ingredients method of cost effectiveness.

[44]Removing the costs of the only partially implemented interactive content puts the costs at $15/student at 100 schools, $12/student at 200 schools, and $10/student at 1000 schools. In our setting, boundary walls and electricity were standard. Upgrading schools to include this infrastructure would increase the costs, but could also confer additional benefits.

[45]Muralidharan, Singh, and Ganimian (2019) do not provide a 200 school cost effectiveness. They found a 0.25 SD effect on math scores per $100 at their evaluated scale and project 0.93 SD per $100 at 50 schools and 1.85 SD per $100 at 1000 schools.

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 capacity and 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, and teacher in-service training. We found that the intervention increased achievement on both the project-specific and provincially standardized math and science tests by about 0.3 standard deviations with under 4 months of exposure. Prior to this study very little was known about improving grade-level content delivery 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, 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, at a mere 200 school scale the cost effectiveness of eLearn would be on par with some of the most cost-effective technology RCTs and beyond 1000 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., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017). From proof of concept to scalable policies: challenges and solutions, with an application. *Journal of Economic Perspectives 31*(4), 73–102.

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 (2019). The use and misuse of computers in education: evidence from a randomized experiment in colombia. Working Paper 4836, World Bank.

Bau, N. and J. Das (forthcoming). Teacher value-added in a low-income country. *American Economic Journal: Policy*.

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.

Bold, T., D. P. Filmer, E. Molina, and J. Svensson (2019). The lost human capital: Teacher knowledge and student achievement in africa. The World Bank Policy Research Working Paper 8849.

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.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers (2006). Missing in action: teacher and health worker absence in developing countries. *Journal of Economic perspectives 20*(1), 91–116.

Duflo, A., J. Kiessel, and A. Lucas (2019). Alternative models of increasing student achievement: Evidence from a nationwide randomized experiment. Working paper.

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.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives work: Getting teachers to come to school. *American Economic Review 102*(4), 1241–78.

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.

Jamison, D. T., B. Searle, K. Galda, and S. P. Heyneman (1981). Improving elementary mathematics education in nicaragua: An experimental study of the impact of textbooks and radio on achievement. *Journal of Educational psychology 73*(4), 556.

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.

Mayo, J. K., E. G. McAnany, and S. J. Klees (1975). The mexican telesecundaria: A cost-effectiveness analysis. *Instructional Science 4*(3), 193–236.

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).

Navarro-Sola, L. (2019). Secondary school expansion through televised lessons: The labor market returns of the mexican telesecundaria. Working paper.

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

The set of potential controls in the LASSO specifications includes total school enrollment, number of grade 8 sections, grade 8 class size, total grade 8 students present at baseline; head teacher, math, and science teacher tenure, experience, and schooling level; indicator variables for whether the school had a library, computer lab, water connection, electricity, latrine, security guard, cement walls, playground, and posters or charts; science and math teacher self-reported time spent preparing lessons, whether they held private tuitions, their total classes taught, their extra classes taught, if they had non-teaching or substitute teacher duties, their time spent in these activities, whether they used technology use in the classroom, whether they used technology at home, whether they used teaching guides, whether they used the Taleemi calendar in teaching, the frequency of students approaching them outside class, their mode of transport, their commute time, if they were contract (versus permanent) teachers, their contract length, their pay scale, and dummies for problems they identified for classroom teaching and learning; student age, household size, number of siblings, parents level of education and occupation, expected schooling, self-reported attendance, and indicator variables for type of transportation to school; whether a student received extra tuitions, completed homework, was assigned readings; dummy variables for different technologies at home; whether parents visited the school and were employed; and the dummy variables for household asset ownership. Further, the squared values of all continuous variables are included.

