American Economic Journal: Economic Policy 2022, 14(2): 61–90
https://doi.org/10.1257/pol.20200713

Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not†

By Sabrin Beg, Waqas Halim, Adrienne M. Lucas, and Umar Saif*

Using two RCTs in middle schools in Pakistan, we show that 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 percent more than the control group, after 4 months of exposure. Students and teachers increased their attendance, and students were more likely to pass the high-stakes government exams. By contrast, providing similar content to students on personal tablets decreased student scores by 0.4 SD. The contrast between the two effects shows the importance of engaging teachers and the potential for technology to do so. (*JEL* I21, I28, J45, O15, O30)

Improving students' knowledge and success on high-stakes exams while working with 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. 2008; 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: bypassing teachers entirely through the use of tutors, assistants, or

*Beg: Department of Economics, University of Delaware (email: sbeg@udel.edu); Halim: IT University Lahore (email: waqashalim@gmail.com); Lucas: Department of Economics, University of Delaware, NBER, J-PAL, and CGD (email: alucas@udel.edu); Saif: UNESCO Chair ICTD (email: umar.saif@cantab.net). C. Kirabo Jackson was coeditor for this article. 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."

†Go to https://doi.org/10.1257/pol.20200713 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

replacement technology, or engaging teachers through extensive training and monitoring, which is typically only successful with careful nongovernmental 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 with 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. eLearn's videos addressed potential deficiencies in teacher capacity, teaching both the teacher and the students the material, and demonstrated a clear and engaging teaching style for teachers to replicate. 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, teachers could project the content to a project-installed LED screen. The intervention was designed to complement existing practices and teachers, not to add employees or to 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. 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.[1] The two interventions had opposite effects. In our preferred specification, eLearn Classrooms increased student achievement by 0.30 standard deviations (SD) and eLearn Tablets decreased student achievement by about 0.43 SD on combined math and science exams designed for the projects. This gain for eLearn Classrooms was 60 percent more than the change in the control group score over the same period, and the decrease for eLearn

[1] 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.

Tablets was 95 percent of the learning that occurred in the control group. Students in the eLearn Classrooms intervention also scored 0.27 SD higher on the combined math and science sections of the standardized test that students take at the end of grade 8. Combining the project and standardized test scores in a single measure of student achievement, the eLearn Classrooms intervention improved test scores 0.26 SD.[2] Further, the eLearn 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. Teachers and students in eLearn Classrooms treatment schools also increased their attendance, unlike those in the eLearn Tablets intervention, and teachers in the eLearn Classrooms treatment schools also increased other self-reported effort.[3]

The positive effect sizes for eLearn Classrooms are the sum of both students' and teachers' learning from the videos and any modifications that teachers made to their teaching practices. Through a conceptual framework, we argue that the difference between the empirical findings of the two interventions was likely due to complementarities between the classroom screens and the digital content and between the screens and the teachers' efforts, which led teachers to improve their teaching practice beyond the minutes of the video. The increased teacher attendance and effort and the heterogeneity findings for the eLearn Classrooms program support the claim of complementarity with teacher effort. 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 for whom the content alone—without a teacher—was potentially the least learning-level appropriate. Finally, we estimate that test score 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, providing technology that augmented the existing classroom experience increased learning, while similar content with less integration decreased learning. Positive effects from a small change that engaged the existing teacher contrast with the idea implicit in many studies, and in the eLearn Tablets intervention, that existing teachers must be bypassed by technology or assistants or substantially retrained, monitored, or incentivized to increase student learning.[4] Further, teachers

[2] This improvement is equivalent to increasing teacher value added by 0.7 to 1.8 SD (Hanushek and Rivkin 2010; Jackson, Rockoff, and Staiger 2014; Azam and Kingdon 2015; Bau and Das 2020).

[3] Based on existing estimates of the relationship between teacher attendance and student achievement, the change in teacher absenteeism is too small to generate the full achievement effects. See Section VB.

[4] For example, see Banerjee et al. (2007, 2017) and Duflo, Kiessel, and Lucas (2021) for assistants; see Banerjee et al. (2007); Lucas et al. (2014); and Banerjee et al. (2017) for training and monitoring; see Banerjee et al. (2007); Linden (2008); Carrillo, Onofa, and Ponce (2010); Mo et al. (2013); Lai et al. (2013); Lai et al. (2015); Lai et al. (2016); and Muralidharan, Singh, and Ganimian (2019) for computer assisted learning; see Jamison et al. (1981); Naslund-Hadley, Parker, and Hernandez-Agramonte (2014); and Johnston and Ksoll (2017) for replacement of teachers through radio broadcasts, audio CDs, or satellite broadcasts; see Duflo, Dupas, and Kremer (2015); Barrera-Osorio and Raju (2017); Mbiti et al. (2019); and Gilligan et al. (2019) for incentives or altered contract schemes.

in the eLearn Classrooms intervention increased their attendance instead of using the assistance to facilitate their absenteeism as in Banerjee et al. (2017).[5] The success of the eLearn Classrooms intervention shows the potential for well-designed technology to create its own impetus for use (Jackson and Makarin 2018), to act as a commitment device for teachers to view the content themselves, to provide a highly effective virtual colleague teaching the same content and grade level (Jackson, Rockoff, and Staiger 2014; Jackson and Bruegmann 2009; Papay et al. 2020), and to integrate questions that provided instant feedback on students' knowledge (Berry et al. 2020). Even though the eLearn Tablets intervention contained materials beyond the textbook, was integrated with the curriculum, could be used at home and school, and solved some of the deficiencies of computers found in previous studies, the intervention decreased test scores, potentially because students lacked guidance for the tablets' effective use, used them for nonscholastic activities, or did not have appropriate spaces at home to use them effectively (Fairlie and Robinson 2013; Escueta et al. 2017; Government of Pakistan 2014).[6]

Second, the two interventions provide additional evidence on the complementarities in the production of learning between elements within an intervention and between the elements and teachers' efforts. Due to complementarities among pieces of a literacy intervention, Kerwin and Thornton (2021) found that when an intervention did not provide writing slates, teachers placed less emphasis on writing, decreasing writing scores. Across our two interventions, the LED screens were the crucial piece, having a strong complementarity with the content and teacher effort—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, or model their additional classroom time after these two effective components.

