NBER WORKING PAPER SERIES

ENGAGING TEACHERS WITH TECHNOLOGY INCREASED ACHIEVEMENT, BYPASSING TEACHERS DID NOT

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

For exceptional research assistance we thank Samantha Sweeney and Attique ur Rehman. For useful comments and suggestions we thank Natalie Bau, Jim Berry, Janet Currie, 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, the 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. This paper combines results from two papers that were previously circulated as “Beyond the Basics: Improving Post-Primary Content Delivery through Classroom Technology” and “Screen Time: Tablets with Interactive Textbooks Did Not Improve Learning.” The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.putty

Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not
Sabrin A. Beg, Adrienne M. Lucas, Waqas Halim, and Umar Saif
NBER Working Paper No. 25704
March 2019, Revised in October 2020
JEL No. C93,I21,I25,I28,O15

ABSTRACT

Using two RCTs in middle schools in Pakistan, we show brief, expert-led, curriculum based videos integrated into the classroom experience improved teaching effectiveness–student test scores in math and science increased by 0.3 standard deviations, 60% more than the control group, after 4 months of exposure. Students and teachers increased their attendance, and students were more likely to pass the government high-stakes exams. In contrast, similar content when provided to students on personal tablets decreased student scores by 0.4SD. The contrast between the two effects shows the importance of engaging existing teachers and the potential for technology to do so.

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

Improving student content knowledge and success on high-stakes exams while working within existing government systems and personnel is a global challenge. Variability in teacher capacity and performance can be a substantial barrier, leading to inadequate learning even for those who are enrolled in school (Andrabi et al. 2007; Muralidharan 2013; Jackson, Rockoff, and Staiger 2014; Bold et al. 2019). Despite hundreds of billions of dollars being spent annually on teacher compensation and teachers' strong impact on student learning, very little is known about how to increase teacher effectiveness (Jackson, Rockoff, and Staiger 2014; Jackson and Makarin 2018; Papay et al. 2020). In response to weak incentive structures, frequent teacher absenteeism, and a legacy of teacher-centered lecturing, policies in lower income countries to remedy teacher deficiencies have primarily taken one of two approaches: 1) by-pass the existing teacher entirely through the use of tutors, assistants, or replacement technology or 2) engage teachers through extensive training and monitoring, typically only successful with careful non-governmental organization (NGO) supervision that exceeds governmental supervisory capacity in most lower income countries (see citations below). This project uses two parallel randomized controlled trials (RCTs) in Punjab, Pakistan to test two alternative models of increasing student achievement: one largely bypassed teachers and encouraged independent learning and the other encouraged existing teachers to become more effective and provided students a more engaging learning environment.

Specifically, we test the impacts of two implementation models of eLearn, a government program to improve student learning in government middle schools in math and science by providing brief videos of expert content. The two models, eLearn Classrooms and eLearn Tablets, started from the premise that both students and teachers could benefit from high quality explanations of concepts in the official science and math curriculum. Electronic copies of the official textbooks, short videos of expert teachers explaining concepts from the official curriculum, a few multiple choice review questions to use after each video, and simulations to demonstrate complex ideas, e.g. photosynthesis, were loaded onto tablet PCs.

The intervention did not provide scripted lessons nor were the video lessons designed to be total substitutes for the teachers. Teachers received a brief two-day training: one day on how to use the multimedia content and one day on how to incorporate it into a more effective, blended teaching practice. This was not new content–the ideas were already in the textbooks that 98 percent of students owned and the multimedia content was already available through a government hosted website. eLearn's videos addressed potential deficiencies in teacher capacity, teaching both the teacher and the student the material, and exposing them to an expert with a clear and engaging teaching style that teachers could replicate themselves. Of particular interest to teachers, the videos modeled how to effectively use written key words, speak clearly, and use follow-up questions to assess comprehension, methods that could be replicated by an in person teacher with effective chalk or white board management. The two interventions addressed deficiencies in separate grade levels: eLearn Classrooms for grade 8 and eLearn Tablets for grade 6.

The mode of delivering the content to the students differed across the two models. eLearn Classrooms integrated the material within the class as only the teachers received the preloaded tablet. To convey the material to students, an LED screen was installed in each classroom where the content could be projected without obscuring the existing white or chalk board.[1] The intervention was designed to complement existing practices and teachers, not add additional employees or act as a substitute for existing personnel. It was a scalable and relatively inexpensive ($9 per student at scale) way to address content knowledge deficiencies and model effective ways to explain complex ideas.

eLearn Tablets used a more independent learning approach. Students received individual tablets with math and science content, but only science teachers received tablets. Teachers had no way to display the content to the entire class. Students could take the tablets home to continue learning and could use them in school even if the teacher was absent. Students

[1] 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.

received training and guidelines on using the tablet at home. Because each student received a tablet, this intervention was more expensive, approximately $131 per student at scale.

We evaluated the two models separately, but simultaneously in the same academic year, through two RCTs that both occurred in Punjab Province, Pakistan.[2] The two interventions had the opposite effect. In our preferred specification, eLearn Classrooms increased student achievement by 0.30 standard deviations (SD) and eLearn Tablets decreased student achievement by about 0.43SD on combined math and science exams designed for the projects. This gain for Classrooms was 60 percent more than the control group score change over the same period and the decrease for Tablets was 95 percent of the learning that occurred in the control group. As grade 8 students take a provincially standardized exam at the end of the academic year, we also test whether the Classrooms intervention increased scores on this high stakes test. Students in the eLearn Classroom intervention also scored 0.27SD higher on the combined math and science sections of this standardized test. When we combine the project and standardized test scores in a single measure of student achievement, the Classrooms intervention improved test scores 0.26SD. This improvement is approximately equivalent to increasing the value-added of a teacher by 1.8SD in Punjab, Pakistan (Bau and Das 2020), 0.7SD in private secondary schools in Uttar Pradesh, India (Azam and Kingdon 2015), or 1.3 to 1.8SD in the US (Hanushek and Rivkin 2010; Jackson, Rockoff, and Staiger 2014). Further, the Classrooms 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.[3] Teachers and students in Classroom treatment schools also increased their likelihood of being

[2]The study schools for the two interventions were distinct and, at the request of our our government partners, not subject to the same randomization. Both Banerjee et al. (2007) and Banerjee et al. (2017) similarly use distinct samples in the same country to test related interventions.

[3]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 student level 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.

at school, unlike those in the Tablets intervention, and teachers in the Classroom treatment schools also increased other self-reported effort.

The positive effect sizes we estimate for Classrooms are the sum of both students and teachers learning from the videos and any modifications that teachers made to their teaching practices. In a conceptual framework we show that the difference between the empirical findings of the two interventions is likely due to complementarities from the classroom screens that led teachers to improve their teaching practice beyond the minutes of the video. The increase in teacher attendance and effort and the heterogeneity findings for the Classrooms program support this claim. The largest test score gains were in the schools with the lowest baseline scores, likely those with the most acute teacher capacity issues, and among students with the lowest baseline scores, those likely well below grade level for whom the content alone without the teacher was likely the least learning-level appropriate. Finally, we estimate that student test scores increases were larger in schools with a teacher with below-median experience and smaller in schools in which teachers had a grade-level specific peer. Teachers with less experience and those without a grade level peer from whom to learn could have had more room to improve their teaching practice based on learning from the high quality teaching examples in the videos.

Our findings have four implications for the literature on improving student achievement.

First, our study contrasts with the implicit idea in many studies, and the one in the Tablets intervention, that existing teachers must be bypassed or substantially re-trained and monitored to increase student learning. We show that a small augmentation in the way content was delivered during the school day–a short video lecture, lasting on average 9 minutes–transformed student achievement while the same content on a student tablet did not. Existing literature provides insight into why using education technology to combine an outside expert with the existing teacher as in the Classrooms intervention could be effective, yet it has not been explicitly tested. The ability to project the videos to the class could have acted somewhat like a commitment device, ensuring that the teachers viewed the full videos themselves, building on findings in the US that providing teachers support and impetus to use materials was more effective than providing materials alone (Jackson and Makarin 2018). During these viewings teachers could see student reactions to the content and the more engaging teaching style of the presenter on the screen–regularly displaying visual aids and key words, posing questions–and model their teaching practice after the most effective elements, almost like having a highly effective colleague to observe, one of the three pillars on which rapidly improving teaching quality rests (Jackson et al. 2014). Further, observing grade level peers teaching the same content are especially important (Jackson and Bruegmann 2009; Papay et al. 2020). The questions that were available with each lesson were integrated with the learning process and provided teachers instant feedback on students' knowledge, allowing teachers to remedy learning deficiencies in real time and were not an extra administrative burden distracting from learning as in Berry et al. (2020). Teachers in the Classrooms intervention increased their attendance instead of using the videos to facilitate their absenteeism, in contrast to other interventions where the additional classroom support was absorbed as a free substitute teacher (Banerjee et al. 2017).[4] This is notable in Pakistan where 14 percent of teachers thought that teacher absence was acceptable if they had left students with something else to do and teachers self-reported being absent 3.2 days per month (World Bank 2017; Bau and Das 2020).[5]

This relatively small external impetus that increased teaching effectiveness is in contrast to previous evidence in low income countries on increasing student achievement that involved much more intense interventions–hiring additional school based teaching staff (e.g. Banerjee et al. 2007 and Banerjee et al. 2017), involving substantial non-governmental organization (NGO) involvement in training and monitoring existing teachers (e.g. Banerjee et al. 2007; Lucas et al. 2014; Banerjee et al. 2017), providing tailored computer programs along with

[4]Based on existing estimates of the relationship between teacher attendance and student achievement, the change in teacher absenteeism is much too small to generate the full achievement effects. See a more in depth discussion in Section 6.2.

[5]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 2020).

extra instructional hours and often a tutor (e.g. Banerjee et al. 2007, Linden 2008, and Muralidharan, Singh, and Ganimian 2019 in India; Mo et al. 2014, Lai et al. 2013, Lai et al. 2015, and Lai et al. 2016 in China; Carrillo et al. 2010 in Ecuador), replacing the majority of each lesson with radio broadcasts, audio CDs, or live interactive broadcasts from a capital city (e.g. Jamison et al. 1981; Naslund-Hadley, Parker, and Hernandez-Agramonte 2014; Johnston and Ksoll 2017), or changing the teacher contract or incentive scheme (e.g. Duflo, Dupas, and Kremer 2015; Barrera-Osorio and Raju 2017; Mbiti et al. 2019; Gilligan et al. forthcoming).[6] [7] The Classrooms intervention is, therefore, less expensive at scale than other interventions with similar effect sizes that involved additional staff and much less expensive at scale than the Tablets intervention. The contrast between the two findings further shows that more expensive interventions are not necessarily better, and can be worse.

