Impact of Internet Access on Student Learning in Peruvian Schools

Kevin Kho
Food and Drug Administration
khokevin@msu.edu

Leah K. Lakdawala
Michigan State University
lkl@msu.edu

Eduardo Nakasone
Michigan State University & IFPRI
eduardo@msu.edu

Selected Paper prepared for presentation at the 2019 Agricultural & Applied Economics Association Annual Meeting, Atlanta, GA, July 21 – July 23

Impact of Internet Access on Student Learning in Peruvian Schools *

Kevin Kho[1], Leah K. Lakdawala[1], and Eduardo Nakasone[1,2]

[1]Michigan State University
[2]International Food Policy Research Institute

Abstract

We investigate the impacts of school-based internet access on pupil achievement in Peru, using a large panel of 5,903 public primary schools that gained internet connections during 2007-2014. We employ an event study approach and a trend break analysis that exploit variation in the timing of internet roll-out up to 5 years after installation. We find that internet access has a moderate, positive short-run impact on school-average standardized math scores, but importantly that this effect grows over time. We provide evidence that schools require time to adapt to internet access by hiring teachers with computer training and that this process is not immediate. These dynamics highlight the need for complementary investments to fully exploit new technological inputs and underscore the importance of using an extended evaluation window to allow the effects of school-based internet on learning to materialize.

*Email addresses for correspondence: khokevin@msu.edu, lkl@msu.edu, eduardo@msu.edu. We thank Jorge Mesinas at the Peruvian Ministry of Education for access to the data from the *Evaluacion Censal de Estudiantes*. We are very grateful to Chris Ahlin, Diether Beuermann, Prashant Bharadwaj, Julian Cristia, Pascaline Dupas, Steven Haider, Scott Imberman, Emilia Tjernström, and Maisy Wong for their insightful and helpful comments and suggestions. We are also thankful to seminar participants at the Michigan Development Day, Michigan State University, and University of Wisconsin-Madison.

1 Introduction

In recent decades, developing countries have achieved large increases in school enrollment, particularly at the primary level. However, most remain far behind developed countries in terms of school quality as measured by student achievement (Glewwe and Kremer 2006). Traditional policies — such as hiring additional teachers or providing textbooks — do not appear to have improved student achievement in developing countries (Kremer et al., 2013).[1] In turn, new approaches to improving school performance, such as Information and Communication Technologies (ICTs), have garnered increasing interest. The promise of boosting modern-day digital competencies, promoting interactive student-centered teaching models, and providing up-to-date learning materials even in remote areas (World Bank, 2018) has encouraged developing countries to invest considerably in ICTs in schools (World Bank, 2018; Escueta et al., 2017; One Laptop per Child, 2016; UNESCO, 2012; Trucano, 2016; International Telecommunication Union, 2014).

Among ICTs, the internet in particular may have important pedagogical uses in developing countries. Internet access can provide underserved students with otherwise unavailable sources of information (Levin and Arafeh, 2002). Similarly, internet can expand teachers' access to references and teaching aids as well as their ability to share information among peers (Jackson and Makarin, 2016; Purcell et al., 2013). However, as with many new technologies, benefits materialize only after a period of learning and adaptation, suggesting the importance of understanding the dynamic effects of ICT interventions over time.

Despite the potential of internet to improve learning, few studies have rigorously evaluated its impacts on student performance in developing countries. Though previous research in developed countries has led to ambivalent conclusions on the effectiveness of internet access as a learning input (Belo et al., 2014; Faber et al., 2015; Gibson and Oberg, 2004; Goolsbee and Guryan, 2006; Machin et al., 2007; OECD, 2015; Vigdor et al., 2014), school-based connectivity can be potentially more important in developing countries due to lower levels of teacher skills, larger class sizes, and limited access to other conventional inputs.[2] Additionally, since the broader literature on ICTs (Escueta et al. 2017 and Bulman and Fairlie 2016) typically examines bundles of interventions such as computer access, learning software, and internet expansion[3], it is not yet clearly understood how

[1] Some notable exceptions exist; for example, see Das et al. (2013) and Muralidharan and Sundararaman (2013).

[2] In a recent paper, Malamud et al. (2018) investigate the impact of home-based internet on Peruvian students' school performance, finding no statistically significant effect on standardized test scores 9 months after the implementation of the program. The authors posit that too little time might be spent on computers at home for any educational benefit to materialize. Relatedly, children might use internet as a tool for entertainment rather than learning. Both of these problems might be reduced when internet is provided at school rather than home.