Third, both eLearn Classrooms and eLearn 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 scaling up a successful program.

Fourth, this was middle-school grade-level content. Grade-level content that engaged the teachers improved grade-level competencies.[7] 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).[8]

Taken together, we show that technology that engaged the teacher increased student achievement while similar content on students' personal devices did the

[5] 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 2018; Bau and Das 2020).

[6] The importance of integration and avoidance of potential pitfalls builds on Barrera-Osorio and Linden (2009); Malamud and Pop-Eleches (2011); Fairlie and Robinson (2013); Mo et al. (2013); Beuermann et al. (2015); Bando et al. (2017); and Cristia et al. (2017).

[7] 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).

[8] Muralidharan, Singh, and Ganimian (2019) focused on the foundational skills of middle school students.

opposite. 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 retraining. 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 deploying high-quality teaching to classroom screens, student test scores increased due to both the content and its complementarity with teachers' efforts that led to increased teacher effectiveness. Providing the same material on tablets to students and teachers without the ability to project it to the class decreased achievement.

I. 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 and 10 and grades 11 and 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 five 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. 2008, 2020; Andrabi, Das, and Khwaja 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, Iqbal, and Sohail 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 in teaching middle school 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.

II. 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 online Appendix A1. We then provide a conceptual framework to explain how the direct and indirect effects differ by intervention.

A. *eLearn*

The overall intervention, eLearn, was designed and implemented by our government partners 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 an expert teacher 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. The video lessons were not designed to be total substitutes for the teachers. 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 nine minutes for eLearn Classrooms and four minutes for eLearn Tablets.[9] These lectures modeled effective teaching through the use of extensive visual aids, displaying phrases that signaled and reinforced important concepts, using follow-up questions, and providing narration to accompany demonstrations—methods that could be replicated by an in-person teacher with effective chalkboard or whiteboard management. Furthermore, the tablets contained about three multiple-choice assessment questions and their answers, to be used after each video. Paired with some videos were an additional three to five minutes of multimedia content, e.g., an interactive animation of photosynthesis. The tablets also contained digital versions of the official textbooks, which almost all students reported having already in hard copy (around 98 percent) even though fewer 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.[10] During the training, teachers received a handbook with teaching tips, suggestions on how to introduce and follow up after the videos, and the multiple choice questions from the tablet. Within each treatment arm, all treatment teachers

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

[10] 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 in 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; 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, in-service teacher training models that increased student achievement included material provision (e.g., Lucas et al. 2014; Kerwin and Thornton 2021; Duflo, Kiessel, and Lucas 2021; Beg, Fitzpatrick, and Lucas 2021). Therefore, the combination of materials—in this case, technology—and training teachers to use them is the most meaningful bundle to evaluate to improve student achievement.

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 the precise details explained in the next two subsections. The parallel implementation of the programs that we evaluated was designed to inform the larger scale-up of the programs, which was to start in 2018.[11]

B. *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, preloaded tablets and classrooms received 40 inch LED television screens. Teachers could use these tablets to watch the videos themselves when preparing for class and project the content on the installed screens. The screens were installed above the existing chalkboard or whiteboard, enabling teachers to continue to use the board in an interactive way with the videos. 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.[12] An additional component of the intervention was designed to engage students and parents at home, but was implemented marginally, at most. (See online Appendix Section A1 for additional details.)

The teacher trainings, hardware installation, and tablet distribution were finished by the start of October, after our 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 four months between the materials' distribution and follow-up testing.[13] Figure 1 displays the study timeline and academic-year timeline for both interventions. See Section IVB for additional details on data collection.

C. *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 eLearn Classrooms intervention, teachers could have easily outsourced their own explanations by relying on those on the tablets, but

[11] This intervention was designed to mimic exactly the planned, full rollout. 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 in online Appendix Section A1.

[12] These teams were neither designed nor equipped to support or improve teaching practices.

[13] As originally designed, the student baseline testing was to occur prior to the June-to-August holiday, and training and installation should have occurred during the holiday break. Once implementation delays were apparent, we delayed the baseline testing, as well.

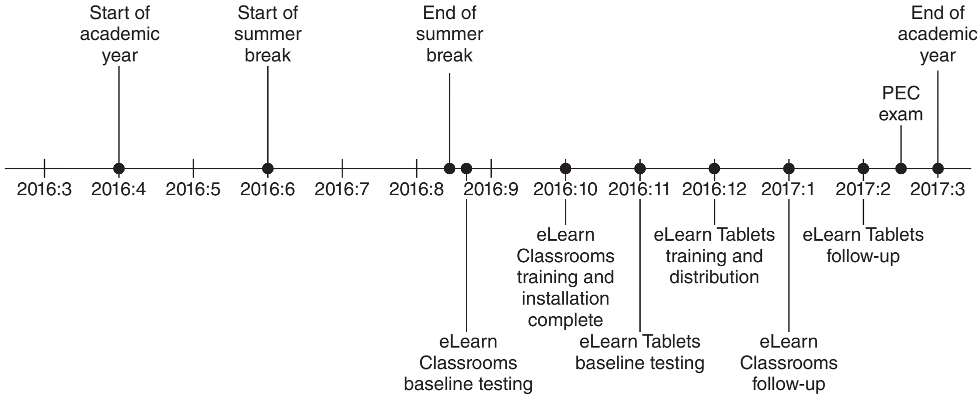

Figure 1. Study and Academic Year Timeline

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

teachers would not have been able to observe what portions of the lessons or methods of explanations students found particularly engaging or whether the students were viewing the correct content, or to receive the instant feedback from students pondering and responding to follow-up questions.

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 the videos to students to watch at home as homework. Tablets were open for students to use for noneducational 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 the devices were working correctly and to answer questions about the devices, but the staff did not provide guidance on how children should be using the devices to support learning. Further, households received a phone number for a helpline that they could call for technical support.

Parents, especially those with limited science and math knowledge, 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. As a typical practice, parents do not often 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 help them use the tablet correctly (Government of Pakistan 2014).

The eLearn Tablets timeline was designed to be similar to the eLearn 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.