Our findings are consistent with studies, both in the US and low income settings, that emphasize the importance of any change being integrated into the school day. Taken together, previous attempts at improving student achievement showed that engaging teachers is difficult and simply providing materials is insufficient. Studies that provided materials or infrastructure without training or monitoring support did not improve student test scores (e.g. Newman et al. 2002, Glewwe et al. 2004, Glewwe, Kremer, and Moulin 2009). The eLearn Tablets intervention is also related to interventions that provided computers directly to children, e.g. one laptop per child.[8] Previous models found no or negative achievement effects when the laptop replaced existing textbooks, were used exclusively at home, or were used primarily at school but without clear curriculum connections versus a 0.17 SD increase in math scores when the laptops included remedial tutoring software (Barrera-Osorio and Linden 2009; Malamud and Pop-Eleches 2011; Fairlie et al. 2013; Mo et al. 2013; Beuermann et al. 2015; Bando et al. 2017; Cristia et al. 2017). Computers could be particularly poorly

[6] Neither Barrera-Osorio and Raju (2017) nor Gilligan et al. (forthcoming) increased test scores for the average student, but both increased persistence in school.

[7] In the US, videos have replaced live in-class teaching through massive open online courses (MOOCs) (Guo, Kim, and Rubin 2014).

[8] 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.

suited for lower income countries as they are overly complicated and have short battery lives, something that could potentially be remedied through the use of tablets (Banerjee et al. 2013). Yet, even though the Tablets intervention contained materials beyond the textbook, was well integrated with the the curriculum, could be used at home and school, and solved at least some of the deficiencies of computers, the intervention decreased test scores, potentially because students lacked guidance for their effective use, used them for other non-scholastic activities, or did not have appropriate spaces at home to use them effectively (Fairlie et al. 2013; Government of Pakistan 2014; Escueta et al. 2017).

Second, the two interventions provide additional evidence on complementarities in education intervention elements and the marginal rate of technical substitution for technology interventions. Due to complementarities among pieces of a literacy intervention, Kerwin and Thornton (2020) found that when the intervention did not provide writing slates, teachers placed less emphasis on teaching writing, resulting in decreased writing scores. Across our two interventions, the LED screens appear to have the strongest complementarities with the content on the tablets–without the ability to project the content, teachers could not learn from student feedback about the effectiveness of the teaching style, receive immediate feedback on student understanding through the questions, nor mimic their additional classroom time after these two effective components. We further provide evidence on the marginal rate of technical substitution between technology and typical classroom or home activities, expanding the type of technology considered in Bettinger et al. (2020). Students in the Tablets intervention were effectively exposed to more technology–they interacted with their tablets without teacher guidance and could use them at home. Their decrease in scores shows this displaced other, more useful activities.

Third, both eLearn Classrooms and Tablets were interventions that were designed by and implemented with the provincial government of Punjab, ensuring a program directly addressing the issues that the government found pressing and increasing the possibility of scale-up of a successful program. This study compares two versions of content delivery, both within existing structures, one that depended on teacher engagement for success–and was successful–and one that bypassed the existing teacher–and was unsuccessful at increasing student test scores. The eLearn Classrooms model established the role of the classroom teacher, not an assistant or outside tutor, and existing supervisory structures, not an NGO, in increasing student test scores. Further, students test score gains translated into score improvements on the government high-stakes test, the exam that parents and the educator sector find more important, and were not limited to remedial learning.

Fourth, this was middle school, grade-level content. Creative grade-level solutions that engage the teachers can improve grade-level competencies, even when the subject level material is more complicated as in middle school.[9] Particularly salient to students in higher grade levels, we find positive effects from eLearn Classrooms on high stakes government exams, 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. Existing evidence on improving middle or secondary school competencies in low income settings 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).[10]

Regardless of income level, because the stock of teachers is relatively fixed, countries need innovative solutions to fill the teacher capacity and effort gaps, improving student achievement without substantial teacher re-training. Most lower-income countries do not have NGOs that can recruit and train assistants nor do they wish to allocate education budgets to such personnel. Teacher incentives have been similarly unpalatable to education sectors. By effectively deploying high quality teaching to classroom screens, student test scores increased due to both the content and its complementarity with teacher effort that led to increased teacher effectiveness. Providing the same material on tablets to students

[9]This is in contrast to the primary school literature that found that grade level content did not improve test scores for the average student (Glewwe et al. 2004, Glewwe, Kremer, and Moulin 2009).

[10]Muralidharan, Singh, and Ganimian (2019) focused on middle school students, but outside of the school day and on their foundational literacy and numeracy skills.

and teachers without the ability to project it to the class decreased achievement.

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. 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.

At the conclusion of middle school, students take 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, our province of focus, 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.

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. 2015; Andrabi et al. 2013). Nationally in 2014, only about 16 percent of middle school students achieved grade-level proficiency in math or science (Government of Pakistan 2014). In our baseline data collection, 51 percent of school principals cited lack of teacher qualifications as a constraint on student learning. The most common middle school science teaching methods in Pakistan have remained stagnant since independence with a focus on rote learning and memorization over conceptual understanding (Pell et al. 2010). This focus on lecture-based instruction and memorization of the items in the textbook is common in developing countries (Glewwe and Muralidharan 2016). Therefore, the issues apparent in teaching middle school science in Pakistan are likely faced by other countries of a similar income level. Modeling effective teaching that teachers could mimic could have a substantial impact.

Despite challenges 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.

3 The eLearn Intervention and Conceptual Framework

We first provide details common to the overall eLearn package then discuss the differences between eLearn Classrooms and eLearn Tablets. Additional details on the interventions appear in Appendix 10.1. We then provide a conceptual framework to explain how the direct and indirect effects differ by intervention.

3.1 eLearn

The overall intervention, eLearn, was designed to improve student learning by providing expert content to enhance existing teachers.

The main component of the interventions was video lectures. Each video lecture was developed and presented by expert teachers to explain a particular math or science concept. All videos directly mapped to the units of the official curriculum and were organized on a tablet within unit folders. Each intervention contained less than 30 hours of content and was designed to be spread over the entire school year. The videos were short, averaging 9 minutes for Classrooms and 4 minutes for Tablets.[11] These lectures modeled effective teaching through the use of extensive visual aids, displaying phrases that signaled and reinforced important concepts, and providing narrative to accompany demonstrations. The tablets further contained about 3 multiple choice assessment questions and their answers for use after each video. Paired with some videos were an additional 3 to 5 minutes of multimedia content, e.g. an interactive animation of photosynthesis. The tablets also contained digital

[11]Based on evidence from Massive Online Open Courses (MOOCs) in the US, 9 minutes is approximately the ideal video length from the perspective of student engagement (Guo, Kim, and Rubin 2014).

versions of the official textbook, which almost all students reported already having in hard copy (around 98 percent) even though less than 32 percent of students reported reading them.

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.[12] Within each treatment arm, all treatment teachers in a district attended the same training regardless of their gender or the gender of their students.

Our intervention took place during the 2016-2017 school year. The timing of the two interventions was slightly different with precise details in the next two subsections. The parallel implementations of the programs that we evaluated were designed to inform the larger scale-up of the programs that was to start in 2018.[13]

3.2 eLearn Classrooms

eLearn Classrooms focused on grade 8 science and math teachers. In eLearn Classrooms, to view and display the video lectures and multimedia content, teachers were given small, pre-loaded tablets, and classrooms received 40 inch LED television screens. Teachers could use these tablets to watch the videos themselves when preparing for lectures and project the content on the installed screens. The screens were installed above the existing chalk or white board, enabling teachers to continue to use the board in an interactive way with the videos.

[12] 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, Barrera-Osorio and Linden 2009, and Bai et al. 2016) 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 2020). 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.

[13] 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.

This classroom technology was designed to augment and complement the teachers' existing teaching techniques. The implementing partner's technology support team visited schools occasionally to ensure equipment was secure and functioning as intended.[14] An additional component of the intervention was designed to engage students and parents at home, but was at most marginally implemented. See Section 10.1 for additional details.

The teacher trainings and hardware installation and tablet distribution were finished by the start of October, after our study baseline data collection. 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 materials distribution and follow-up testing.[15] Figure 1 displays the study and academic year timeline for both interventions. See Section 5.2 for additional details on data collection.

[Figure 1 about here]

3.3 eLearn Tablets

During the same academic year as the eLearn Classrooms implementation, other schools implemented eLearn Tablets. In these schools grade 6 Science teachers received all of the same pieces of the intervention, modified for the grade 6 curriculum, except the screens. Therefore, they lacked the ability to project material to the classroom. Instead, students received their own tablets with both science and math content. As with the Classrooms intervention, teachers could have easily outsourced their own explanations by relying on those on the tablets, but teachers would not have been able to observe which portions of the lessons or methods of explanations students found particularly engaging, whether the students were even viewing the correct content, or received the instant feedback from students pondering follow-up questions.

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

[15] As originally designed, the student baseline testing was to occur 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.

Students received training and guidelines on using the tablet for self-paced learning at home and school. Students used the tablets to watch videos in class, and teachers assigned students to watch the videos at home as homework. Tablets were open for students to use for non-educational purposes such as playing games and watching movies. Parents were not trained on how to monitor or effectively encourage tablet use for educational purposes. Parents received occasional phone calls from technical support staff to ensure devices were working correctly and answer questions about the devices themselves, but not the content or how their children should be using the devices to support their learning. Further, households received a phone number of a helpline that they could call for technical support.

Parents, especially those with limited science and math knowledge themselves, might have had a hard time supporting their students in at home learning. In our sample, about 40 percent of mothers and 25 percent of fathers had no formal schooling themselves. As a typical practice, parents often did not check on their children's homework (Government of Pakistan 2014). Further, students might not have had suitable furniture, such as a table or desk, at home to correctly use the tablet (Government of Pakistan 2014).

The Tablets timeline was designed to be similar to the Classrooms intervention, but faced additional delays. Tablet distribution and associated training were completed in December 2016, after our baseline data collection in November 2016. The follow-up surveys and exam occurred in February 2017. This timeline is displayed in Figure 1.

3.4 Conceptual Framework

The effects of the two interventions on student achievement was the sum of the direct effect of the intervention components plus any indirect effects due to students' or teachers' behavior changes. Equation 1 puts this idea in a basic achievement framework where student achievement is function of both inputs and effort, which can also react to changes in inputs. Formally,

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

where A is student achievement, x are inputs, and $e(x)$ is an effort function that can react to changes in inputs. Our interventions changed the inputs available, x. The total change in A is then

$$\frac{dA}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial e} * \frac{\partial e}{\partial x} \tag{2}$$

As long as $\frac{\partial f}{\partial e} > 0$, then any effort increases as a result of the interventions will increase achievement with the opposite happening for decreases in effort.

In the Classrooms intervention, the total effect on achievement was the direct changes that resulted from having the content available, the teacher watching the content herself, and students and teachers watching the content on the screen together plus any indirect changes to effort that either the teachers or students initiated in reaction to this intervention.