[3] Some notable exceptions analyze the individual impact of computer access (Beuermann et al., 2015; Cristia et al., 2017; Barrera-Osorio and Linden, 2009; Mo et al., 2013; Toyama, 2015; de Melo et al., 2013; Sharma, 2014; Meza-Cordero, 2017; Bai et al., 2016) or learning software (Bando et al., 2016; Banerjee et al., 2007; Carrillo et al., 2010; He et al., 2008; Linden, 2008; Muralidharan et al., 2016) in developing countries. However, there is little evidence on the impact of internet access.

internet on its own influences learning.[4]

Moreover, most prior studies of internet access — and of ICTs more generally — have been based on short-term observation of small samples, and are thus only designed to detect somewhat larger and immediate treatment effects. Importantly, such studies may overlook potential longer term impacts that may follow from an initial learning period, during which teachers, students, and administrators adapt to new technology. Hence, detecting gains in learning that may arise over such a learning period requires a longer evaluation window.

We examine the impact of internet access on student performance in the universe of public primary schools in Peru that initially acquired internet between 2007 and 2014, emphasizing its dynamic effects in schools over time. During our sample period, about 933,000 students gained access to internet. We link administrative data on school-based access to internet with school-average math scores from a large-scale national test that covers nearly the universe of second graders in public schools in Peru. We construct a panel dataset of 5,903 schools that gained internet during our study period. To fully exploit the longitudinal structure of our data and identify dynamic effects, we employ an event study framework in addition to a trend break analysis — approaches which also allow us to detect and control for pre-existing trends in student performance. Since we observe a large panel of schools over eight years, we are also able to assess how other determinants of school performance change over time, tracing out the dynamics of student, teacher, and school-level inputs. This allows us to discuss potential channels through which internet affects school performance, as well as to thoroughly explore the possibility that other confounding factors drive our results.

Using within-school variation in the timing of internet installation, we find that internet access leads to initial modest math score improvements of 0.042 to 0.076 standard deviations in the first 18 months after installation. Importantly, this advantage grows significantly over time (at a rate of about 0.047 standard deviations per year on top of an initial level improvement), reaching 0.29 standard deviations five periods after installation.

We posit that this growth in our estimated impacts over time reflects an adaption period, during which schools must learn to integrate new technologies. Namely, we observe that schools respond to internet access by hiring teachers with formal training in digital skills, and that this process follows only gradually. In particular, schools are 2.1 percentage points more likely to have a computer-trained teacher in the first year after installing internet, and 9.6 percentage points more likely by the fifth year after installation — a doubling of the pre-internet likelihood of having computer-trained teachers. Hence, the fact that the gradual growth over time in test scores shadows growth in the staffing of computer-trained teachers may suggest that complementary investment in staff

[4] Previous work (e.g., Cristia et al., 2014; Bet et al., 2014; Sprietsma, 2007) has assessed programs providing school-based internet as part of broader ICT expansion schemes. However, these papers do not aim to discern the effect of internet separately from that of other technologies.

computer proficiency is needed to fully exploit internet-enabled classroom capabilities.[5]

Furthermore, our data offers suggestive evidence for two potential channels through which internet access improves test scores. First, gains in test scores are predominantly driven by schools with high student-to-teacher ratios, suggesting that internet-related activities may supplement the limited individualized attention that teachers can provide in large classes. Second, gains in test scores are largest for schools with relatively low teacher qualifications (as measured by the per student count of teachers holding a pedagogical or university degree) which is consistent with internet resources compensating for or addressing deficits in teacher training.[6]

Additionally, our main findings are robust to a number of alternative explanations. Concerning potential endogeneity in the timing of internet access, we find that, conditional on year and school fixed effects and a set of time-varying school characteristics (e.g., school size, infrastructure, and resources), schools receiving access to internet do not exhibit positive pre trends in performance or different pre internet scores compared with those that do not. Second, we also find that our results are not explained by concurrent changes in other inputs (e.g., infrastructure, textbooks, or computers) or by pre-existing trends that differ by geographic areas, administrative units, internet installation year, or even by each individual school (i.e. school-specific trends). Third, while our main specifications are based on an unbalanced sample of schools, our results are very similar when using different sample restrictions (including a sample of non-attriting schools). Lastly, analyzing student composition within schools shows that our findings cannot be driven by endogenous sorting of students.

We contribute novel insights and perspective to a nascent body of research in developing countries on the educational benefits of school-based internet access, as well as to a wider literature concerning ICTs as schooling inputs. Primarily, the size and time span of our data present opportunities to complement and contextualize existing knowledge from randomized control trials (RCTs), which largely comprise the current work relating ICTs and academic performance. Whereas RCTs are mostly constrained to observe short term effects (rarely beyond one academic year), we use our extended study period to analyze the effects of internet access up to 5 years after it is introduced to schools. Our results indicate that this longer evaluation window is highly relevant to understanding the impact of internet access, due to the dynamic effects of internet on learning over time.