D. *Conceptual Framework*

The effects of the two interventions on student achievement is the sum of the direct effect of the intervention components plus any intervention-induced behavior

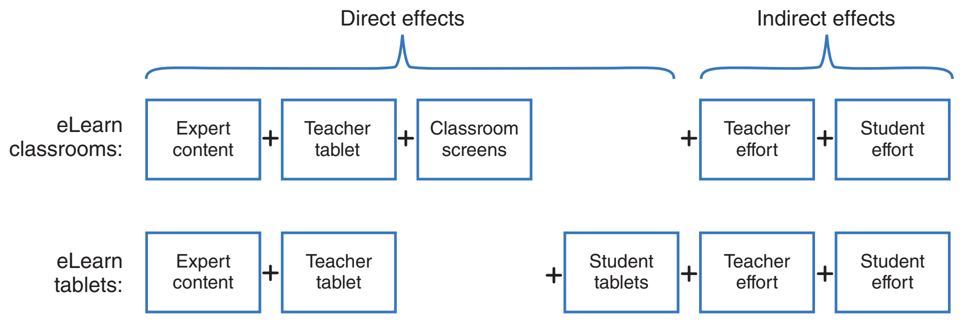

Figure 2. Graphical Conceptual Framework

changes by students or teachers.[14] Figure 2 displays the similarities and differences between the interventions graphically. Both interventions had expert content available and provided teachers a tablet on which they could view content themselves. The eLearn Classrooms intervention had classroom screens, while the eLearn Tablets intervention provided tablets directly to students. The direct effect on student knowledge was likely similar across these two modalities of viewing the content. Both interventions had the potential to induce differential behavior changes in teachers and students. The total effect of each intervention is the sum of the direct effects of the intervention components and any resulting indirect behavioral changes, positive or negative.

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 to the status quo control group. First, when comparing the eLearn Classrooms intervention to the control group, the magnitude of the effect is the sum of the direct and indirect effects from the eLearn 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 I); therefore, whether this intervention would be successful was an empirical question. Second, the effect of the eLearn Tablets intervention relative to the control group is the sum of the direct and indirect effects of that intervention. As with eLearn Classrooms, while the eLearn Tablets intervention attempted to remedy some of the pitfalls of previous technology interventions, whether it would be successful was 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 elucidate the sources of the positive or negative treatment effects that we found relative to the control group. Almost all of the intervention components of the two interventions were similar: the availability of the content, the teachers' tablets and training, and students' ability to watch the content

[14] In a basic achievement framework, student achievement is function of both inputs and effort, which can also react to changes in inputs. Formally, $A = f(x, e(x))$, 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 $dA/dx = (\partial f/\partial x) + (\partial f/\partial e) \times (\partial e/\partial x)$, where $\partial f/\partial x$ are the direct effects of the intervention and $(\partial f/\partial e) \times (\partial e/\partial x)$ are the indirect effects due to an effort response to the changes in inputs.

in the classroom (whether on a classroom screen or a personal tablet). Therefore, any differences between the impacts would have to come from the following sources: (i) students in the eLearn Tablets intervention could take their devices home; (ii) the components of the two interventions could have different complementarities with each other; (iii) the indirect effect due to student behavior changes could vary by treatment; and (iv) the indirect effect due to teacher behavior changes could vary by treatment. Both of these indirect effects could vary by treatment due to the complementarity between teacher or student effort and the specific pieces of each intervention.

If the effects across the two interventions are the same, then the sum of these four effects across the two interventions is equivalent.

If the eLearn Classrooms effects are larger, then its direct and indirect effects must exceed those in eLearn Tablets. The potential differential positive effect would come from the complementarity of the classroom screen and the digital content and the complementarity between the screen and teacher effort. The screens allowed teachers to learn from student reactions and model an effective teaching style during the nonvideo portions 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 from high-quality teaching (Jackson and Bruegmann 2009; Jackson, Rockoff, and Staiger 2014; Papay et al. 2020) and having an impetus to use new teaching techniques (Jackson and Makarin 2018). The screens in eLearn Classrooms could have facilitated this learning and mimicking by teachers, increasing the effectiveness of classroom lessons beyond the minutes of content that was available in both treatments.

By contrast, if the eLearn Tablets effects are larger, then the additional, direct benefit of being able to use the tablets at home plus related indirect benefits, such as students being additionally engaged with the material and potentially spending more time studying, were larger than the benefits from the eLearn Classrooms intervention.

III. Empirical Strategy

The primary conceptual difficulty in assessing the effects of various inputs into the education production function is the nonrandom 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 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

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

where y_{is} is outcome y for student i in school s, α is the constant term, $treatment_s$ is an indicator variable equal to one if the school was an eLearn treatment school (eLearn Classrooms or eLearn Tablets, depending on the sample), X_{is} is a vector of school-level 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 dummy variables in the X_{is} vector.[15] 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 controls 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 and strata.[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 interventions artificially improved scores only on this test, which makes the results for eLearn Tablets particularly disheartening. A distinguishing feature of eLearn Classrooms is that we have two types of tests for that intervention. We link students to their official PEC exam scores and test the impact of the intervention on these scores, 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, gender, and school characteristics that might have made the program more or less effective. 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' efforts and other inputs.

Our additional provision of technological inputs and teacher training could have crowded out other 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. In additional specifications, we test for these changes. 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 only, we further estimate the effect of the intervention on student self-reported effort and technology use.

Additionally, based on data collected from teachers, for eLearn 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, whether they used technology to teach their classes, whether they had been

[15] The strata are school and gender by district for eLearn Classrooms and school and gender by district by previous-year school test score tercile for eLearn Tablets.

[16] The variables we consider are listed in online Appendix Section A2.

[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 VB for more details.

part of any training, whether they held private tutoring sessions outside of school, whether they performed other official duties, and whether they 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 the 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.

IV. Sample Selection, Randomization, and Data

A. *Sample Selection and Randomization*

Our study took place within the 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 Punjabi 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 were similar to the average school in Punjab based on infrastructure and test scores. First, while the conditions of a boundary wall and electricity might be binding or indicate particularly wealthy schools in other contexts, in Punjab 93 percent of schools had electricity and 97 percent had 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 at the school level was stratified by district and gender.[21] 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 eLearn Classrooms. Twenty schools were randomly chosen to receive tablets and associated training, stratified by district, gender, and previous-year test score tercile.[22]

[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.