The Tablets intervention shared some of the direct effects–the content was similarly available and the teacher could have watched it on her own as in the Classrooms intervention. Students in this arm could watch the content during class on their personal tablet screens, likely with the same benefit as seeing the content on a larger more distant screen as students did in the Classrooms version. Students could also take their tablets home and experience additional direct benefit of watching the content there. Students and teachers in this intervention could not watch the content together as a group. Figure 2 graphically displays the similarities and differences between the interventions.

[Figure 2 about here]

This study provides multiple useful comparisons to understand the relative importance of direct and indirect effects of the two interventions relative to each other and the status quo control group.

When comparing the Classrooms intervention to the control group, the magnitude of the effect is the sum of the direct and indirect effects from the Classrooms intervention. Previous attempts at improving student learning through either technology or directly engaging the classroom teacher without any changes in incentives or monitoring have not always been successful (see discussion in Section 1), therefore whether this intervention would be successful was an empirical question.

Similarly, the effect of the Tablets intervention relative to the control group is also the sum of the direct and indirect effects for that intervention, another empirical question.

The additional benefit of this study is that we can then compare the magnitudes of the effect sizes across the two interventions to learn something about the sources of the positive or negative treatment effects that we found relative to the control group. Almost all of the direct effects of the two interventions were similar–the availability of the content, the teacher tablets, and students' ability to watch the content in the classroom. Therefore, any differences between the impacts would have to come from the following sources: 1) students in the Tablets intervention had an additional direct benefit that they could use the content to study at home, 2) the indirect effect due to student behavior changes could vary by treatment, and 3) the indirect effect due to teacher behavior changes could vary by treatment.

If the effects across the two interventions are the same, then the two sums of these three effects on student achievement is the equivalent.

If the Classroom effects are larger, then the positive indirect effects that accrue in Classrooms must exceed the additional direct benefits of Tablets and its indirect effects. The only potential differential positive indirect effect in the Classrooms intervention was from the complementarities that the classroom screens generated. The screens allowed teachers to learn from student reactions and model the effective teaching style during the non-video portion of their lessons, potentially increasing student interest in a way that watching the videos as a solitary endeavor did not. Previous work has emphasized the importance of teachers being able to observe and learn form high quality teaching (Jackson and Bruegmann 2009; Jackson et al. 2014; Papay et al. 2020) and having an impetus to use new teaching techniques (Jackson and Makarin 2018) in improving student test scores, two indirect aspects in Classrooms but not Tablets.

In contrast, if the Tablets effects are larger, then the additional direct benefit of being able to use the tablets at home plus related indirect benefits was larger than any indirect benefits in the Classrooms intervention.

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 parallel randomized controlled trials of our interventions.

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. As the the treatment schools for the two interventions were randomly selected from different samples, we estimate the effects separately with the same empirical specification. Formally we estimate

$$y_{is} = \alpha + \beta treatment_s + X_{is}^{'}\Gamma + \varepsilon_{is} \tag{3}$$

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 (eLearn Classroom or eLearn Tablets, depending on the sample), 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 implement 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 student level baseline test scores.[16] We also provide specifications with additional hand-picked controls.

Our primary outcomes of interest are student test scores. We first test for the impact on exams designed specifically for this project. Even though these bespoke exams followed the official curriculum, one could worry that the intervention artificially improved scores only on this test, which makes the results for Tablets particularly disheartening. A distinguishing feature of eLearn Classrooms is that we have two types of tests for that intervention. 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, a margin that is relevant for longer term outcomes.[17] 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.

For both interventions we estimate whether the intervention affected the likelihood that the student was present on the day of the follow-up.[18]

For eLearn Classrooms, additional data allow us to test for potential mechanisms. For Classrooms only we further estimate the effect of the intervention on student self-reported

[16] The variables we consider are listed in the Appendix Section 10.2.

[17] 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.

[18] A student being present is also our measure of attrition. Our findings are robust to attrition correction. See Section 6.

effort and technology use.

Additionally, based on data collected from teachers, for Classrooms alone we estimate a similar model, allowing i to index the teacher instead of the student. The outcomes of interest for teachers are whether they used technology to prepare for classes, used technology to teach their classes, had been part of any training, held private tutoring sessions outside of school, performed other official duties, and were approached by students for help outside of class time. We also estimate the effect of the program on the teacher's average number of classes taught, how many hours they spent preparing for class, and how many extra classes they taught in a month during school hours to cover grade 8 syllabus. Finally, for both interventions 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.[19] 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 and tablets. 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

[19]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.

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 schools separately for eLearn Classrooms and eLearn Tablets.

For eLearn Classrooms, we selected 60 schools, an equal number of boys' and girls' schools, for the sample.[20] 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.

For eLearn Tablets, we selected 75 schools from among those schools not selected for Classrooms. Twenty schools were randomly chosen to receive tablets and associated training, stratified by district.[21] Four schools either opted out or were eliminated from the study due to poor network connectivity, leaving 71 total schools, 19 treatment and 52 control 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.

Primary Data Collection

The data collected across the two interventions was similar, but on different schedules and with some additional details collected from the Classrooms schools. The baseline surveys solicited information from head teachers, the relevant grade math and science teachers, and randomly selected students from the relevant grade present on the day of the baseline. All

[20] We would have liked to have a larger sample but were only able to raise enough money for a 60 school sample. A grant directly to the government covered the implementation but not full evaluation costs. 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 was especially compelling and rendered a larger study unnecessary and would not fund it. Our government partners deemed randomization at a level lower that the school, e.g. classroom or teacher, politically infeasible. Further, only about half of our schools had multiple sections of the relevant grade level.

[21] The government only had funding to supply 20 schools with tablets in the first year. These 20 schools were designed to be the first phase of a multi-phase roll-out. The rest of the roll-out did not occur due to the negative results of the evaluation.

present students in the relevant grade took our project mathematics and science tests that followed the established curriculum while testing higher order conceptual and problem solving abilities than the official provincial tests that rely heavily on rote memorization.[22] In the Classrooms baseline we tested 2,999 students and conducted 1,690 student interviews across 59 schools in in late August 2016, two instructional months into the 2016-2017 academic year, prior to the teacher training or availability of the new technology. In the Tablets baseline we tested and interviewed 3,614 students across 71 schools in November 2016.[23] Enumerators told schools that we would be visiting them near the end of the school year, but they did not provide an exact date. We administered follow-up surveys and exams in January 2017 for Classrooms schools and February 2017 for Tablets schools. The same students were again surveyed and tested, if present.[24] Head teachers and subject teachers were again surveyed. The school year ended in March. 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.[25] Our findings are similar using raw test scores.

For the Classrooms sample we further test for changes in technology use and other student and teacher behaviors.

Administrative Data

We use administrative data to create additional measures of student achievement for the students in the Classrooms arm. For each student we have the administrative student by

[22] Additional test details appear in Appendix Section 10.3.

[23] The surveys and program implementation for both interventions were originally designed to occur prior to the June to August summer holiday but implementation funding was delayed.

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

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

subject level PEC exam results and whether the student passed the PEC from the Punjab Examination Commission.[26] Students completed the PEC exams in mid-February. We merged these data to the students in our sample using students' and fathers' full names.[27] Because we do not have item level responses, these scores are scaled with a mean 0 and standard deviation of 1. Further, 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, as a third measure of achievement. The two exams—project-specific and administrative—were both designed to cover material from the same curriculum. 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.

For both interventions we use administrative data on teacher attendance 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.[28] These data were available at the school level only. Therefore, they measure the percentage of all teachers in the school present during the visit.

Summary Statistics

Table 1 displays means and standard deviations of student (Panel A), teacher (Panel B), and school (Panel C) characteristics across the two interventions and the treatment and control schools. Almost all of the measures are statistically indistinguishable by treatment

[26]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.

[27]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.

[28]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 and students with tablets in some grade 6 classrooms. We cannot reject that this might have influenced their overall assessment of teacher attendance in a school, but believe it to be unlikely.

status with two exceptions at the 10 percent level in the Classrooms intervention: treatment students report being absent more often in the previous month by 0.3 days 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).[29] Given we are testing 14 outcomes across two samples, 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. Appendix Figure A1 shows the baseline test score distributions. Across both samples only one student scored a 0 on the baseline exam.

[Table 1 about here]

6 Results

We first test for the effects of the program on students' test scores for both the project-specific and PEC exams. Then, we explore possible mechanisms behind the achievement results including student attendance, an interesting outcome itself as well as our measure of attrition.[30]

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 student level baseline test scores as control variables. The results of this estimation appear in Table 2. Panel A is a parsimonious specification that includes only the strata and student level baseline test scores as control

[29] Even though most of the teachers have an advanced degree, teacher content knowledge does not necessarily translate into the ability to explain content to students (Lu et al. 2019). Further, the skills necessary to earn a university degree might not be applicable to the middle school curriculum. In Punjab, Bau and Das (2020) found that a teacher having a college degree is associated with only a 0.2 SD 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.

[30] We were unable to reach 15 percent of our baseline samples during our endlines. In Section 6.2 we test for differential attrition by treatment status and provide Lee (2009) bounds.

variables. eLearn Classrooms increased achievement by 0.26 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 Classrooms on the math and science portions of the standardized government PEC tests. This treatment increased this PEC score by 0.22 SD. These two effects are statistically significant at the ten percent level. In column 3, we combine these tests into a single score measure. Classrooms increased the combined test score by 0.27 SD, statistically significant at the 5 percent level. In column 4 we estimate the effect of Classrooms on the likelihood that a student passed the PEC exam, and find a positive, statistically insignificant effect.[31] In column 5, we estimate the effect of the Tablets intervention on a the combined Math and Science score. In contrast to the results in the first four columns, the Tablets intervention decreased test scores by 0.42 SD.

[Table 2 about here]

Given our small sample size and slight imbalances from Table 1, we include additional student, teacher, and school control variables in Panel B as determined by the 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 farther from 0 with stronger statistical significance. The Classrooms treatment increased test scores on the project exam by 0.30 SD (column 1), had a 0.27 SD effect on the PEC score (column 2), and increased the combined score by 0.26 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 ($p< 0.05$) 5 percentage points. As in Panel A, the Tablets intervention decreased student test scores, by 0.43 SD when using the LASSO

[31] 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.

controls (Panel B, Column 5).[32] [33]

During this same period control group students increased their project test scores by 0.49 SD in the Classrooms sample and 0.45 in the Tablets sample.[34] Therefore, based on Panel B, Column 1, the Classrooms intervention increased the project test score by 60 percent relative to the gains in the control group. In contrast, the Tablets intervention eliminated almost all (95 percent) of the gains during the three instructional months of the intervention.[35] These differences point towards strong positive indirect effects of the classroom screens due to the complementarities between the installed screens and other behaviors. In the rest of the section we test for other changes.

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

[32] In additional specifications we include other controls instead of using LASSO, first controlling only for additional student-level covariates and then controlling for additional student, teacher, and school level covariates. These results appear in Appendix Table A1 and are similar to those presented in Table 2.

[33] Because of our sample size we calculated p-values adjusted for randomization inference. The corresponding p-values for the coefficients of interest in Table 2 are, from Column 1 to Column 5, 0.057, 0.014, 0.054, 0.100, and 0.015.