Additionally, the large scale of our sample — containing about 6,000 public schools — provides power to detect the short-run impacts of internet, which appear to be modest in size. Namely, we discover average math score gains of 0.042-0.076 standard deviations in the first year of internet access, statistically significant at the 5% level. While several evaluations of programs distributing

[5]Similarly, evaluations of laptop provision in the U.S. (Hull and Duch 2016) and computer assisted learning in China (Mo, Zhang, Wang, Huang, Shi, Boswell and Rozelle 2015) estimate that the effects of ICT interventions grow over time.

[6]These findings complement evidence from the developed country context, where providing teachers with online access to "off the shelf" lesson plans improves students' math achievement, and where benefits were larger for weaker teachers (Jackson and Makarin 2016).

computers in developing countries report similarly sized short-run effects (ranging from 0.052-0.088 standard deviations 5-22 months after initial access; Bet et al. 2014, Barrera-Osorio and Linden 2009, Cristia et al. 2017, Beuermann et al. 2015, Mo et al. 2013), none of these studies are able to statistically distinguish effects from zero, perhaps in part because they analyze far smaller samples of schools (ranging from 13 to 318 schools). Since we investigate a massive national policy (which affected public primary schools serving roughly 900,000 children), we can assess conditions relevant for internet provision programs. We find that the effects of internet access are largest in schools with low levels of existing resources, particularly those that are understaffed and whose teachers are less qualified.

Finally, we explicitly identify the gains that internet access produces over hardware resources alone. Anecdotally, the usefulness of school computers without internet access has been limited by lack of access to information (National Public Radio 2012) and the inability to obtain routine maintenance and software updates — particularly in remote, difficult-to-reach locations (One Laptop per Child 2011). Indeed, our data suggest that computers alone (in schools without internet access) have only modest impacts on student learning. To the best of our knowledge, the scale, longitudinal length, and setting of this study, along with the comprehensiveness of the available data uniquely address important gaps in the existing literature. More broadly, our work contributes to understanding the role of internet access in economic development. In consideration of prior research connecting faster internet to higher employment, incomes, and wealth in African countries (Hjort and Poulsen 2017), increased human capital production may factor importantly in this progress.

The paper proceeds as follows. In Section 2, we describe the educational setting in Peru and *Plan Huascarán*, the source of variation in internet installation for many pubic schools during our sample period. We also provide details about the two administrative datasets we merge for our analysis, the *Censo Escolar* and the *Evaluación Censal de Estudiantes*. Section 3 describes our event study and trend break estimation strategies and presents the main estimates of the impact of internet access on test scores. In Section 4, we investigate the robustness of our results to a number of plausible confounding factors: changes in other school resources, potential regional shocks correlated with internet access and student performance, differential pre-trends in test scores, and changes in sample composition (both in terms of schools and students). In Section 5 we show that the dynamic patterns in test score impacts may be explained by an adaptation period, during which schools hire computer-trained teachers. We also shed light on two potential mechanisms through which internet access generates gains in learning by examining heterogeneous effects among different subsamples. Section 6 summarizes and discusses our findings.

2 Setting and data

2.1 Education and ICT Access in Peru

Education in Peru is compulsory and free through the public school system beginning at age 3 and continuing until the end of secondary school. In the past few decades, Peru has greatly increased access to primary school (grades 1-6, approximately age 6-11), raising the net enrollment rate from 85.6% in 1980 to 97.9% in 2015 (The World Bank 2016). At the same time, however, the education budget has seen little growth, and thus greater enrollment over time has eroded per-student resources (Saavedra and Suarez 2002). The World Bank (2012) finds that, within Latin America, only the Dominican Republic has a lower education expenditure-to-GDP ratio than Peru.

This dearth of resources has limited the quality of education, as evidenced by Peru's performance in the OECD's Program for International Student Assessment (PISA) — an international standardized test among 15 year olds — in 2012 and 2015. In 2012, Peru ranked last out of 65 participating countries in all three evaluated subjects, with results revealing that most Peruvian students have serious deficiencies in math (75% deficient), science (69%), and reading (60%). In 2015, Peru jumped to the 64th place (out of 70 countries in the evaluation), nonetheless demonstrating that substantial progress remains to be made. Widespread under-preparedness is evident as early as primary school. In 2007, the Ministry of Education began administering yearly standardized tests, the National Student Assessment or *Evaluacion Censal de Estudiantes* (henceforth ECE, described below), to all second graders registered in classes with five or more students. The inaugural results of the ECE in 2007 showed that only 7% of students acquired skills mandated by the national curriculum in math (Appendix Figure A.1). Despite improvement since then in test scores and in the proportion of students meeting expected skill levels, the quality of schooling has continued to prove inadequate for many children; even by 2014, fewer than a quarter of second graders achieved proficiency in math.