[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 attempted to fund a larger follow-up, funding agencies determined that compelling evidence from this study rendered a larger study unnecessary and would not fund it.

[21] Our government partners deemed randomization at a level lower than the school, e.g., classroom- or teacher-level, politically infeasible. Further, only about half of our schools had multiple sections of the relevant grade level.

[22] 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 multiphase rollout. The rollout did not occur due to the negative results of the evaluation.

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.

B. *Data*

We use two sources of data: primary data hand-collected at each study school and administrative data.

Primary Data Collection.—The data collected across the two interventions were similar, but on different schedules and with some additional details collected from the eLearn Classrooms schools. The baseline surveys solicited information from head teachers, i.e., school principals, the relevant grade's math and science teachers, and randomly selected students from the relevant grade present on the day of the baseline. All 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, which rely heavily on rote memorization.[23] In the eLearn Classrooms baseline, we tested 2,999 students and conducted 1,690 student interviews across 59 schools 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 eLearn Tablets baseline we tested and interviewed 3,614 students across 75 schools in November 2016.[24] Enumerators told schools that we would be visiting them again near the end of the school year, but they did not provide an exact date. We administered follow-up exams in January 2017 for eLearn Classrooms schools and February 2017 for eLearn Tablets schools. The same students were again tested, if present. In eLearn Classrooms schools, students, head teachers, and subject teachers were again surveyed. The school year ended in March. Figure 1 above provides the 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 eLearn Classrooms sample, we further test for changes in technology use and other student and teacher behaviors.

Administrative Data.—We used administrative data to create additional measures of student achievement for the students in the eLearn Classrooms arm. From the Punjab Examination Commission, we have the administrative student-by-subject-level PEC exam results and whether the student passed the PEC.[26] Students

[23] Additional test details appear in online Appendix Section A.A3.

[24] 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.

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

[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.

completed the PEC exams in mid-February. We merged these data with the students in our sample using students' and their fathers' full names.[27] Because we do not have item-level responses, these scores are scaled with a mean of 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 used 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 a 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 in the treatment and control schools.[29] Almost all of the measures are statistically indistinguishable by treatment status, with two exceptions at the 10 percent level in the eLearn 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 3 treatment schools not having a computer lab).[30] Given we are testing 14 outcomes across 2 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.

[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 VB. The unmatched 7 percent includes students who registered for the PEC but we were unable to match to records, who did not register for the exam, or who changed schools. Of those not matched, and therefore more likely to have changed schools or dropped out, about a quarter of students were not present at our follow-up survey. When considering only those present at our follow-up survey, 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 or what schools were in the treatment or control groups. 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 we believe it to be unlikely.

[29] Summary statistics and baseline balance checks for additional variables appear in online Appendix Table A1. Online Appendix Figure A1 shows the baseline test score distributions. Across both samples, only one student scored a zero on the baseline exam.

[30] 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 an 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.

Table 1—Summary Statistics

	eLearn Classrooms			eLearn Tablets		
	Treatment (1)	Control (2)	Difference T − C (3)	Treatment (4)	Control (5)	Difference T − C (6)
Panel A. Student characteristics						
Combined math and science score	−0.06	0.07	−0.12	0.02	−0.01	0.03
	(0.99)	(1.01)	(0.19)	(1.10)	(0.96)	(0.28)
Age	13.90	13.87	0.03	12.12	12.17	−0.05
	(1.24)	(1.23)	(0.10)	(1.38)	(1.43)	(0.13)
Days absent last month	1.50	1.17	0.33	1.40	1.66	−0.25
	(2.41)	(1.79)	(0.17)	(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)	(0.05)	(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)	(0.06)	(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)	(0.03)	(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)	(0.08)	(0.49)	(0.47)	(0.12)
Years of teaching experience	10.72	10.67	0.04	15.03	14.52	0.51
	(8.52)	(9.10)	(1.73)	(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)	(5.48)	(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)	(0.11)	(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)	(0.07)	(0.51)	(0.50)	(0.15)
Panel C. School characteristics						
Total enrollment in grade	63.1	63.2	−0.1	60.1	52.1	7.97
	(16.4)	(13.5)	(3.9)	(42.3)	(29.8)	(10.2)
Sections in grade	1.40	1.35	0.05	1.55	1.40	0.15
	(0.50)	(0.48)	(0.13)	(0.76)	(0.82)	(0.20)
School has a computer lab	0.90	1.00	−0.1	0.60	0.65	−0.05
	(0.31)	(0)	(0.06)	(0.50)	(0.48)	(0.13)

Notes: Columns 1, 2, 4, 5: Standard deviations appear in parentheses. Columns 3 and 6: Cluster-robust standard errors appear in parentheses. Panel C: Enrollment and number of sections are for the relevant grade (grade 8 for eLearn Classrooms and grade 6 for eLearn Tablets).

V. 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.[31]

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

Table 2—Achievement Effects

	eLearn Classrooms				eLearn Tablets
	Standardized combined math and science test score				
	Project (1)	PEC (2)	Combined project and PEC (3)	Pass the PEC (4)	Standardized combined math and science test score (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^2	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: Standard errors clustered at the school level appear in parentheses. 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, the 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: PEC high-stakes test. Column 3: PCA of project exam and PEC score. Column 4: Linear probability model.

A. *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. 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 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 eLearn 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 10 percent level. In column 3, we combine these tests into a single score measure. eLearn 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 eLearn Classrooms on the likelihood that a student passed the PEC exam and find a positive, statistically insignificant effect.[32] In column 5, we estimate the effect of the eLearn Tablets intervention on the combined

[32] 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.

math and science score. In contrast to the results in the first four columns, the eLearn Tablets intervention decreased test scores by 0.42 SD.[33]

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 with precision and to ensure we are not attributing underlying differences between the groups to a treatment effect. These results are somewhat farther from zero with stronger statistical significance. The eLearn 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)—all statistically significant at the 5 percent level, 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 eLearn Tablets intervention decreased student test scores by 0.43 SD when using the LASSO controls (panel B, column 5).[34,35]

During this same period, control group students increased their project test scores by 0.49 SD in the eLearn Classrooms sample and 0.45 SD in the eLearn Tablets sample.[36] Therefore, based on panel B, column 1, the eLearn Classrooms intervention increased the project test score by 60 percent relative to the gains in the control group. By contrast, the eLearn Tablets intervention eliminated almost all (95 percent) of the gains during the 3 instructional months of the intervention. These differences point toward strong positive effects of the classroom screens due to the complementarities between the installed screens and other behaviors. In the rest of this section we test for other changes.