[34] To calculate the control group increase in test scores across the two rounds, we use the IRT adjusted and standardized scores for all control group students who completed both the baseline and the endline. We then compare the means between the two rounds of test scores displaying the difference in those means in the table.

[35] Appendix Table A2 contains the subject-specific project exam effects for both Classrooms and Tablets. For the Classroom intervention it also contains the results for the subject-specific PEC scores 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 insignificant positive effect on the other test scores.

itself, it is also a measure of respondent attrition. We use an indicator variable equal to 1 if the student was present at follow-up as the outcome, y_{is}, in Equation 1. As with all our other binary outcomes, we use a linear probability model. The results appear in Table 3.

[Table 3 about here]

Students in the Classrooms treatment group were about 4 percentage points more likely to be present at follow-up (column 1), demonstrating increased student effort and engagement as a result of the intervention. 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 A3 we follow Lee (2009) and provide treatment bounds. Our test score findings are robust to this attrition adjustment.

A second potential point of attrition is whether a baseline student took the PEC exam. Columns 3 and 4 of Table 3 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 and no statistically significant differential affect by baseline test score and treatments status.[36]

The Tablets intervention did not statistically significantly change the likelihood that a student was present during the follow-up (columns 5 and 6). Nevertheless, we provide treatment bounds based on Lee (2009) in Appendix Table A3 columns 3 and 4, finding our results robust to this bounding exercise.

[36] 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.

We further test for treatment effects on teacher attendance. As an objective measure of teacher attendance 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 1 and 2 we estimate the effect of the classrooms 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 monthly observations for each school. Our intervention increased the portion of teachers present in the school by about 1 percentage point (column 1).[37] This finding is in contrast to the concern prior to the implementation that teachers would be more more likely to be absent as the videos could be virtual substitutes. In this context, teacher attendance was high even in the control group (94 percent). Teachers being present to teach is one component of overall teacher effectiveness as absent teachers cannot be effective at increasing their students' learning.[38]

[Table 4 about here]

As this is a monthly measure, we can test the evolution over time in teacher attendance. In column 2 we test whether this response changed over time. The point estimate on the

[37]Since we do not have PMIU data at the individual level, this is the effect on teacher attendance for the whole school. Treatment teachers were less than 10 percent of teachers in a school. This change in absenteeism is likely not the result of a Hawthorne effect, i.e. teachers changing their behavior because they were being observed. Both treatment and control teachers were part of the same experiment and equally observed by the enumeration team as a part of this study. They were further equally observed by the PMIU as part of normal, monthly data collection activities.

[38]The increase in student achievement was likely not only caused by increased teacher attendance as our achievement results are much larger than those implied by the modest gains to attendance. If sample treatment teachers attended the average amount recorded in the control group, they could have at most increased their attendance 6 percentage points. A 6 percentage point increase in attendance resulting in a 0.3SD increase in test scores is well outside existing estimates. When teachers increased their attendance by 21 percentage points in Duflo et al. (2012), student test scores increased 0.17SD. With an increase in teacher attendance of 8 percentage points, Cilliers et al. (2018) found no change in student test scores. The estimates in Herrmann and Rockoff (2012) and Gershenson (2016) imply a 6 percentage point change in absenteeism would result in at most a 0.02SD change in test scores. Das et al. (2007) did not find a statistically significant relationship between teacher absenteeism and student test scores. For a subsample, they estimated a 5 percent increase in teacher absence reduces test scores 4 to 8 percent of average gains over the year. In our context that would be a 5 percentage point increase in attendance and a 0.02SD to 0.04SD increase in learning. Therefore, while teacher attendance did increase, the change was likely too small to generate the entire test score increases.

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.[39]

When we perform the same exercise for Tablets schools, we find that the Tablets intervention did not change teacher attendance (columns 3 and 4).

6.3 Technology Use in Classrooms

To measure whether and how teachers used the Classrooms technology we use three sources: data collected by the tablets, teachers' survey responses, and students' survey responses.

The tablets recorded data on time of use and number of items used each month. The data collected by the tablets reported that on average each school accessed 74 of 192 videos (39 percent), 11 of 50 simulations (22 percent), and 152 of 600 questions (25 percent).[40] At an average video length of just over 9 minutes, this implies just over 11 hours of video content. The use of the three was positively, but not perfectly correlated. Schools on average accessed 2 questions for each video played and 1 simulation for each 9 videos played.[41] Therefore, schools appear to use the intervention as a bundle, as intended.

Figure 3 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–teachers accessed 81 percent of videos, 70 percent of simulations, and 90 percent of questions during school hours. Even at its lowest point in February the average school was still accessing some content. Recall that students took the PEC exam in mid-February, therefore teaching time was both interrupted and structured differently in that month.[42]

[39] Based on the point estimates, the portion present would revert to the non-treatment level after 4 months.

[40] Each school accessed at least 27 videos and 27 questions. One school did not access any of the simulations.

[41] Each video had on average three questions and every four videos had a simulation that could accompany it. The r-squared on bivariate regressions between videos and questions accessed is 0.60, between videos and simulations accessed is 0.21, and between questions and simulations accessed is 0.19.

[42] Teachers were encouraged to use all of the content but the actual use was left to their discretion. From

[Figure 3 about here]

Teachers further self reported their own technology use. We first test whether the intervention changed the teachers' technology use on the extensive margin. Table 5 contains these results. Teachers were 33 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).[43]

[Table 5 about here]

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."

Even though most of the use of the tablets was during school hours, the tablets cannot report whether the content was displayed to the students. The students were asked how frequently their teachers displayed the content. Over 95 percent of students reported that their teachers displayed the videos and used the assessment questions at least once a week. About 45 percent of the students reported daily use of the assessment questions and over half reported daily use of the videos. These responses are highly correlated with the data from the tablets. Therefore, teachers and students likely viewed this content together.

Teams from our implementing partner conducted two spot check visits to each school during the intervention. During these visits, 83 percent of eLearn Classroom schools used at least one teacher tablet during the visit. Therefore, technology use in the classroom was an important part of the intervention.

Because of concerns over student privacy, we do not have data from the tablets used in the Tablets intervention.

---

the experimental design, we cannot know whether the ideal amount of use is closer to the November peak or the December and January levels, or whether heavy use in November provided teachers sufficient modeling to be more effective themselves and they no longer chose to rely on the videos. Further, the November peak could be due to learning how to navigate the software and selecting videos in error.

[43] At the baseline, of those teachers who used technology in the classroom, about 70 percent used a mobile phone and 20 percent used a computer or the internet.

6.4 Other Changes

We additionally collected self-reported data on changes in other inputs and effort from teachers and students in the Classrooms sample. According to data collected during the training, all treatment teachers attended the training. Column 3 of Table 5 shows that treatment teachers self-reported attending 0.33 more in-service teacher training events during the school year.

We tested for additional changes in teacher effort that might have occurred as a result of the intervention. Teachers in the Classrooms treatment spent more minutes per day planning lessons (9.1 minutes off a control group base on 58) and were 16 percentage points (off a control group mean of 45 percent) more likely to have students approach them outside of class for additional help, demonstrating both more student and teacher effort and an increased level of comfort in the relationship between students and teachers. We find statistically insignificant changes in the likelihood of holding private tutoring sessions, the number of regular classes taught per week, and the number of extra classes per month to cover the syllabus. Full point values appears in Appendix Table A4. Therefore, teachers in treatment schools increased their use of technology, their observed effort (attendance), and 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).[44]

[44] Self-reported absenteeism is an imperfect measure of actual attendance as students might misreport and those present are a selected sample. This should be symmetric across treatment statuses.

7 Heterogeneity

Because the intervention content was at the level of the curriculum and some of the students were likely behind grade level, the intervention could have differential effects by baseline test score. We test for this possibility by including an interaction term between the treatment indicator and baseline test score quartile as additional regressors in Equation 1 with the project test, the PEC, and the combined project and PEC score as three separate outcomes.[45]

The point values for Classrooms appear in Appendix Table A5, Panel A, columns 1 to 3. Based on this specification, the Classrooms intervention had a U shaped impact on student project scores. For the project exam, students in the lowest quartile gained the most and students in the second quartile gained the least. For students not in the lowest quartile, we fail to reject that the overall effect is 0.[46] We find a similar U shaped relationship for the PEC–students in lowest and top two quartiles had positive test scores gains due to the intervention. Not surprisingly, the combined project and PEC score reflects this same relationship with statistically significant score improvements only for those students in the top or bottom quartile at baseline. Appendix Figure A2 shows the non-parametric treatment effects across the distribution.

The results for the Tablets intervention appear Table A5, Panel A, column 4. Across all baseline quartiles, the intervention caused test scores to decrease with no statistically significant differences between the lowest and other quartiles.

Ideally one might like to know heterogeneity by school quality using school value added as a proxy. Unfortunately, we cannot calculate that with available data. Instead we test for heterogeneity by the average school level baseline test score, understanding that this likely encompasses much more than school value added, including household wealth and whether the community prioritizes schooling. To test for heterogeneity on this margin we include

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

[46]In Appendix Table A6 we estimate the effect of the intervention by difficulty of the test questions (see Appendix Section 10.3 for more details). The test score gains are statistically significant and about 0.2SD for below median questions and statistically insignificant for above median questions.

interactions between the treatment and the school level score quartile as additional regressors in Equation 1. These results appear in Panel B of Appendix Table A5. For Classrooms, students in third quartile schools are the only ones who had statistically significant test scores gains on the Project exams while students in the lowest quartile schools increased their test scores on the PEC (columns 1 and 2). When combining the two scores, the students in the lowest quartile schools increased their test scores (column 3). For Tablets, student test scores decreased in schools of the lowest two quartiles and increased in the third quartile, with no change for the top quartile (column 4). Students from lower scoring schools might have had lower levels of schooling and household infrastructure that would prevent them from using the tablets effectively at school or home. The contrast between the two findings for the lowest quartile schools–student scores increased in the Classroom schools and decreased in the Tablets schools–shows the importance of and potential for teacher engagement for increasing student test scores in schools that had been lower performers with likely fewer teaching and household resources.

We further test for heterogeneity by school gender, replacing the test score interaction 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. We do not find any consistent benefit accruing to schools of one gender over another. See additional discussion in Appendix Section 10.4

We finally test for heterogeneity by two characteristics of schools that could determine whether teachers were more likely to benefit from having an additional virtual peer to mimic. We find that students in schools with a teacher who had below median experience have test score increases statistically different from 0, while those in schools with more experienced teachers do not show test score increases (Appendix Table A7, Column 1). Further, test score gains were concentrated in schools in which teachers did not already have a grade-level subject peer teacher (Appendix Table A7, Column 2).

8 Cost Effectiveness

Classrooms

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 for an intervention like Classrooms where the intervention is at the classroom and not student level. Adding an additional student to an existing classroom in the Classrooms intervention is free, understanding that at some point a class would become too large for a single teacher. 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.[47] 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 other attributes discussed in Appendix Section 10.1 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.[48]

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

[48] 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.