In the early 2000s, the Peruvian government launched *Plan Huascarán*, which produced much of the variation in school internet access observed during our sample period. This project aimed to "incorporate information and communication technologies to increase the coverage, quality, decentralization, democratization, and equity of the Peruvian education system." Project planners ambitiously aimed to install hardware and internet in 32,000 schools and to train 180,000 teachers by 2020. *Plan Huascarán* targeted primary, secondary, and integrated schools (i.e., those that teach both primary and secondary classes) under public management, particularly in rural (or peri urban) and high poverty areas. Officially, selection into the program was rationed, with each Local Educational Management Unit (UGEL) allowed to submit a set number of its schools, adhering to a set proportion of primary, secondary, and integrated schools (see Appendix Figure A.2 for an excerpt of the official Ministry of Education flow chart that outlined the specific prioritization protocol under *Plan Huascarán*).[7] The largest allocated proportion (50%) was set aside for primary schools. As prerequisites for program selection, schools needed to have electricity and a computer lab with anti-theft measures (i.e., perimeter fencing). Within each UGEL and level (e.g., primary), prioritization among qualified schools was officially based on the size of the student population, with larger schools receiving higher priority. Lists of eligible schools were aggregated to the regional

[7] An excerpt from the translated Ministry of Education directive regarding the prioritization protocol for *Plan Huascarán* is provided in Appendix Figure A.3.

level and then submitted to *Plan Huascarán* headquarters, accompanied by data sheets on the characteristics of each school listed, a sketch and description of each school's computer facilities, and the discussion minutes from each UGEL.[8] Officially, no school was integrated into the project without all required information.

As a consequence of initiatives such as *Plan Huascarán* and the One Laptop per Child program (OLPC, undertaken by the Peruvian government in 2008)[9], the ratio of students to computers in primary schools fell dramatically from 240 to 6 between 2000 and 2014 . In parallel, the government has steadily increased access to internet in schools (as described in Section 2.2.1). In 2013, the Ministry of Education announced plans to triple the number of schools having internet access.

2.2 Data

Our analysis uses school-level data from two sources administered by the Ministry of Education: the *Censo Escolar* (CE), an annual census of schools, and the *Evaluación Censal de Estudiantes* (ECE), an annual standardized test of second graders' skills.

2.2.1 Censo Escolar (CE) and School-based Internet Access

Each year, all school principals are required to submit two forms with their updated information to the Ministry of Education. Between April and July, principals complete a form on enrollment (by grade and age), teachers (by qualification), available supplies and materials (e.g., books, computers, and laboratories), and infrastructure (e.g., access to utilities, building characteristics, and internet connectivity). Between December and February, another form is completed on year-end pupil outcomes (e.g., number of pupils transferring to other schools).[10] We refer to the CE for data on school characteristics such as internet access, enrollment, teachers, educational materials and resources, and physical infrastructure.[11] Between 2007 and 2014, around 29,500 public primary schools reported administrative information in the CE annually.

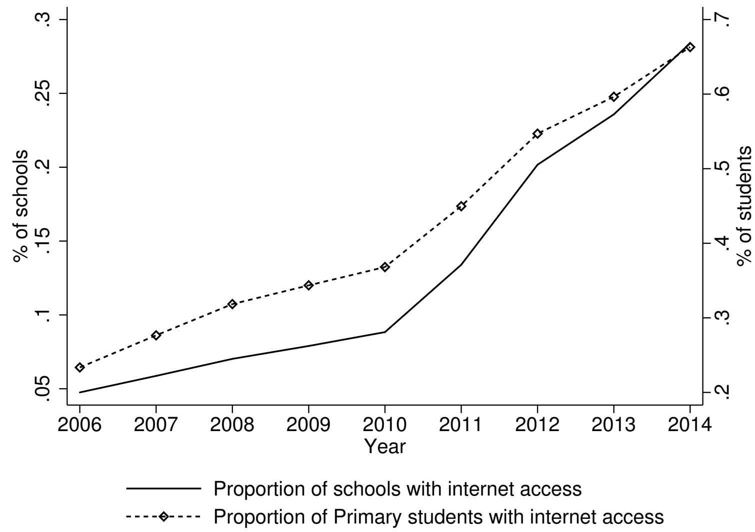
[8] A translated version of the school data sheet is provided in Appendix Figure A.4.

[9] Peru has been the single largest buyer of OLPC laptops and to date has distributed close to one million laptops, mainly targeting school children in poor areas of the country. As our analysis to follow accounts for the total number of computers in a school, including OLPC laptops, we indirectly control for the influence of OLPC. For a discussion of the OLPC program in Peru, see Trucano (2012). In general, impact evaluations of OLPC in Peru suggest that the provision of laptops did not improve student performance (Beuermann et al. 2015; Cristia et al. 2017).