B. *Attrition and Attendance*

The achievement results in Section VA 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.

[33] Since their test scores at the end of the year determine their eligibility for further schooling, grade 8 students could be more motivated than grade 6 students to learn. Therefore, the effect size of an intervention like eLearn Tablets on grade 8 students might be less negative.

[34] In additional specifications we included 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 online Appendix Table A2 and are similar to those presented in Table 2. Online Appendix Table A3 contains the subject-specific project exam effects for both eLearn Classrooms and eLearn Tablets. For the eLearn Classrooms 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. 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. eLearn Classrooms caused a smaller, statistically insignificant positive effect on the combined test score of the other subjects.

[35] 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.

[36] 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 tests. We then compare the means between the two rounds of test scores, displaying the difference in those means in the table.

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 × 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: Linear probability models, including all students who took the baseline test. Standard errors clustered at the school level appear in parentheses. Additional controls determined by LASSO.

As a first measure of observable effort, we estimate separately whether students who were present in the baseline were similarly present at the endline. While an interesting outcome in 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 of our binary outcomes, we use a linear probability model. The results appear in Table 3.

Students in the eLearn Classrooms treatment group were about 4 percentage points more likely to be present at the follow-up (column 1), demonstrating increased student effort and engagement as a result of the intervention.[37] Relative to the control group mean of 85 percent, this is about a 5 percent increase in the likelihood of being present. While finding that our intervention increased attendance is encouraging, 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 whether 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 small, statistically insignificant point value. Nevertheless, in online Appendix Table A4 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 in 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 treatment status.[38]

[37] As all study students were surveyed by the enumeration team regardless of treatment status, this differential change in attendance, which only occurs among the treated students in the eLearn Classrooms intervention, is not likely due to being part of a study.

[38] 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 whom we could not match.

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 × months of treatment		−0.00555		−0.0144
		(0.00336)		(0.00910)
Observations	274	274	189	189
Control group mean	0.94		0.93	

Notes: The portion of all teachers present in the school during an unannounced spot check, measured monthly for each school. Standard errors clustered at the school level appear in parentheses. Additional controls determined by LASSO.

The eLearn 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 online Appendix Table A4 columns 3 and 4, finding our results robust to this bounding exercise.

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 of Table 4, we estimate the effect of the eLearn 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).[39,40] This finding is in contrast to the concern expressed prior to the implementation that teachers would be 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.

As this is a monthly measure, we can test the evolution of teacher attendance over time. In column 2 we test whether this response changed over time. The main effect

[39] 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 equally observed by the enumeration team as a part of this study. Furthermore, they were equally observed by the PMIU as part of normal monthly data collection activities.

[40] 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 increases in teacher attendance. If sample treatment teachers attended school the average amount recorded in the control group, they could have at most increased their attendance by 6 percentage points. A 6 percentage point increase in attendance resulting in a 0.3 SD increase in test scores is well outside estimates on the effects of teacher attendance on student test scores from Das et al. (2007); Duflo, Hanna, and Ryan (2012); Cilliers et al. (2018); Herrmann and Rockoff (2012); and Gershenson (2016).

remains positive (and statistically significant at the 5 percent level). The point estimate on the interaction between treatment and months of treatment is negative and statistically significant at the 10 percent level. Therefore, the intervention appears to have increased teacher effort as measured by teacher attendance, but this increased attendance might have diminished over time.[41]

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

C. *Technology Use in eLearn Classrooms*

To measure whether and how teachers used the eLearn 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).[42] 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 two questions for each video played and one simulation for every nine videos played.[43] Therefore, schools used the intervention as a bundle, as intended.

Figure 3 displays the average number of videos, questions, and simulations accessed at the school level every month. Across all three items, use peaked in November—the first full month of the intervention—and during-school use (solid blue line) exceeded use outside of school hours (dashed red line) 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.[44]

Teachers 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).[45]

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

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

[43] Each video had, on average, three questions, and every four videos had an accompanying simulation. The R^2 on bivariate regressions between the number of videos and questions accessed is 0.60; between videos and simulations accessed it is 0.21, and between questions and simulations accessed it is 0.19.

[44] Teachers were encouraged to use all of the content, but the actual use was left to their discretion. From the experimental design, we cannot know whether the ideal amount of use is closer to the November peak or the December and January levels, 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 inexperience with the software and selecting videos in error.

[45] 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.