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

[Appendix Table A1 about here]

[Appendix Table A2 about here]

[Appendix Table A3 about here]

[Appendix Table A4 about here]

Figure 1: Study and Academic Year Timeline

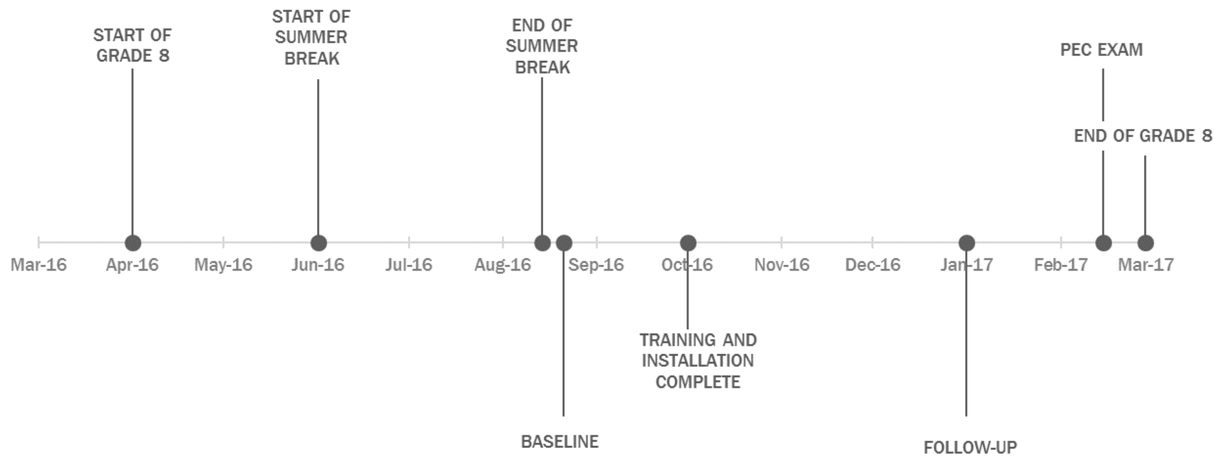
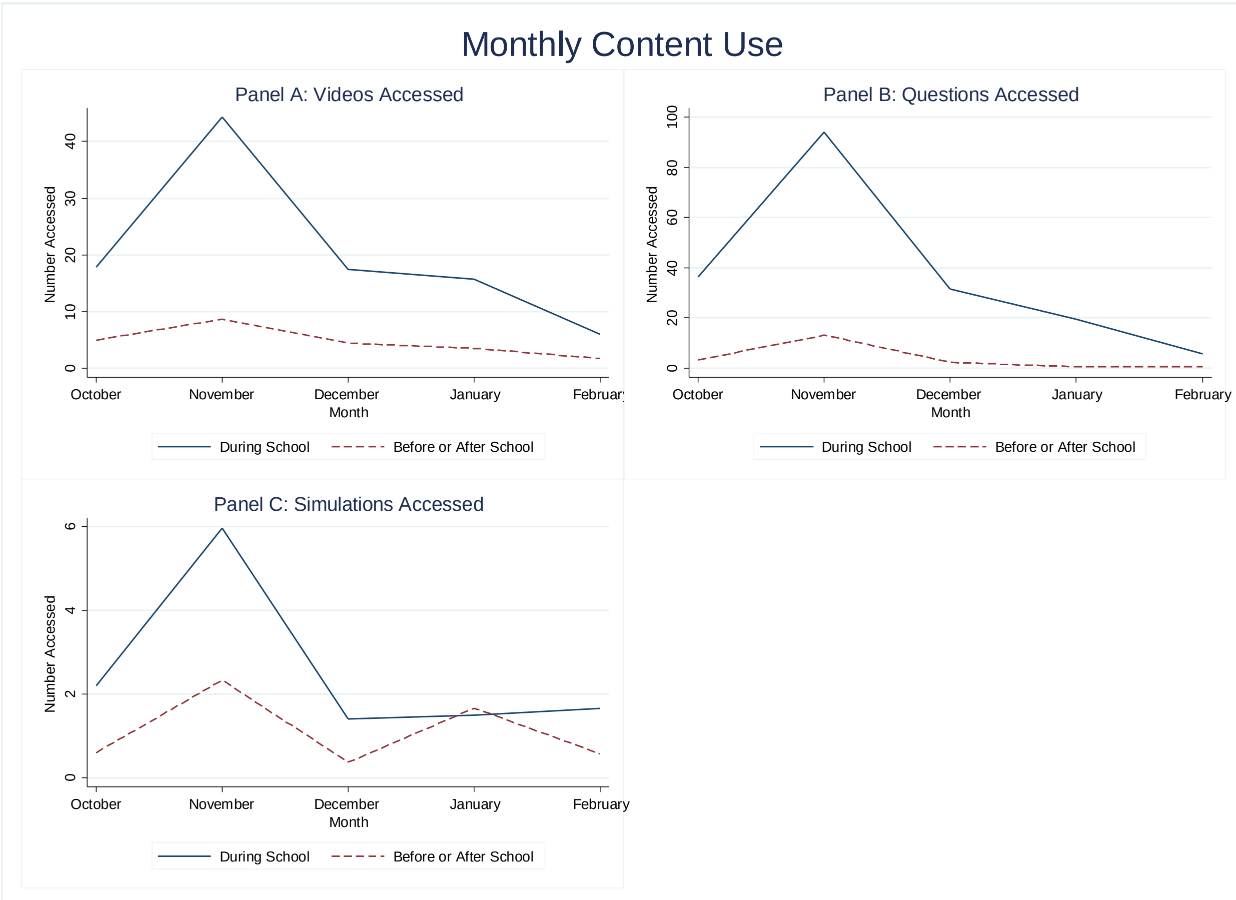