[10] The school year in Peru runs from March to December.

[11] While this information is self-reported by school principals, the Ministry of Education applies different filters and verifies the consistency of the data with secondary data sources. The CE forms are submitted to the Ministry of Education electronically and include consistency rules to avoid reporting errors. Once the electronic forms are submitted, the Ministry of Education validates the information with teacher payrolls, delivery records of materials, and historic information on enrollment. To provide further evidence that schools do not strategically misreport information (for example, inflating enrollment), we compare second grade enrollment as reported in the CE to the number of students scheduled to take the exam in the ECE data. The median discrepancy in these numbers is 0 and the average is 0.33 students (i.e. on average 0.33 more second grade students are reported in the ECE than in the CE); since the CE is reported earlier in the school year than the ECE, this discrepancy could arise naturally due to students entering after the CE forms have been submitted.

Figure 1: Internet Connectivity in Primary Public Schools, 2006-2014



NOTE: The stock of schools that gained internet connections is based on the first year in which they report internet access in the Peruvian *Censo Escolar*.

We use information from the CE to determine the timing of initial internet connection among schools in our sample. Administrators report in the first semester of every year whether their school currently has access to internet. Though some schools report gaps in internet access, the data do not allow us to distinguish between temporary outages and longer-term disruptions to connectivity.[12] Based on this information, we determine the first year in which a school reports gaining access to internet and interpret this as the time of connection. In our estimation framework, this implies a conservative estimate of the impact of internet access, because we treat schools that might have permanently lost their connections as still being connected. Another benefit of using initial internet connection rather than current access is that we avoid bias due to endogenous changes in access. We estimate that 7,089 schools — and the 933,000 students in these schools — at some point gained internet connectivity between 2007 and 2014. This implies that the rate of internet connection in schools increased from 5% to 30% and that the share of students with internet connection in their schools jumped from 23% to 66% (Figure 1).

Most of the observed expansion in internet connectivity during this period was due to *Plan Huascarán*. In Appendix Table A.1 we verify that the official qualification and prioritization rules set by the Ministry of Education do in fact predict actual installation. Schools received priority primarily based on quotas by province (Local Education Management Units, UGELs), high poverty status, location in a rural versus urban area, public versus non-public management, the presence

[12] Out of the 30,338 public primary schools with at least one year of data in the CE between 2006 and 2014, 13.2% report not having access after having access in a previous year; about 20% of those schools regain internet access at a later point.

of required infrastructure (including electricity, a computer lab, and anti-theft measures such as perimeter fencing), and enrollment. Column 1 includes these characteristics and year effects to control for aggregate trends in internet connectivity. To approximate the status of "adequate infrastructure, in good condition," we include indicators of whether the school has a library and administrative offices. To capture high poverty status, we include district-level fixed effects. We also include UGEL fixed effects to account for the UGEL-specific quotas. As expected, most prioritization characteristics positively predict internet access, though location in urban areas is not statistically significant. In column 2, we add school fixed effects to match our main specification (described in Section 3), thus dropping terms for prioritization characteristics that are time-invariant within schools (e.g., UGEL, district, and location). Even with school fixed effects, facilities such as the existence of a computer room, administrative offices, and a library positively predict internet access. This pattern is consistent across specifications 3 and 4, which additionally control for perimeter fencing (available only for 2010 and later) and information from school data sheets (number of computers used for instruction, number of computers used for administrative purposes, and number of teachers), respectively. Since these factors predict internet access and are also likely to influence student performance directly, we control for all of these measures in our main specifications (except perimeter fencing due to data limitations).

2.2.2 Evaluación Censal de Estudiantes (ECE)

The Ministry of Education also mandates the *Evaluación Censal de Estudiantes* (ECE), a yearly standardized assessment of second graders' skills, which is administered in late November or early December (before the end of the school year). In order to ensure uniform testing environments — and to prevent content leaks or influence from school personnel — the Ministry hires independent staff to administer the test in all schools simultaneously. As the same test is given to all schools, neither the content nor the testing environment varies by school characteristics. Furthermore, the ECE was designed for comparability of results over time: experts defined the current and future skill categories prior to the test's first administration. Hence, since its inauguration in 2007, the ECE has assessed the same skill sets with consistent relative focus. In our data, we do not observe student-level ECE scores (or *any* student-level characteristics); instead we observe only school-level averages. To account for differences in difficulty across cohorts and year-to-year changes in test score dispersion, we standardize ECE scores across the universe of tested schools within each year.