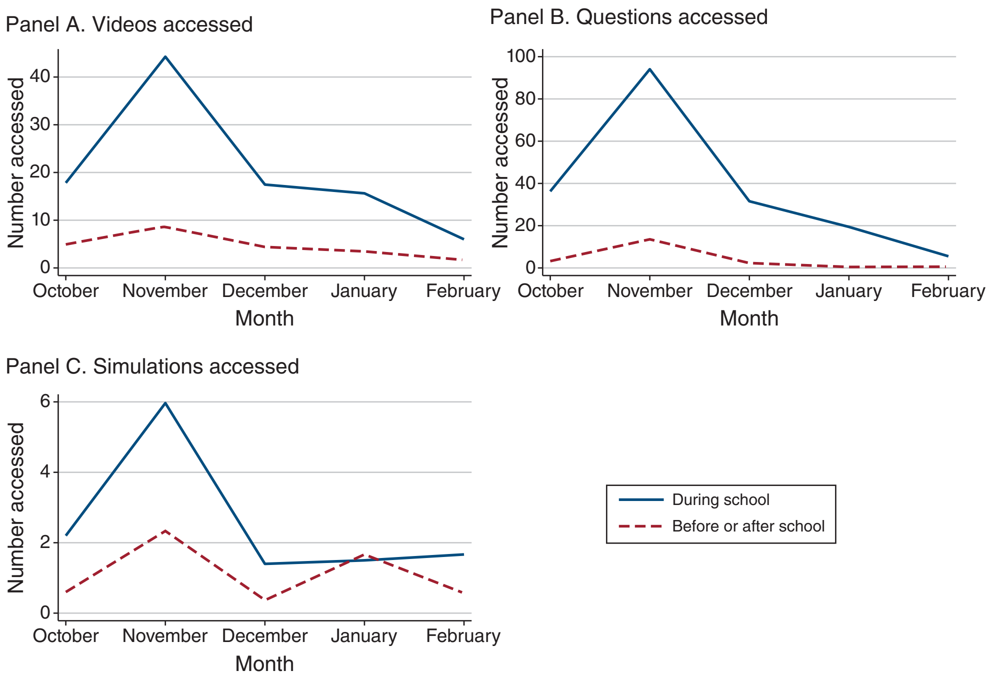

Figure 3. Monthly Content Access in the eLearn Classrooms Intervention

Notes: Average number of items accessed at the school level based on data collected by the tablets in the eLearn Classrooms intervention. The program was implemented in October.

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

	Teacher uses technology		Number of in-service trainings this year
	To prepare for lessons (1)	In the classroom (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: Standard errors clustered at the school level appear in parentheses. Additional controls determined by LASSO. Columns 1 and 2: Linear probability models.

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 teacher 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 were highly correlated with the data from the tablets. Therefore, teachers and students likely viewed this content together.

Implementing partner teams conducted two spot-check visits to each school during the intervention. During these visits, 83 percent of eLearn Classrooms schools used at least 1 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 eLearn Tablets intervention.

D. *Other Changes*

We additionally collected self-reported data on changes in other inputs and effort from teachers and students in the eLearn 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 eLearn Classrooms treatment spent 9.1 minutes more per day planning lessons (compared to a control group base of 58 minutes) and were 16 percentage points more likely (compared to a control group mean of 45 percent) 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 online Appendix Table A5. 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 students spent studying, the self-reported number of days students were absent in the last week, whether students 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).[46]

E. *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

[46] 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.

regressors in equation (1) with the project test, the PEC, and the combined project and PEC score as 3 separate outcomes.[47]

The point values for eLearn Classrooms appear in online Appendix Table A6, panel A, columns 1 to 3. Based on this specification, the eLearn 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 zero.[48] We find a similar U-shaped relationship for the PEC: students in the lowest and the top two quartiles had positive test score gains due to the intervention. Online Appendix Figure A2 shows the non-parametric treatment effects across the distribution.

The results for the eLearn Tablets intervention appear in Table A6, 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.

Since the interventions might be particularly suited to schools that were lower quality before implementation, one might like to know heterogeneity by school quality using school value added as a proxy. Unfortunately, we cannot calculate that with the 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 prioritized schooling. To test for heterogeneity on this margin, we include interactions between the treatment and the school-level score quartile as additional regressors in equation (1). These results appear in panel B of online Appendix Table A6. For eLearn Classrooms, students in third-quartile schools were the only ones who had statistically significant test score 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 eLearn 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 at home. The contrast between the two findings for the lowest-quartile schools—student scores increased in the eLearn Classrooms schools and decreased in the eLearn Tablets schools—shows the importance of and potential for teacher engagement to increase student test scores in schools that had been lower performers with likely fewer teaching and household resources.

Even though this intervention sought to improve test scores regardless of student gender, improving girls' outcomes is a particularly salient outcome in Pakistan, where only 38 percent of grade 8 students are girls (National Education Assessment System 2016). To test for heterogeneity by school gender, we replace the test score interaction with one for treatment times female school. Recall that all schools

[47] 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.

[48] In online Appendix Table A7, we estimate the effect of the intervention by the difficulty of the test questions (see online Appendix Section A4 for more details). The test score gains are statistically significant and about 0.2 SD for questions of below-median difficulty and statistically insignificant for questions of above-median difficulty.

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 online Appendix Section A4.

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 had test score increases statistically different from zero, while those in schools with more experienced teachers did not show test score increases (online Appendix Table A8, column 1). Further, test score gains were concentrated in schools in which teachers did not already have a grade-level subject peer teacher (online Appendix Table A8, column 2).

VI. Cost Effectiveness

A. *eLearn 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 eLearn Classrooms, where the intervention is at the classroom level and not the student level. Adding an additional student to an existing classroom in the eLearn 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 in our study, would be US$9 per student.[49] 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 online Appendix Section A1 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 fully 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 per student, a 200 school intervention would have an average cost of $20

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

per student, and a 1,000 school intervention would have an average cost of $11 per student.[50]

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, Conner, 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 1,000 school scale. The cost effectiveness at 200 schools exceeds the cost effectiveness of the other technology interventions reported in Kremer, Conner, and Glennerster (2013), and at 1,000 schools it exceeds the cost effectiveness of Muralidharan, Singh, and Ganimian (2019).[51] A program that linked school committees to local governments in Indonesia was more cost effective (Pradhan et al. 2011). 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 did not include any out-of-school time for students.

B. *eLearn Tablets*

Despite eLearn Tablets decreasing student achievement, we provide the cost estimates using the same method as we used for eLearn Classrooms. The largest difference between the two programs' costs were the costs of a tablet for each student, phone calls to students' households 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 $131, assuming a two-year tablet life. As implemented, the fixed costs were $20 per student at the 20 school scale. As with the eLearn Classrooms intervention, the larger the program, the more these fixed costs decrease per student. Even at the largest scale possible, the intervention would still have marginal costs of $131 per student, assuming schools of similar size to the schools in our intervention. The contrast between the two findings further shows that more expensive interventions are not necessarily better, and can be worse.

VII. 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, which digitized

[50] Removing the costs of the only partially implemented interactive content puts the costs at $15 per student at 100 schools, $12 per student at 200 schools, and $10 per student at 1,000 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.