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).[49] A program that linked school committees to local governments in Indonesia (Pradhan 2012) was more cost effective. None of the other available studies attempted to transform what was happening in a middle school classroom. A second measure to consider in cost effectiveness is student time. Most other effective technology interventions included out of school time, in some cases multiple hours per week. Our intervention does not include any out of school time for students.

Tablets

Despite Tablets decreasing student achievement, we provide the cost estimates using the same method as we used for Classrooms. The largest difference between the two programs' costs were the costs of a tablet for each student, phone calls to students' household to support the technology, and staffing a help-desk to respond to households' questions. Ignoring the fixed costs, the marginal cost per student at a new school was $130.74 assuming a two year tablet life. As implemented, the fixed costs were $20/student at the 20 school scale. As with the Classroom intervention, the larger the program, the more these fixed costs decrease per student. Even at the largest scale possible, the intervention would still have the $130.74 fixed costs per student, assuming schools of similar size to our intervention schools.

[49] 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.

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. We tested this potential through two separate, simultaneous RCTs in Punjab, Pakistan, as part of the Punjab government's eLearn project. As part of eLearn project the government of Punjab digitized text books and created videos, multiple choice questions, and simulations to accompany the existing curriculum. In the eLearn Classrooms version of the project, grade 8 science and math teachers received this content on personal tablets and LED screens were installed in classrooms for teachers to project the content to the class. In eLearn Tablets, grade 6 science teachers and all grade 6 students received this math and science content. We find that the Classrooms 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. These effects are the sum of the students learning directly from the content and teachers modifying their teaching practices to mimic the effective teaching that they observed on the screen. Given the large effect sizes and short total content–29 hours divided into approximately 9 minute videos–student learning from videos alone is likely not the only channel. The results from the Tablets intervention support this hypothesis–the availability of similar content, adjusted for grade level, decreased student test scores by about 0.4SD. Therefore, the content alone without its integration into an effective teaching practice was not sufficient to increase student test scores. Within the Classrooms intervention, schools with the lowest average baseline score and students with the lowest baseline scores gained the most even though the content was grade-level and potentially well above the learning-level for these students, showing that interventions that improve teaching effectiveness, even if the content is grade-level, do not necessarily exacerbate existing score heterogeneity. Increased effort is reflected in increased teacher and student attendance. This study reinforces the potential for effective peers to model high quality teaching and improve teacher effectiveness that has been shown in the US (Jackson and Bruegmann 2009; Papay et al. 2020). Prior to this study very little was known about improving teacher effectiveness without intensive monitoring, supervision, training, or incentives or improving grade-level content delivery in developing country middle schools.

Returning back to the conceptual framework outlined in Section 3.4, since the direct effects were likely similar, or potentially favored the Tablets schools, the difference in achievement effects are likely due to the different indirect effects of the two intervention. The classroom time in Classrooms likely became more effective, and teachers and students demonstrated increased effort in other ways, while in Tablets students likely were distracted by their tablets either at home or school, displacing other useful academic activities or sleeping. In Classrooms, teachers watched the videos and used the comprehension questions with the students, learning from the expert teachers what students found effective and receiving immediate feedback on their students' comprehension. These indirect effects point to strong complementarities between the Classrooms intervention and teachers' effort. As we find positive effects for the Classrooms intervention and negative effects for the Tablets intervention, the change in classroom effectiveness was an important component of the overall effect of the Classrooms intervention.

Finally, at a mere 200 school scale the cost effectiveness of eLearn Classrooms 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. At a cost of $131 per student even at scale, Tablets was substantially more expensive, and less effective.

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, especially for students of low baseline learning levels, potentially overcoming existing teacher capacity constraints.

References

Aker, Jenny C, Ksoll, Christopher, & Lybbert, Travis J. 2012. Can mobile phones improve learning? Evidence from a field experiment in Niger. *American Economic Journal: Applied Economics*, **4**(4), 94–120.

Andrabi, Tahir, Das, Jishnu, Khwaja, Asim Ijaz, Vishwanath, Tara, & Zajonc, Tristan. 2007. Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate. *World Bank, Washington, DC*.

Andrabi, Tahir, Das, Jishnu, & Khwaja, Asim Ijaz. 2013. Students today, teachers tomorrow: Identifying constraints on the provision of education. *Journal of public Economics*, **100**, 1–14.

Andrabi, Tahir, Das, Jishnu, Khwaja, Asim Ijaz, Singh, Niharika, & Ozyurt, S. 2015. Upping the Ante: The Equilibrium Effects of Unconditional Grants to Private Schools. *Unpublished paper*.

Azam, Mehtabul, & Kingdon, Geeta Gandhi. 2015. Assessing teacher quality in India. *Journal of Development Economics*, **117**, 74 – 83.

Bai, Yu, Mo, Di, Zhang, Linxiu, Boswell, Matthew, & Rozelle, Scott. 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.

Bando, Rosangela, Gallego, Francisco, Gertler, Paul, & Fonseca, Dario Romero. 2017. Books or laptops? The effect of shifting from printed to digital delivery of educational content on learning. *Economics of Education Review*, **61**, 162–173.

Banerjee, Abhijit, Cole, Shawn, Duflo, Esther, & Linden, Leigh. 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*.

Banerjee, Abhijit, Glewwe, Paul, Powers, Shawn, & Wasserman, Melanie. 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*.

Banerjee, Abhijit, Banerji, Rukmini, Berry, James, Duflo, Esther, Kannan, Harini, Mukerji, Shobhini, Shotland, Marc, & Walton, Michael. 2017. From proof of concept to scalable policies: challenges and solutions, with an application. *Journal of Economic Perspectives*, **31**(4), 73–102.

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

Barrera-Osorio, Felipe, & Ganimian, Alejandro J. 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, Felipe, & Linden, Leigh L. 2019. *The use and misuse of computers in education: evidence from a randomized experiment in Colombia.* Working Paper 4836. World Bank.

Barrera-Osorio, Felipe, & Raju, Dhushyanth. 2017. Teacher performance pay: Experimental evidence from Pakistan. *Journal of Public Economics*, **148**, 75 – 91.

Bau, Natalie, & Das, Jishnu. 2020. Teacher Value-Added in a Low-Income Country. *American Economic Journal: Policy*, **12**, 62–96.

Bellei, Cristián. 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, Alexandre, Chernozhukov, Victor, & Hansen, Christian. 2014. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, **28**(2), 29–50.

Berry, James, Kannan, Harini, Mukherji, Shobhini, & Shotland, Marc. 2020. Failure of frequent assessment: An evaluation of Indias continuous and comprehensive evaluation program. *Journal of Development Economics*, **143**, 102406.

Bettinger, Eric, Fairlie, Robert W, Kapuza, Anastasia, Kardanova, Elena, Loyalka, Prashant, & Zakharov, Andrey. 2020 (April). *Does EdTech Substitute for Traditional Learning? Experimental Estimates of the Educational Production Function.* Working Paper 26967. National Bureau of Economic Research.

Beuermann, Diether W., Cristia, Julian, Cueto, Santiago, Malamud, Ofer, & Cruz-Aguayo, Yyannu. 2015. One Laptop per Child at Home: Short-Term Impacts from a Randomized Experiment in Peru. *American Economic Journal: Applied Economics*, **7**(2), 53–80.

Bold, Tessa, Filmer, Deon P, Molina, Ezequiel, & Svensson, Jakob. 2019. *The Lost Human Capital: Teacher Knowledge and Student Achievement in Africa.* The World Bank Policy Research Working Paper 8849.

Burdett, Newman. 2017 (December). *Review of High Stakes Examination Instruments in Primary and Secondary School in Developing Countries.* Tech. rept. RISE-WP-17/018. RISE Working Paper.

Cai, Xiqian, Lu, Yi, Pan, Jessica, & Zhong, Songfa. 2019. Gender Gap under Pressure: Evidence from China's National College Entrance Examination. *Review of Economics and Statistics*, **101**(2), 249–263.

Carrillo, Paul, Onofa, Mercedes, & Ponce, Juan. 2010. *Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador.* Inter-American Development Bank Working Paper IDB-WP-223.

Chaudhury, Nazmul, Hammer, Jeffrey, Kremer, Michael, Muralidharan, Karthik, & Rogers, F Halsey. 2006. Missing in action: teacher and health worker absence in developing countries. *Journal of Economic perspectives*, **20**(1), 91–116.

Cilliers, Jacobus, Kasirye, Ibrahim, Leaver, Clare, Serneels, Pieter, & Zeitlin, Andrew. 2018. Pay for locally monitored performance? A welfare analysis for teacher attendance in Ugandan primary schools. *Journal of Public Economics*, **167**, 69–90.

Cristia, Julian, Ibarrarán, Pablo, Cueto, Santiago, Santiago, Ana, & Severín, Eugenio. 2017. Technology and Child Development: Evidence from the One Laptop per Child Program. *American Economic Journal: Applied Economics*, **9**(3), 295–320.

Das, Jishnu, Dercon, Stefan, Habyarimana, James, & Krishnan, Pramila. 2007. Teacher shocks and student learning evidence from Zambia. *Journal of Human resources*, **42**(4), 820–862.

Duflo, Annie, Kiessel, Jessica, & Lucas, Adrienne. 2020. *Alternative Models of Increasing Student Achievement: Evidence from a Nationwide Randomized Experiment*. Working paper.

Duflo, Esther, Hanna, Rema, & Ryan, Stephen P. 2012. Incentives work: Getting teachers to come to school. *American Economic Review*, **102**(4), 1241–78.

Duflo, Esther, Dupas, Pascaline, & Kremer, Michael. 2015. School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics*, **123**, 92–110.

Escueta, Maya, Quan, Vincent, Nickow, Andre Joshua, & Oreopoulos, Philip. 2017. *Education Technology: An Evidence-Based Review*. National Bureau of Economic Research Working Paper 23744.

Fairlie, Robert W., & Robinson, Jonathan. 2013. Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. *American Economic Journal: Applied Economics*, **5**(3), 211–40.

Gershenson, Seth. 2016. Performance standards and employee effort: Evidence from teacher absences. *Journal of Policy Analysis and Management*, **35**(3), 615–638.

Gilligan, Daniel O., Karachiwalla, Naureen, Kasirye, Ibrahim, Lucas, Adrienne M., & Neal, Derek. forthcoming. Educator Incentives and Educational Triage in Rural Primary Schools. *Journal of Human Resources*.

Glewwe, Paul, & Muralidharan, Karthik. 2016. Improving school education outcomes in developing countries: Evidence, Knowledge gaps, and Policy Implications. *Pages 653–743 of: Handbook of Economics of Education*, vol. 5. North Holland.

Glewwe, Paul, Kremer, Michael, Moulin, Sylvie, & Zitzewitz, Eric. 2004. Retrospective vs. prospective analyses of school inputs: the case of flipcharts in Kenya. *Journal of Development Economics*, **74**, 251–268.