The ECE gauges the academic performance of the vast majority of second graders in Peru, targeting all public and private schools that meet two criteria: 1) having at least five second graders enrolled during the test year, and 2) using Spanish as the primary language of instruction. The rationale for the first criterion is entirely budgetary, as smaller schools are often in remote areas and would take considerable resources to reach. As it stands, the ECE already requires about 40,000 field workers each year. Schools teaching in indigenous languages are covered under a separate testing schedule. In total, 16,000 – 19,000 primary public schools participated per year (55% to 65% of all primary schools; see Figure A.5a). About 27% – 39% of schools were exempt under the minimum enrollment or language criteria. The remaining schools (between 4% and 10%) were not tested due to logistical problems. The coverage of the test was nonetheless very broad: since the smallest schools were excluded by definition, and since schools in native language tend to have modest enrollments, between 83% and 90% of all second graders in the country were tested in the ECE in a given year (Appendix Figure A.5b).

Appendix Table A.2 displays the summary statistics for the full sample of schools (columns 1-3) and for the subsample of schools that also appear in the ECE (columns 4-6). In general, the two samples appear to be similar on observable characteristics. As a proportion, there are more schools that do not report ECE scores among the unconnected schools (columns 3 and 6). This is because the unconnected schools are smaller and are therefore less likely to be covered under the ECE (required only for schools with at least 5 second grade students).

2.2.3 Estimation Sample

Our empirical strategies exploit the timing of internet connection within schools. Therefore, we restrict our sample to only those schools that help identify the effects of internet access conditional on school fixed effects (i.e., those which installed internet during the study period, 2007-2014) and exclude all schools without changes in internet access during this period (i.e., those that already had internet before 2007 and those that did not gain access by 2014).[13] This leaves us with 7,083 schools, roughly a quarter of all public primary schools in Peru (Appendix Table A.2, column 2). We then merge this information with annual school-level average math scores from the ECE, resulting in 6,527 matched schools (Appendix Table A.2, column 5).

Columns 4-6 of Appendix Table A.2 present summary statistics from 2007 (or each school's earliest available year in our sample) for the 25,624 schools that appear in both the CE and the ECE in our sample period. We divide the sample into schools that already had access to internet before 2007 ("early adopters"), those that became connected between 2007 and 2014 (our estimation sample), and those that had not gained access by 2014 ("non-adopters").

We highlight two key observations from Appendix Table A.2. First, only 1,359 (5.3%) of schools were internet-equipped by 2007. Though 17,738 schools (69.2%) remained unconnected by 2014, 6,527 schools (25.5%) gained access during our study period. The sharp expansion in internet connectivity during this period allows us to form insights from a large number of schools despite our sample restrictions.[14] Second, schools that gained internet from 2007-2014 generally fall "between" the early adopters and non-adopters in various measures of school quality. Namely, early adopters appear to be schools with higher performance, larger enrollment, and better infrastructure and edu-

[13] In Section 3.1, we explain why our estimation sample does not include schools that gain internet access prior to 2007 or do not have access as of 2014.

[14] Within the group of schools that gained connection between 2007-2014, there is considerable variation in the timing of access for our analysis: 4,915 schools are observed for at least one period prior to internet connection, 5,424 are observed 1-2 years after internet connection, and 3,316 are observed 3-5 years after internet connection. This allows us to implement the event study approach described in Section 3.1.

cational inputs (e.g., piped water, libraries, administrative offices, teachers, classrooms, computers, and textbooks). Conversely, non-adopters systematically appear worse in these areas. Thus, our estimation sample focuses neither on the best nor on the worst performing schools.

As we use a school-fixed effects approach, we further restrict our estimation sample to schools that are observed at least twice during the sample period. All in all, there are 5,903 schools that were tested in the ECE, that gained internet in our period of analysis and that are observed at least twice during our window of analysis. Appendix Table A.2 illustrates how the sample of schools that gain internet access between 2007 and 2014 changes as we make successive restrictions. In column 2, we start with the 7,083 schools that install internet during our sample period according to the CE. The fifth column is comprised of all schools that can then be matched to ECE data (6,527). The final (seventh) column contains our estimation sample, i.e., the subset of schools that give the ECE at least twice during the sample period (5,903). Schools in the estimation sample are slightly more likely to be urban, more likely to have teachers for each specific grade, larger in terms of enrollment, and equipped with more resources (classrooms, computers, and textbooks). However, they appear to be similar to the full sample of internet-connected CE schools on other important observable dimensions, such as the presence of a library and the number of teachers.

In Figures 2b-2i below, we plot each "treatment cohort's" average math performance over time (by year of initial internet connection).[15] For reference, Figures 2a and 2j represent the performance of schools that gained access prior to 2007 (the start of our sample period) and that had not gained access to internet by the end of our analysis period, respectively. Generally, schools that connected later or remained unconnected exhibit lower average test scores, indicating that variation in internet access *across* schools is not random.