[51] Muralidharan, Singh, and Ganimian (2019) do not provide cost effectiveness for 200 schools. 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 1,000 schools.

textbooks 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 eLearn 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 eLearn Tablets intervention support this hypothesis—the availability of similar content, adjusted for grade level, decreased student test scores by about 0.4 SD. Therefore, the content alone, without its integration into an effective teaching practice, was not sufficient to increase student test scores. Within the eLearn Classrooms intervention, schools with the lowest average baseline scores and students with the lowest baseline scores gained the most even though the content was at the grade level and potentially well above the learning level for these students, showing that interventions that improve teaching effectiveness, even if the content is at 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 United States (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 in schools in lower-income countries, nor was much known about improving learning outside of primary schools. 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. 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.

Returning to the conceptual framework outlined in Section IID, the difference in achievement effects are likely due to the complementarities between the installed classroom screen and teacher effort and the screen and the digital content. The classroom time in eLearn Classrooms likely became more effective and teachers and students demonstrated increased effort in other ways, while with eLearn Tablets students likely were distracted by their tablets either at home or school, displacing other useful academic activities or sleep. With eLearn Classrooms, teachers watched the videos and used the comprehension questions with the students, learning what students found effective from expert teachers and receiving immediate feedback on their students' comprehension. As we find positive effects for the eLearn Classrooms intervention and negative effects for the eLearn Tablets intervention, the change in classroom effectiveness was an important component of the overall effect of the eLearn 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 1,000 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, eLearn 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 existing teaching practices increased efforts by students and teachers and substantially increased middle school learning, especially for students of low baseline learning levels, potentially overcoming existing teacher capacity constraints.

REFERENCES

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2013. "Students Today, Teachers Tomorrow: Identifying Constraints on the Provision of Education." *Journal of Public Economics* 100: 1–14.

Andrabi, Tahir, Jishnu Das, Asim I. Khwaja, Selcuk Ozyurt, and Niharika Singh. 2020. "Upping the Ante: The Equilibrium Effects of Unconditional Grants to Private Schools." *American Economic Review* 110 (10): 3315–49.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, Tara Vishwanath, and Tristan Zajonc. 2008. *Learning and Educational Achievements in Punjab Schools* (*LEAPS*)*: Insights to Inform the Education Policy Debate.* Washington, DC: World Bank.

Azam, Mehtabul, and Geeta Gandhi Kingdon. 2015. "Assessing Teacher Quality in India." *Journal of Development Economics* 117: 74–83.

Bai, Yu, Di Mo, Linxiu Zhang, Matthew Boswell, and Scott Rozelle. 2016. "The Impact of Integrating ICT with Teaching: Evidence from a Randomized Controlled Trial in Rural Schools in China." *Computers and Education* 96: 1–14.

Bando, Rosangela, Francisco Gallego, Paul Gertler, and Dario Romero Fonseca. 2017. "Books or Laptops? The Effect of Shifting from Printed to Digital Delivery of Educational Content on Learning." *Economics of Education Review* 61: 162–73.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31 (4): 73–102.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235–64.

Barrera-Osorio, Felipe, and Alejandro J. Ganimian. 2016. "The Barking Dog That Bites: Test Score Volatility and School Rankings in Punjab, Pakistan." *International Journal of Educational Development* 49: 31–54.

Barrera-Osorio, Felipe, and Leigh L. Linden. 2009. "The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia." World Bank Policy Research Working Paper 4836.

Barrera-Osorio, Felipe, and Dhushyanth Raju. 2017. "Teacher Performance Pay: Experimental Evidence from Pakistan." *Journal of Public Economics* 148: 75–91.

Bau, Natalie, and Jishnu Das. 2020. "Teacher Value Added in a Low-Income Country." *American Economic Journal: Economic* Policy 12 (1): 62–96.

Beg, Sabrin, Waqas Halim, Adrienne M. Lucas, and Umar Saif. 2021. "Replication Data for: Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not." Harvard Dataverse, V2, UNF:6:fpaSKwID+YXoTg8dulNYzg== [fileUNF]. https://doi.org/10.7910/DVN/Y1UZTT.

Beg, Sabrin, Anne Fitzpatrick, and Adrienne M. Lucas. 2021. "Improving Public Sector Service Delivery: The Importance of Management." https://conference.iza.org/conference_files/edu_2021/lucas_a26750.pdf.

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): 629–40.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29–50.

Berry, James, Harini Kannan, Shobhini Mukherji, and Marc Shotland. 2020. "Failure of Frequent Assessment: An Evaluation of India's Continuous and Comprehensive Evaluation Program." *Journal of Development Economics* 143: 102406.

Beuermann, Diether W., Julian Cristia, Santiago Cueto, Ofer Malamud, and Yyannu Cruz-Aguayo. 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, Deon Filmer, Ezequiel Molina, and Jakob Svensson. 2019. "The Lost Human Capital—Teacher Knowledge and Student Achievement in Africa." World Bank Policy Research Working Paper 8849.

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

Cilliers, Jacobus, Ibrahim Kasirye, Clare Leaver, Pieter Serneels, and Andrew Zeitlin. 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, Pablo Ibarrarán, Santiago Cueto, Ana Santiago, and Eugenio Severín. 2017. "Technology and Child Development: Evidence from the One Laptop per Child Program." *American Economic Journal: Applied Economics* 9 (3): 295–320.

Das, Jishnu, Stefan Dercon, James Habyarimana, and Pramila Krishnan. 2007. "Teacher Shocks and Student Learning: Evidence from Zambia." *Journal of Human Resources* 42 (4): 820–62.

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas. 2021. "Experimental Evidence on Alternative Policies to Increase Learning at Scale." NBER Working Paper 27298.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics* 123: 92–110.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–78.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. "Education Technology: An Evidence-Based Review." NBER Working Paper 23744.

Fairlie, Robert W., and Jonathan Robinson. 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): 615–38.

Gilligan, Daniel O., Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M. Lucas, and Derek Neal. 2019. "Educator Incentives and Educational Triage in Rural Primary Schools." *Journal of Human Resources*. https://dx.doi.org/10.3368/jhr.57.1.1118-9871R2.

Glewwe, P., and K. Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." In *Handbook of the Economics of Education,* Vol. 5, edited by Eric A. Hanushek, Stephen J. Machin, and Ludger Woessmann, 653–743. Amsterdam: North Holland.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1 (1): 112–35.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics* 74 (1): 251–68.