Government of Pakistan. 2014. *National Assessment Report*. Tech. rept.

Guo, Philip J, Kim, Juho, & Rubin, Rob. 2014. How video production affects student engagement: An empirical study of MOOC videos. *Pages 41–50 of: Proceedings of the first ACM conference on Learning@ scale conference*.

Hanushek, Eric A, & Rivkin, Steven G. 2010. Generalizations about using value-added measures of teacher quality. *American Economic Review*, **100**(2), 267–71.

Herrmann, Mariesa A, & Rockoff, Jonah E. 2012. Worker absence and productivity: Evidence from teaching. *Journal of Labor Economics*, **30**(4), 749–782.

Jackson, C. Kirabo. 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.

Jackson, C Kirabo, & Bruegmann, Elias. 2009. Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, **1**(4), 85–108.

Jackson, C Kirabo, Rockoff, Jonah E, & Staiger, Douglas O. 2014. Teacher effects and teacher-related policies. *Annu. Rev. Econ.*, **6**(1), 801–825.

Jackson, Kirabo, & Makarin, Alexey. 2018. Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment. *American Economic Journal: Economic Policy*, **10**(3), 226–54.

Jamison, Dean T, Searle, Barbara, Galda, Klaus, & Heyneman, Stephen P. 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, Jamie, & Ksoll, Christopher. 2017. *Effectiveness of Interactive Satellite-Transmitted Instruction: Experimental Evidence from Ghanaian Primary Schools*. Center for Education Policy Analysis Working Paper 17-08.

Kerwin, Jason T., & Thornton, Rebecca L. 2020. Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures. *The Review of Economics and Statistics*, **0**(ja), 1–45.

Kremer, Michael, Conner, Brannen, & Glennerster, Rachel. 2013. The Challenge of Education and Learning in the Developing World. *ScienceMag*, **340**.

Lai, Fang, Zhang, Linxiu, Hu, Xiao, Qu, Qinghe, Shi, Yaojiang, Qiao, Yajie, Boswell, Matthew, & Rozelle, Scott. 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.

Lai, Fang, Luo, Renfu, Zhang, Linxiu, Huang, Xinzhe, & Rozelle, Scott. 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, Fang, Zhang, Linxiu, Bai, Yu, Liu, Chengfang, Shi, Yaojiang, Chang, Fang, & Rozelle, Scott. 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.

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

Linden, Leigh L. 2008. *Complement or substitute?: The effect of technology on student achievement in India.*

Lu, Meichen, Loyalka, Prashant, Shi, Yaojiang, Chang, Fang, Liu, Chengfang, & Rozelle, Scott. 2019. The impact of teacher professional development programs on student achievement in rural China: evidence from Shaanxi Province. *Journal of Development Effectiveness*, **11**(2), 105–131.

Lucas, Adrienne M, & Mbiti, Isaac M. 2014. Effects of school quality on student achievement: Discontinuity evidence from Kenya. *American Economic Journal: Applied Economics*, **6**(3), 234–263.

Lucas, Adrienne M, McEwan, Patrick J, Ngware, Moses, & Oketch, Moses. 2014. Improving Early-Grade Literacy In East Africa: Experimental Evidence From Kenya And Uganda. *Journal of Policy Analysis and Management*, **33**(4), 950–976.

Malamud, Ofer, & Pop-Eleches, Cristian. 2011. Home computer use and the development of human capital. *The Quarterly journal of economics*, **126**(2), 987–1027.

Mbiti, Isaac, Muralidharan, Karthik, Romero, Mauricio, Schipper, Youdi, Manda, Constantine, & Rajani, Rakesh. 2019. Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. *The Quarterly Journal of Economics*, **134**(3), 1627–1673.

Mo, Di, Swinnen, Johan, Zhang, Linxiu, Yi, Hongmei, Qu, Qinghe, Boswell, Matthew, & Rozelle, Scott. 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, Karthik. 2013. Priorities for primary education policy in Indias 12th five-year plan. *Pages 1–61 of: India Policy Forum*, vol. 9. National Council of Applied Economic Research.

Muralidharan, Karthik, Singh, Abhijeet, & Ganimian, Alejandro J. 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, **109**(4), 1426–1460.

Naslund-Hadley, Emma, Parker, Susan W, & Hernandez-Agramonte, Juan Manuel. 2014. Fostering early math comprehension: Experimental evidence from Paraguay. *Global Education Review*, **1**(4).

Navarro-Sola, Laia. 2019. *Secondary School Expansion through Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria*. Working paper.

Newman, John, Pradham, Menno, Rawlings, Laura B., Ridder, Geert, Coa, Ramiro, & Evia, Jose Luis. 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.

Papay, John P., Taylor, Eric S., Tyler, John H., & Laski, Mary E. 2020. Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data. *American Economic Journal: Economic Policy*, **12**(1), 359–88.

Pell, Anthony William, Iqbal, Hafiz Muhammad, & Sohail, Shahida. 2010. Introducing science experiments to rote-learning classes in Pakistani middle schools. *Evaluation & Research in Education*, **23**(3), 191–212.

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

Pradhan, Menno, Suryadarma, Daniel, Beatty, Amanda, Wong, Maisy, Alishjabana, Armida, Gaduh, Arya, & Artha, Rima Prama. 2011. *Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia*. The World Bank.

10 Appendix

10.1 Additional Program Details

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

eLearn

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

eLearn Classrooms

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

As designed, at the conclusion of each unit, students should have received an SMS on their households' mobile phones that an Intelligent Tutoring System (ITS) module was available for use. This ITS system would have allowed students to use text messages to receive and respond to review questions. While almost all study students reported having at least one mobile phone in their household and 80 percent reported having more than one, due to delays this system was barely implemented during our study. Only one third of the treatment schools received at least one module, consistent with the 25 percent of students in the treatment group who reported receiving at least one module. Among those students who received at least one module, only 10 percent, or 2.5 percent of the overall treatment group, received more than 3 of the 22 intended modules. Finally, the mobile carrier who broadcast the SMSs incorrectly charged the students to respond to these texts, leading to low take-up even among those who were reached. An interactive voice response system (IVR) was to call parents to inform them that their child was absent and allow parents to respond with the reason for the absence. The IVR system was not operational during our period of study. Therefore, while at home engagement was designed to be a component of the intervention, this piece was at most minimal.

eLearn Tablets

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

10.2 Additional LASSO Details

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

10.3 Additional Testing Details

Our project-specific exams were designed by subject experts, not particularly involved with the design and implementation of the program, to cover the standard curriculum and be conceptual and less prone to rote memorization, criticisms of the PEC exam and other similar provincial exams in Pakistan (School Education Department 2013; Burdett 2017). Each test included 80 grade level questions. No study teachers had access to the test and students were not allowed to keep any testing materials. Appendix Figure A1 displays the baseline test score distribution for the project exam for eLearn Classrooms.

[Appendix Figure A1 about here]

10.4 Additional Figures and Estimations

Appendix Table A1 uses alternative controls than those that appear in the main paper.

[Appendix Table A1 about here]

Appendix Table A2 provides the subject specific test score estimates.

[Appendix Table A2 about here]

Appendix Table A3 provides the Lee (2009) bounds for the achievement estimates.

[Appendix Table A3 about here]

Appendix Table A4 tests for changes in additional teacher inputs.

[Appendix Table A4 about here]

Appendix Table A5 tests for heterogeneity in effects by student baseline test score, school baseline test score, and school gender. The first two are discussed above in Section 7. When considering gender, for the Classrooms intervention project test, the main effect is positive, and even though the interaction effect is negative, we reject at the 10 percent level that the sum is 0–the intervention increased test scores for both male and female students (Appendix Table A5, Panel C, column 1). In contrast, for the PEC score, the main effect is not statistically significant, the interaction is positive, and we reject (at the 1 percent level) that the coefficients sum to 0. The combined project and PEC score has both a positive (and statistically significant) main effect and (statistically insignificant) interaction effect. We reject at 1 percent that the overall effect on female students is 0. Therefore, while the intervention was not designed to favor students of a particular gender, instead providing expert content to assist all students, female students' scores increased on the PEC exam but males students' scores did not.[50]

For the Tablets intervention, the main effect is negative and statistically significant, the interaction is negative and insignificant, and we reject that the overall effect on female students is 0. (Table A5, Panel C, column 1).

As the heterogeneous score gains by gender in the Classrooms intervention 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 Classrooms data. At the school

[50]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 under performed on the high-stakes exams. In our setting, female students gain more from the intervention on the high-stakes PEC exam.

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.[51]

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

While the two differences in other outcomes 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 females. 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 men. Therefore, while we cannot

[51] The findings of heterogeneity by baseline test score holds within gender.

directly test the mechanisms, these findings are consistent with the importance of a gender matching role model in future aspirations, which potentially led parents of boys to be more likely to visit school and boys to aspire to higher education.

[Appendix Table A5 about here]

Appendix Figure A2 shows the non-parametric treatment effects. We examine heterogeneity non-parametrically as a function of baseline test scores. The test score lines for treatment and control lines are kernel-weighted locally-smoothed means of the endline test scores at each percentile of the baseline test-score distribution. The treatment effect for each baseline percentile is calculated as the difference between the treatment and control. The 95% confidence intervals are estimated using bootstrapping. The x-axis is the percentile of the residual of a regression of baseline test score on student and school characteristics (the same characteristics from the handpicked controls regression). The y-axis is the residual of a regression of the endline score on the same student and school characteristics.

[Appendix Figure A2 about here]

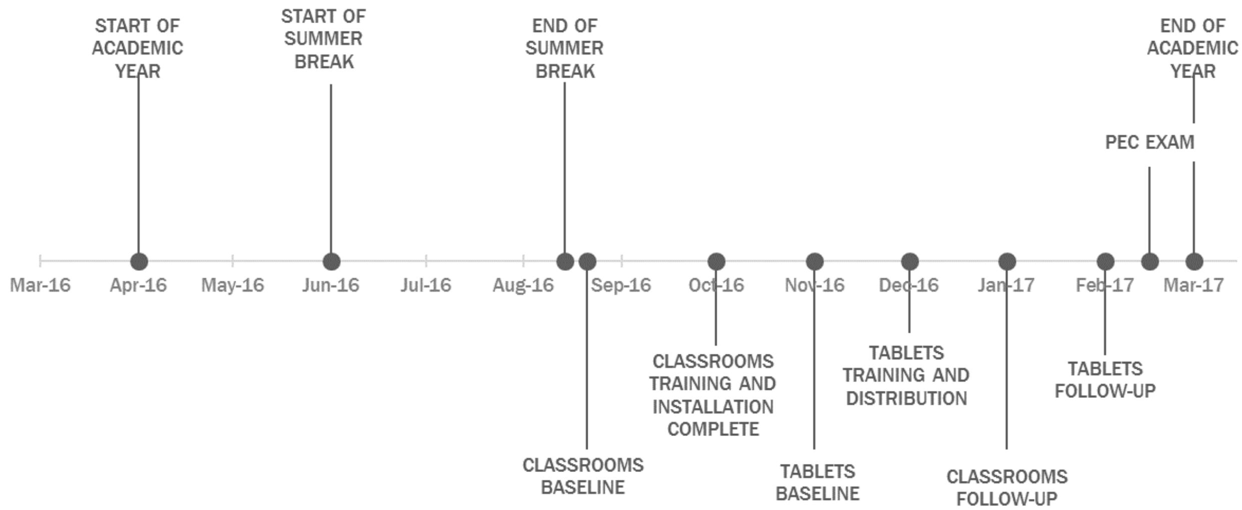
Appendix Table A6 contains separate estimates the effect sizes for questions by quartile of question difficulty. First, we used IRT to determine the difficulty of each question in the baseline. Then, we created a separate test score for each student based only on the questions in each quartile of difficulty, i.e. the easiest quartile, the second quartile of difficulty, the third quartile of difficulty, and the most difficult questions. We estimated the treatment effect separately for each newly created test score.