Figure 2: Monthly Content Use by Teachers



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

Table 1: Summary Statistics

	Treatment (1)	Control (2)	Difference T-C (3)
Panel A: Student Characteristics			
Combined Math and Science Score	-0.056 (0.99)	0.068 (1.01)	-0.12 (0.190)
Age	13.90 (1.24)	13.87 (1.23)	0.03 (0.102)
Days Absent Last Month	1.495 (2.41)	1.167 (1.79)	0.327* (0.173)
Has a Computer at Home	0.43 (0.50)	0.41 (0.49)	0.02 (0.0484)
Mother Has No Formal Schooling	0.35 (0.48)	0.33 (0.47)	0.02 (0.0576)
Father Has No Formal Schooling	0.17 (0.38)	0.20 (0.40)	-0.02 (0.0341)
Panel B: Teacher Characteristics			
Has an Advanced Degree	0.75 (0.44)	0.80 (0.40)	-0.05 (0.0806)
Years of Teaching Experience	10.72 (8.52)	10.67 (9.10)	0.05 (1.727)
Minutes per Day Planning Lessons	40.67 (33.75)	33.64 (27.85)	7.03 (5.477)
Use Technology to Prepare for Class	0.58 (0.50)	0.60 (0.49)	-0.02 (0.106)
Use Technology in Class	0.14 (0.35)	0.17 (0.38)	-0.03 (0.0650)

	Treatment (4)	Control (5)	Difference T-C (6)
Panel C: School Characteristics			
Total Enrollment in Grade 8	63.10 (16.36)	63.21 (13.52)	-0.11 (3.901)
Grade 8 Sections	1.40 (0.50)	1.35 (0.48)	0.06 (0.128)
Grade 8 Students Present	49.10 (15.33)	50.31 (11.59)	-1.21 (3.531)
School Has a Computer Lab	0.90 (0.31)	1.00 (0)	-0.100* (0.0557)
Math Teacher Has a College Degree	0.33 (0.48)	0.14 (0.35)	0.195* (0.109)
Math Teacher Has Masters Degree	0.50 (0.51)	0.66 (0.48)	-0.16 (0.129)
Math Teacher Has a PhD	0.00 0.00	0.00 0.00	0.00 (0)
Math Teacher Years of Experience	12.14 (8.81)	14.62 (9.67)	-2.48 (2.410)
Science Teacher Has a College Degree	0.07 (0.25)	0.14 (0.35)	-0.07 (0.0799)
Science Teacher Has Masters Degree	0.90 (0.31)	0.76 (0.44)	0.14 (0.0982)
Science Teacher Has a PhD	0.00 0.00	0.03 (0.19)	-0.03 (0.0345)
Science Teacher Years of Experience	9.40 (7.54)	7.78 (6.33)	1.61 (1.809)