However, *within* a cohort of schools becoming connected in a given year (2007-2014), there do not appear to be trends in scores prior to internet access. This suggests that within cohorts of treated schools, the timing of access is unrelated to test score trends on average. Furthermore, Figure 2 suggests that performance gains among treated schools are modest initially and only become sizable in the medium term. In contrast, the relative math performance of schools that never connected to the internet appears to have stagnated over the period of analysis. Furthermore, it appears that schools with internet connectivity prior to 2007 continued to experience increases in average test scores during our period of analysis.[16] This pattern motivates the strategies employed in our main analysis (Section 3) to identify the dynamic effects of internet.

[15]Recall that within each year, school level averages are normalized across *all* schools giving the ECE, including those not in our main estimation sample.

[16]We cannot determine the timing of internet connectivity prior to 2007. While we can identify schools that had internet installed by 2007, we do not have information about the specific year in which they gained connectivity.

Figure 2: Standardized Average Math Scores over Time, by Year of Initial Internet Access

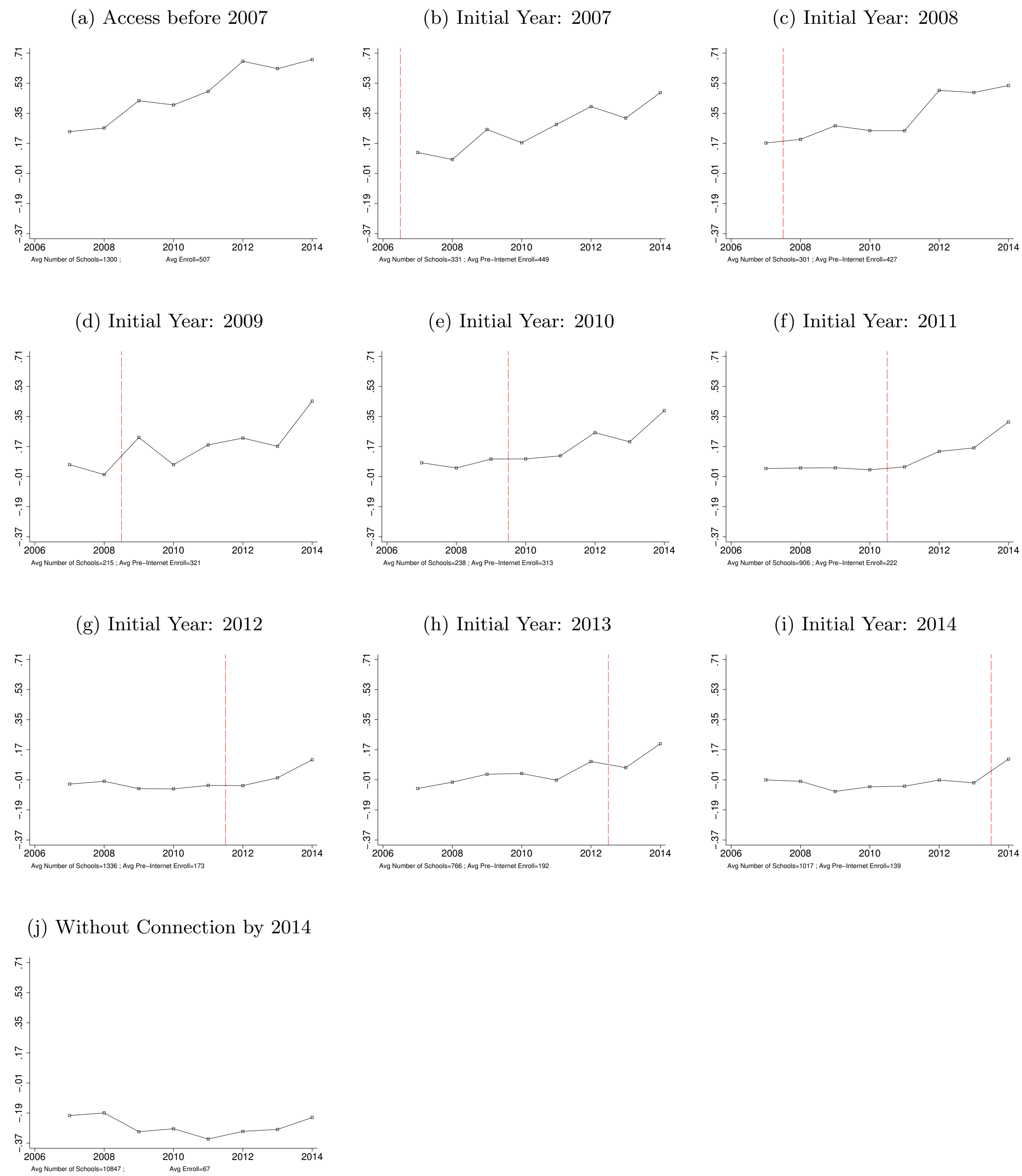


Figure 2a plots the standardized test scores for all public schools that participated in the ECE and had an internet connection prior to 2007. It also shows the average enrollment in these schools (2006-2014). Figures 2b-2i plot the standardized test scores over time separately for groups of schools that participated in the ECE based on the year of initial internet connection. The sample includes all public schools that initially gained internet access between 2007 and 2014 and with at least two observations within the period.For each group of schools, we also show the average enrollment before internet connection (e.g., for schools that gained internet connection in 2010, this is the average enrollment between 2006 and 2009). Figure 2j plots the standardized test scores for all public schools that did not have an internet connection by 2014 and shows their average enrollment between 2006 and 2014 .

3 Empirical strategies and results

3.1 Event study specification

In order to analyze dynamic impacts of internet access over time, we estimate the following event study specification:

$$Y_{ir} = \sum_{t=-3}^{5} \beta_t \mathbf{1}\{E_{ir} = t\} + \gamma X_{ir} + \alpha_i + \theta_r + \varepsilon_{ir}, \quad t \neq -1 \tag{1}$$

Our primary outcome of interest is the school-level average of standardized math scores for second grade students[17] in school i in year r (Y_{ir}) (normalized across the universe of Peruvian schools within each year). α_i and θ_r are school and year fixed effects.[18] α_i captures all time-invariant observed and unobserved school-level determinants of performance. θ_r accounts for year-to-year aggregate changes in performance: for example, changes to the national curriculum, national policies regarding teacher contracts, general increases in educational investments, and visibility of international assessments (such as PISA). X_{ir} is a set of time-varying school characteristics that includes total enrollment, number of second grade students scheduled for testing, facilities (piped water, library, administrative offices), and resources per student (classrooms, computers and teachers).