[Appendix Table A6 about here]

Appendix Table A7 tests for heterogeneity by whether one of the teachers in the school had below-median years of teaching experience (Column 1) or whether at least one of the study teachers had a teacher peer who taught the same grade-level and subject.

[Appendix Table A7 about here]

Figure 1: Study and Academic Year Timeline



Notes: Academic year milestones appear above the line with intervention events below the line.

Figure 2: Graphical Conceptual Framework

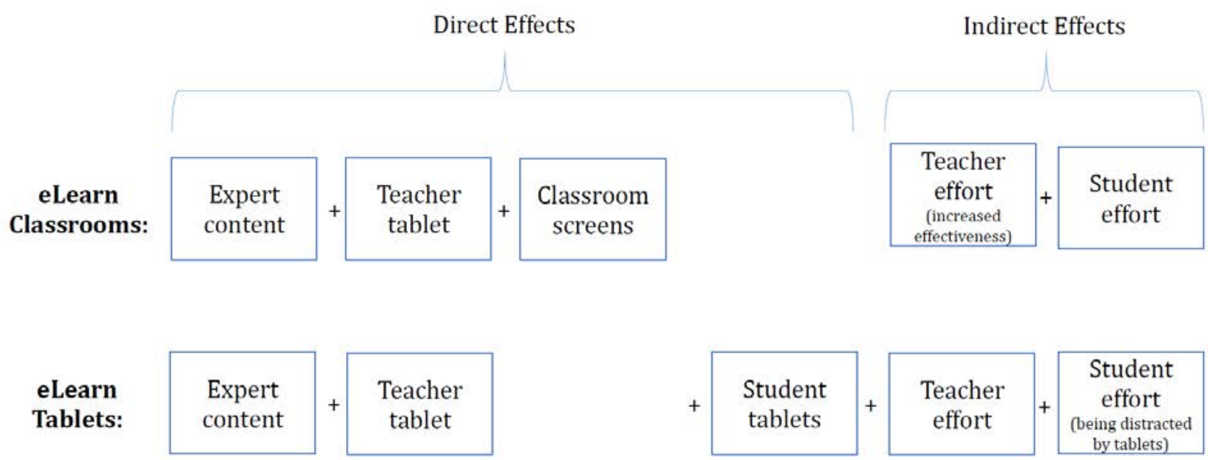
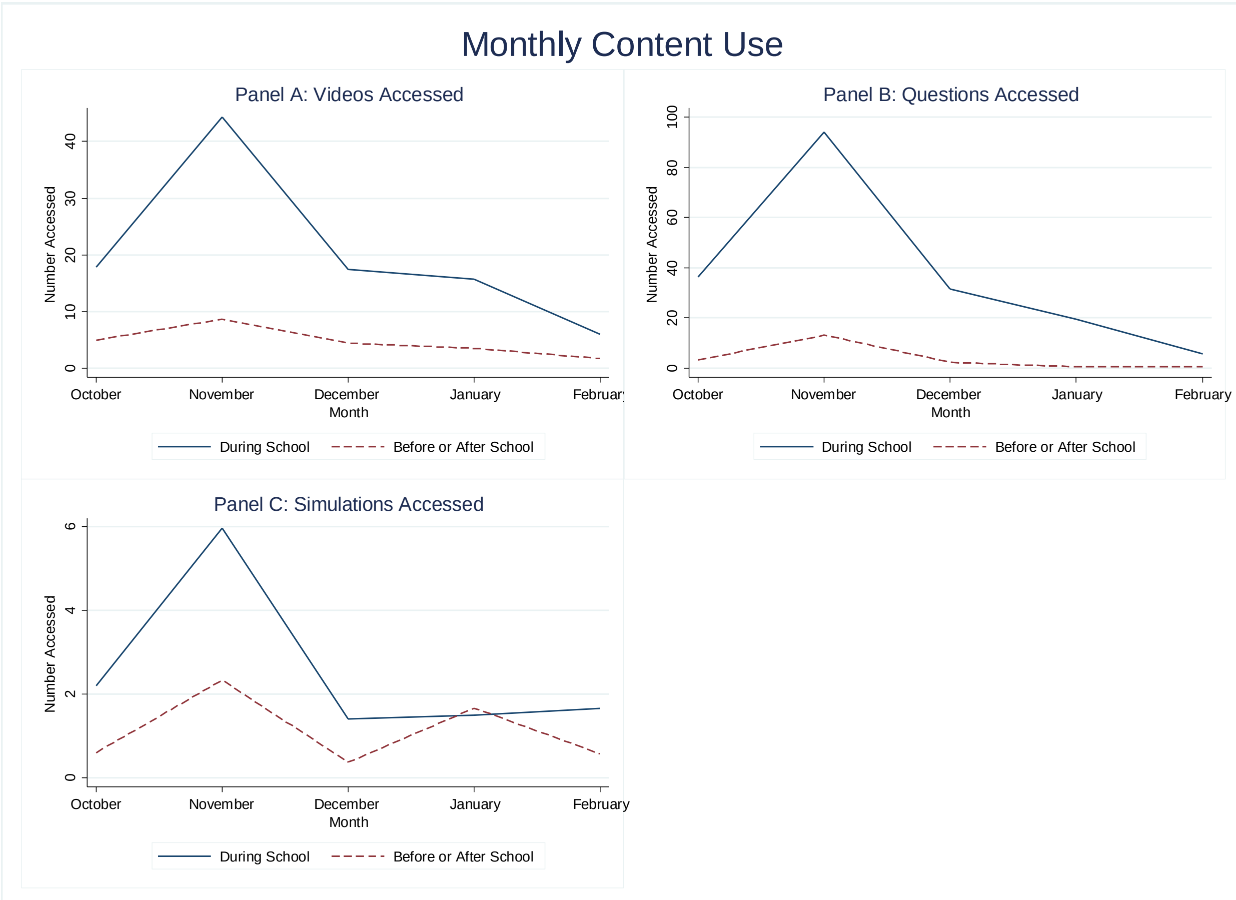


Figure 3: Monthly Content Use by Teachers in the Classrooms Intervention



Notes: Based on data collected by tablets in the Classrooms intervention. The program was implemented in October.

Table 1: Summary Statistics

	eLearn Classrooms			eLearn Tablets		
	Treatment	Control	Difference T-C	Treatment	Control	Difference T-C
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Student Characteristics						
Combined Math and Science Score	-0.056	0.068	-0.12	0.019	-0.008	0.027
	(0.99)	(1.01)	(0.190)	(1.10)	(0.96)	(0.28)
Age	13.90	13.87	0.03	12.12	12.17	-0.05
	(1.24)	(1.23)	(0.102)	(1.38)	(1.43)	(0.13)
Days Absent Last Month	1.50	1.17	0.327*	1.40	1.66	-0.25
	(2.41)	(1.79)	(0.173)	(2.03)	(3.40)	(0.20)
Has a Computer at Home	0.43	0.41	0.02	0.23	0.22	0.01
	(0.50)	(0.49)	(0.0484)	(0.42)	(0.41)	(0.06)
Mother Has No Formal Schooling	0.35	0.33	0.02	0.36	0.44	-0.08
	(0.48)	(0.47)	(0.0576)	(0.48)	(0.50)	(0.07)
Father Has No Formal Schooling	0.17	0.20	-0.02	0.22	0.26	-0.05
	(0.38)	(0.40)	(0.0341)	(0.41)	(0.44)	(0.05)
Panel B: Teacher Characteristics						
Has an Advanced Degree	0.75	0.80	-0.05	0.62	0.67	-0.05
	(0.44)	(0.40)	(0.0806)	(0.49)	(0.47)	(0.12)
Years of Teaching Experience	10.72	10.67	0.05	15.03	14.52	0.51
	(8.52)	(9.10)	(1.727)	(11.65)	(10.52)	(2.88)
Minutes per Day Planning Lessons	40.67	33.64	7.03	4.86	5.87	-1.01
	(33.75)	(27.85)	(5.477)	(12.80)	(12.85)	(3.02)
Use Technology to Prepare for Class	0.58	0.60	-0.02	0.62	0.57	0.05
	(0.50)	(0.49)	(0.106)	(0.49)	(0.50)	(0.14)
Use Technology in Class	0.14	0.17	-0.03	0.48	0.47	0.01
	(0.35)	(0.38)	(0.0650)	(0.51)	(0.50)	(0.15)
Panel C: School Characteristics						
Total Enrollment in Grade	63.10	63.21	-0.11	60.1	52.1	7.97
	(16.36)	(13.52)	(3.901)	(42.3)	(29.8)	(10.2)
Sections in Grade	1.40	1.35	0.06	1.55	1.40	0.15
	(0.50)	(0.48)	(0.128)	(0.8)	(0.8)	(0.2)
School Has a Computer Lab	0.90	1.00	-0.100*	0.60	0.65	-0.05
	(0.31)	(0)	(0.0557)	(0.5)	(0.5)	(0.1)

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. Enrollment and number of sections for relevant grade: grade 8 for Classrooms and Grade 6 for Tablets.

Table 2: Achievement Effects

	eLearn Classrooms				eLearn Tablets
	Standardized Combined Math and Science Test Score				Standardized Combined Math and Science Test Score
	Project	PEC	Combined Project and PEC	Pass the PEC	
	(1)	(2)	(3)	(4)	(5)
Panel A: Limited Controls					
Treatment	0.256*	0.221*	0.269**	0.0384	-0.420**
	(0.135)	(0.129)	(0.119)	(0.0277)	(0.163)
Observations	2,622	2,766	2,463	2,766	3,058
R-Squared	0.13	0.25	0.20	0.06	0.36
Panel B: LASSO Controls					
Treatment	0.296**	0.269***	0.263**	0.0468**	-0.426***
	(0.130)	(0.103)	(0.127)	(0.0234)	(0.149)
Observations	2,622	2,766	2,463	2,766	3,058
Average Control Group Change or Mean	0.49	0.00	0.00	0.92	0.45

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

Table 3: Student Attendance and Attrition

	eLearn Classrooms				eLearn Tablets	
	Present at Follow-up (Took follow-up exam)		Matched to and Completed PEC		Present at Follow-up (Took follow-up exam)	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0407**	0.0496**	-0.0142	-0.0094	0.0333	0.0306
	(0.0199)	(0.0218)	(0.0104)	(0.0106)	(0.0399)	(0.0354)
Treatment X Baseline Score		0.0077		0.0133		0.0342
		(0.0157)		(0.0111)		(0.0284)
Observations	2,999	2,999	2,999	2,999	3,614	3,614
Control Group Mean	0.85		0.93		0.85	

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. Linear probability models. Includes all students who took the baseline test.