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: Achievement Effects

	Standardized Combined Math and Science Test Score			
	Project	PEC	Combined Project and PEC	Pass the PEC
	(1)	(2)	(3)	(4)
Panel A: Limited Controls				
Treatment	0.256*	0.221*	0.269**	0.0384
	(0.135)	(0.129)	(0.119)	(0.0277)
Observations	2,622	2,766	2,463	2,766
R-Squared	0.13	0.25	0.20	0.06
Panel B: LASSO Controls				
Treatment	0.292**	0.258**	0.286**	0.0475**
	(0.135)	(0.103)	(0.133)	(0.0228)
Observations	2,622	2,766	2,463	2,766
Average Control Group Change or Mean	0.49	0.00	0.00	0.92

Notes : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Includes all students who took a baseline test and the test at the top of the column. Panel A: Controls are strata and baseline test scores. As students take the PEC only once, previous year's school level PEC included in Columns 2 and 3. Panel B: Additional controls selected by LASSO method. Column 1: Project exams. Control group change in the final row. Columns 2-4: Control group mean in the final row. Column 2: Punjab Examination Council high stakes test. Column 3: PCA of project exam and PEC score. Column 4: Linear probability model.

Table 3: Attendance and Attrition

	Students:				Teachers:	
	Present at Follow-up (Took follow-up exam)		Matched to and Completed PEC		Portion Present	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0427**	0.0488**	-0.0063	-0.0059	0.0114*	0.0221**
	(0.0194)	(0.0218)	(0.0112)	(0.0107)	(0.00671)	(0.0100)
Treatment X Baseline Score		0.00785		0.0117		
		(0.0158)		(0.0113)		
Treatment X Months of Treatment						-0.00555*
						(0.00336)
Observations	2,999	2,999	2,999	2,999	274	274
Control Group Mean	0.85		0.93		0.94	

Notes : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Additional controls determined by LASSO. Columns 1-4: Linear probability models. Includes all students who took the baseline test. Columns 5-6: The portion of all teachers present in the school during an unannounced spot check. Measured monthly for each school.

Table 4: 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	0.302***	0.787***	0.285**
	(0.0767)	(0.0657)	(0.137)
Observations	115	115	115
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 determined by LASSO. Columns 1 and 2: Linear probability models.

Appendix Table A1: Subject Specific Achievement Effects

	Standardized Test Score				
	Project Exams		PEC Exams		
	Math	Science	Math	Science	All Other Subjects
	(1)	(2)	(3)	(4)	(5)
Treatment	0.196*	0.283*	0.166	0.148	0.0555
	(0.106)	(0.151)	(0.117)	(0.0958)	(0.100)
Observations	2,622	2,622	2,766	2,766	2,766
Average Control Group Change	0.18	0.53			

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

Appendix Table A2: Lee (2009) Bounds

	Standardized Combined Project Math and Science Test Score	
	(1)	(2)
Treatment	0.262**	0.304**
	(0.126)	(0.135)
Observations	2,551	2,551

Notes : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Sample size adjusted for attrition following Lee (2009). Controls determined by LASSO.

Appendix Table A3: 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
	(1)	(2)	(3)	(4)	(5)
Treatment	7.743 (6.182)	0.109* (0.0627)	0.135 (1.367)	1.705 (1.385)	0.0788 (0.0677)
Observations	115	115	115	115	115
Control Group Mean	34.2	0.08	33.2	2.7	0.45

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

Appendix Table A4: Heterogeneous Achievement Effects

	Standardized Combined Math and Science Test Score		
	Project	PEC	Combined Project and PEC
	(1)	(2)	(3)
Panel A: By Baseline Test Score			
Treatment	0.266*	0.260**	0.193**
	(0.138)	(0.103)	(0.0971)
Treatment X Baseline Score	-0.0919	0.0215	-0.0153
	(0.102)	(0.0636)	(0.0557)
Observations	2,622	2,766	2,463
Panel B: By School Gender			
Treatment	0.329	0.202	0.208
	(0.203)	(0.198)	(0.147)
Treatment X Female School	-0.0793	0.0892	0.0449
	(0.251)	(0.218)	(0.172)
Observations	2,622	2,766	2,463
F-test of coefficients on Treatment + Treament X Female School=0			
p-value	0.12	0.006	0.03
Average Control Group Change	0.49		

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