Government of Pakistan. 2014. *National Assessment Report.* Islamabad: UNICEF.

Guo, Philip J., Juho Kim, and Rob Rubin. 2014. "How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos." In *Proceedings of the First ACM Conference on Learning at Scale Conference,* 41–50. New York: Association for Computing Machinery.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100 (2): 267–71.

Herrmann, Mariesa A., and Jonah E. Rockoff. 2012. "Worker Absence and Productivity: Evidence from Teaching." *Journal of Labor Economics* 30 (4): 749–82.

Jackson, C. Kirabo. 2010. "Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago." *Economic Journal* 120 (549): 1399–1429.

Jackson, C. Kirabo, and Elias Bruegmann. 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, Jonah E. Rockoff, and Douglas O. Staiger. 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Economics* 6: 801–25.

Jackson, Kirabo, and Alexey Makarin. 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., Barbara Searle, Klaus Galda, and Stephen P. Heyneman. 1981. "Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement." *Journal of Educational Psychology* 73 (4): 556–67.

Johnston, Jamie, and Christopher Ksoll. 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., and Rebecca L. Thornton. 2021. "Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures." *Review of Economics and Statistics* 103 (2): 251–64.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340 (6130): 297–300.

Lai, Fang, Linxiu Zhang, Yu Bai, Chengfang Liu, Yaojiang Shi, Fang Chang, and Scott Rozelle. 2016. "More is Not Always Better: Evidence from a Randomised Experiment of Computer-Assisted Learning in Rural Minority Schools in Qinghai." *Journal of Development Effectiveness* 8 (4): 449–72.

Lai, Fang, Renfu Luo, Linxiu Zhang, Xinzhe Huang, and Scott Rozelle. 2015. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing." *Economics of Education Review* 47: 34–48.

Lai, Fang, Linxiu Zhang, Xiao Hu, Qinghe Qu, Yaojiang Shi, Yajie Qiao, Matthew Boswell, and Scott Rozelle. 2013. "Computer Assisted Learning as Extracurricular Tutor? Evidence from a Randomised Experiment in Rural Boarding Schools in Shaanxi." *Journal of Development Effectiveness* 5 (2): 208–31.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–1102.

Linden, Leigh L. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India." http://www.leighlinden.com/Gyan_Shala_CAL_2008-06-03.pdf.

Lucas, Adrienne M., and Isaac M. Mbiti. 2014. "Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya." *American Economic Journal: Applied Economics* 6 (3): 234–63.

Lucas, Adrienne M., Patrick J. McEwan, Moses Ngware, and Moses Oketch. 2014. "Improving Early-Grade Literacy in East Africa: Experimental Evidence from Kenya and Uganda." *Journal of Policy Analysis and Management* 33 (4): 950–76.

Lu, Meichen, Prashant Loyalka, Yaojiang Shi, Fang Chang, Chengfang Liu, and Scott Rozelle. 2019. "The Impact of Teacher Professional Development Programs on Student Achievement in Rural China: Evidence from Shaanxi Province." *Journal of Development Effectiveness* 11 (2): 105–31.

Malamud, Ofer, and Cristian Pop-Eleches. 2011. "Home Computer Use and the Development of Human Capital." *Quarterly Journal of Economics* 126 (2): 987–1027.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *Quarterly Journal of Economics* 134 (3): 1627–73.

Mo, Di, Johan Swinnen, Linxiu Zhang, Hongmei Yi, Qinghe Qu, Matthew Boswell, and Scott Rozelle. 2013. "Can One-to-One Computing Narrow the Digital Divide and the Educational Gap in China? The Case of Beijing Migrant Schools." *World Development* 46: 14–29.

Muralidharan, Karthik. 2013. "Priorities for Primary Education Policy in India's 12th Five-Year Plan." In *India Policy Forum,* Vol. 9, edited by Shekhar Shah, Barry Bosworth, and Arvind Panagariya, 1–61. Delhi, India: National Council of Applied Economic Research.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109 (4): 1426–60.

Naslund-Hadley, Emma, Susan W. Parker, and Juan Manuel Hernandez-Agramonte. 2014. "Fostering Early Math Comprehension: Experimental Evidence from Paraguay." *Global Education Review* 1 (4): 135–54.

National Education Assessment System. 2016. *National Assessment Report 2016*. Islamabad, Pakistan: Ministry of Federal Education and Professional Training.

Navarro-Sola, Laia. 2019. "Secondary School Expansion through Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria." https://laianaso.github.io/laianavarrosola.com/Navarro-Sola_JMP.pdf.

Papay, John P., Eric S. Taylor, John H. Tyler, and Mary E. Laski. 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, Hafiz Muhammad Iqbal, and Shahida Sohail. 2010. "Introducing Science Experiments to Rote-Learning Classes in Pakistani Middle Schools." *Evaluation and Research in Education* 23 (3): 191–212.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4): 1289–1324.

Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Armida Alishjabana, Arya Gaduh, and Rima Prama Artha. 2011. "Improving Educational Quality through Enhancing Community Participation—Results from a Randomized Field Experiment in Indonesia." World Bank Policy Research Working Paper 5795.

World Bank Group. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

This article has been cited by:

1. Liang Choon Wang, Michael Vlassopoulos, Asad Islam, Hashibul Hassan. 2024. Delivering Remote Learning Using a Low-Tech Solution: Evidence from a Randomized Controlled Trial in Bangladesh. *Journal of Political Economy Microeconomics* **2**:3, 562-601. [Crossref]
2. Anjali Chandra, Subha Mani. 2024. Improving children's learning in developing countries: what works and what doesn't. *New Zealand Economic Papers* **135**, 1-5. [Crossref]
3. Emilio Borghesan, Gabrielle Vasey. 2024. The Marginal Returns to Distance Education: Evidence from Mexico's Telesecundarias. *American Economic Journal: Applied Economics* **16**:1, 253-285. [Abstract] [View PDF article] [PDF with links]
4. Joana Cardim, Teresa Molina-Millán, Pedro C. Vicente. 2023. Can technology improve the classroom experience in primary education? An African experiment on a worldwide program. *Journal of Development Economics* **164**, 103145. [Crossref]