Table 4: Teacher Attendance

	Portion Present			
	eLearn Classrooms		eLearn Tablets	
	(1)	(2)	(3)	(4)
Treatment	0.0114*	0.0221**	-0.00107	0.0146
	(0.00671)	(0.0100)	(0.0113)	(0.0117)
Treatment X Months of Treatment		-0.00555*		-0.0144
		(0.00336)		(0.00910)
Observations	274	274	189	189
Control Group Mean	0.94		0.93	

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. The portion of all teachers present in the school during an unannounced spot check. Measured monthly for each school.

Table 5: Changes in Inputs - Technology and Training--eLearn Classrooms

	Teacher Uses Technology		Number of In-service Trainings This Year
	To Prepare for Lessons	In the Classroom	
	(1)	(2)	(3)
Treatment	0.333***	0.780***	0.326**
	(0.0728)	(0.0674)	(0.128)
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: Achievement Effects--Alternative Controls

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

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

Appendix Table A2: Subject Specific Achievement Effects

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

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

Appendix Table A3: Lee (2009) Bounds

	Standardized Combined Project Math and Science Test			
	eLearn Classrooms		eLearn Tablets	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)
Treatment	0.271**	0.305**	-0.467***	-0.397***
	(0.127)	(0.134)	(0.155)	(0.152)
Observations	2,551	2,551	2,939	2,939

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 A4: Changes in Other Teacher Inputs--eLearn Classrooms

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

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

Appendix Table A5: Heterogeneous Achievement Effects

	Standardized Combined Math and Science Test Score			
	eLearn Classrooms			eLearn Tablets
	Project	PEC	Combined Project and PEC	
	(1)	(2)	(3)	(4)
Panel A: By Student Baseline Test Score				
Treatment	0.346**	0.241*	0.237*	-0.555***
	(0.169)	(0.146)	(0.128)	(0.215)
Treatment X 2nd Quartile	-0.287**	-0.0393	-0.0864	-0.222
	(0.139)	(0.0808)	(0.0893)	(0.161)
Treatment X 3rd Quartile	-0.178	-0.00229	-0.0669	-0.238
	(0.151)	(0.0997)	(0.0968)	(0.257)
Treatment X Top Quartile	-0.0595	0.109	-0.0161	-0.123
	(0.291)	(0.160)	(0.145)	(0.334)
Observations	2,622	2,766	2,463	3,058
p-values of F-tests of coefficients on Treatment + Treatment X Quartile sum to 0				
Quartile 2	0.70	0.20	0.27	0.00
Quartile 3	0.19	0.04	0.16	0.00
Top Quartile	0.24	0.00	0.05	0.01
Panel B: By School Baseline Test Score				
Treatment	0.406	0.761***	0.783***	-0.469**
	(0.278)	(0.227)	(0.237)	(0.226)
Treatment X 2nd Quartile	-0.384	-1.214***	-1.283***	-0.124
	(0.401)	(0.323)	(0.328)	(0.405)
Treatment X 3rd Quartile	0.192	-0.998***	-0.868***	1.634***
	(0.431)	(0.303)	(0.289)	(0.555)
Treatment X Top Quartile	-0.0654	-0.730**	-0.775**	0.780
	(0.441)	(0.326)	(0.314)	(0.532)
Observations	2,622	2,766	2,463	3,058
p-values of F-tests of coefficients on Treatment + Treatment X Quartile sum to 0				
Quartile 2	0.92	0.02	0.01	0.05
Quartile 3	0.05	0.13	0.53	0.02
Top Quartile	0.34	0.85	0.96	0.48
Panel C: By School Gender				
Treatment	0.328*	0.210	0.279*	-0.407**
	(0.197)	(0.197)	(0.166)	(0.184)
Treatment X Female School	-0.0622	0.0934	0.0891	-0.163
	(0.247)	(0.214)	(0.195)	(0.261)
Observations	2,622	2,766	2,463	3,058
p-value of F-test of coefficients on Treatment + Treatment X Female School sum to 0				
p-value	0.099	0.00	0.00	0.00
Average Control Group Change	0.49			0.45

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

Appendix Table A6: Achievement Effects by Question Difficulty

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

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

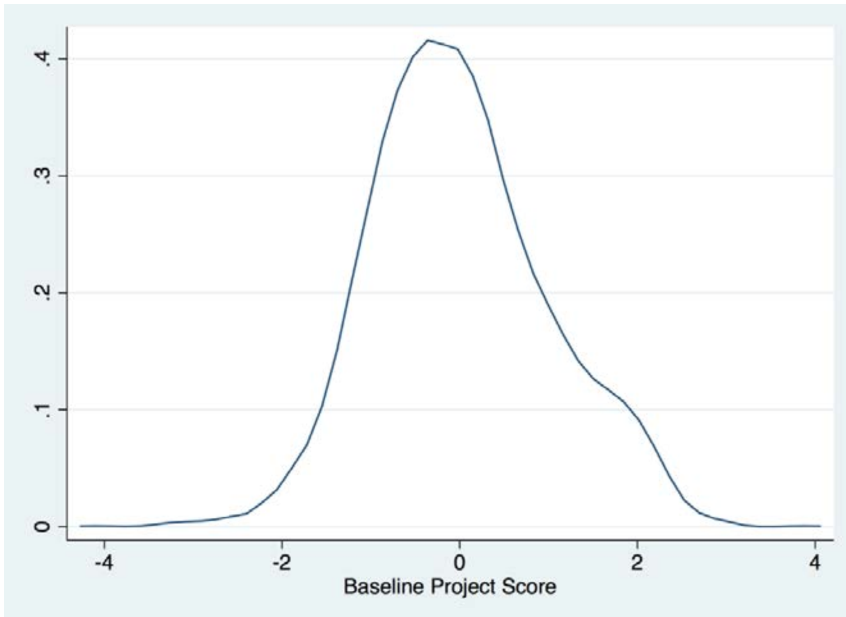
Appendix Table A7: Achievement Effects by Teacher Experience and Peers - eLearn Classrooms

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

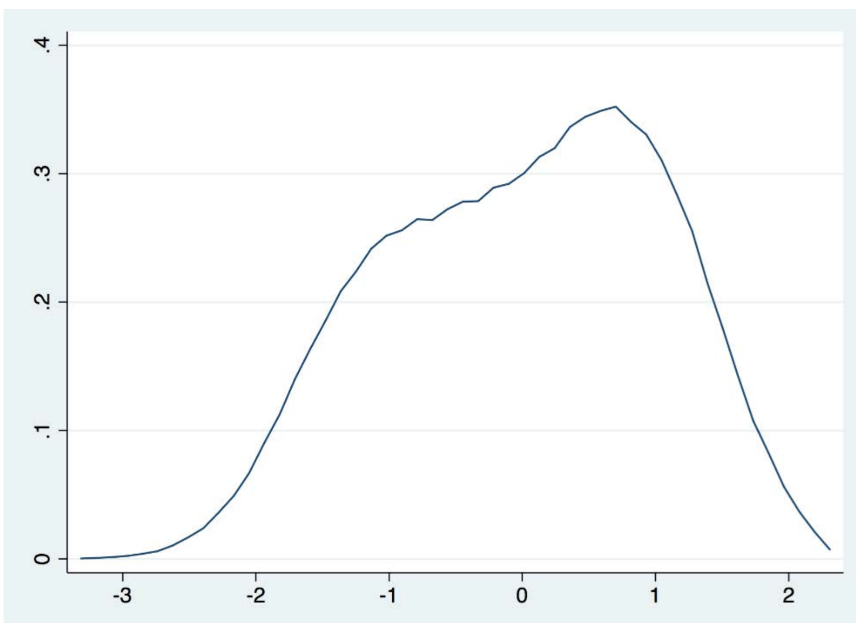
Notes : * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the school level appear in parenthesis. Includes all students who took the test at both baseline and endline. Controls include strata, student baseline test scores, and those selected by the LASSO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

Appendix Figure A1: Baseline Test Score Distributions

Panel A: eLearn Classrooms



Panel B: eLearn Tablets

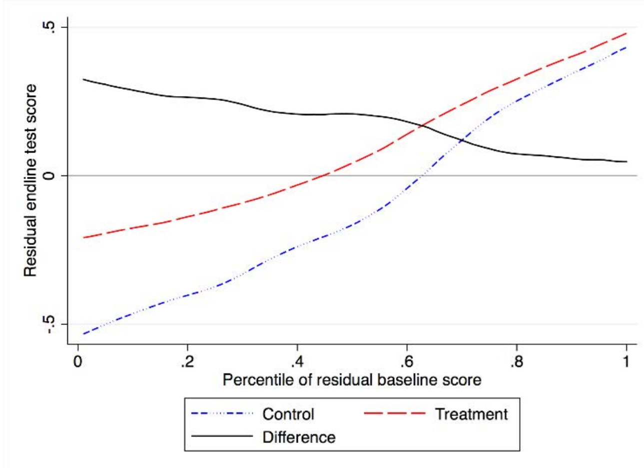


Baseline Project Score

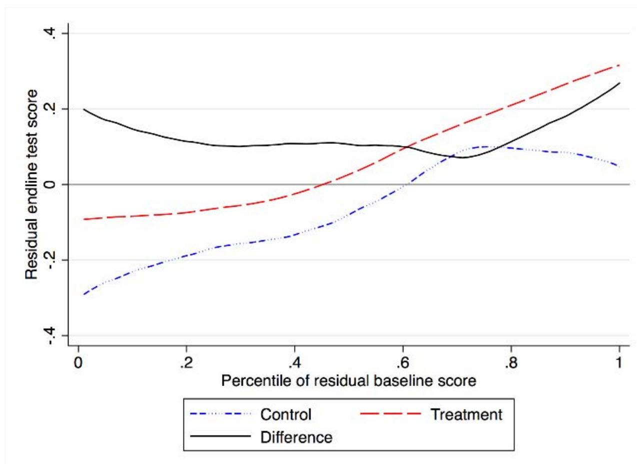
Notes: Baseline test score distributions.

Appendix Figure A2: Nonparametric Treatment Effects

Panel A: eLearn Classrooms Project Test



Panel B: eLearn Classrooms PEC Exam



Panel C: eLearn Tablets