Let I_i denote the year in which school i gains internet connection (the first year in the dataset in which i reports internet access in the CE). E_{ir} represents time relative to internet access for each school; specifically, $E_{ir} = r - I_i$. The coefficients on the set of event study dummy variables β_t capture the path of test scores relative to the year before a school receives internet access (i.e., relative to $t = -1$). It is worth highlighting one important feature in the timing of the two datasets we use. The CE reports internet access in the beginning of the school year, while the ECE is a year-end test. Any school that installs internet after the CE (April-July) does not report internet access until the following calendar year. If internet installation occurs before the ECE exams (end

[17] Though the ECE also tests reading ability, we do not examine effects on reading scores in this paper. In a review of literature examining Computer Assisted Learning (CAL), Parr and Fung (2000) observe that "the best results appear to be for basic maths skills; there is little evidence of gains in reading." While only a few studies at present compare the impact of ICTs on learning across different subjects, most of this evidence suggests that math scores are more responsive to the provision of technology (see the results in Banerjee et al. 2007, Carrillo et al. 2010, Mo et al. 2013, Mo, Huang, Shi, Zhang, Boswell and Rozelle 2015, Muralidharan et al. (2016), and Texas Center for Educational Research 2009). Some practitioners also argue a priori that ICTs should affect math performance more so than reading. The chief schools officer at Rocketship Education, a large charter school that heavily utilizes teaching software, argues that "math software is much further along than literacy... To isolate the basic skill of literacy is just much trickier to do... For a computer to know whether or not there's a proper self-to-text connection is a lot trickier than finding out if they have the right answer to math problem" (Barseghian 2011).

[18] A regression that includes school fixed effects, event study time dummies, and a full set of calendar year fixed effects results in perfect multicollinearity. We therefore pool two year effects (which should be close to zero, given that school-level scores are normalized within each year to a mean of zero and standard deviation of one). Results are robust to our pooling choice, and are similar when pooling two pre-internet event study indicators (e.g., t=-3 and t=-2) or dropping year effects altogether.

of November - December), students are exposed to internet access during at least part of the year *prior* to reporting initial access in the CE. Therefore in merging internet information from the CE to test scores from the ECE, we match test scores from the ECE to the internet status in the CE of the following calendar year. This means that some schools acquire internet access in $t = 0$ (if installation occurred *before* submitting the CE information) while others acquire it in $t = 1$ (if installation occurred *after* submitting the CE information). Unfortunately, school-level information is not available for either the month of installation or completion of the CE, and so we are unable to tell how many schools receive internet in $t = 0$ versus $t = 1$. Thus, in interpreting estimates of β_t it is important to keep in mind that $t = 0$ is a partially treated year for some schools and a pre-treatment year for others, while $t = 1$ is a partially treated year for some schools and a (fully) treated year for others.

By exploiting variation in the timing of internet access *within* schools (as well as additionally controlling for aggregate year effects and a set of time-varying characteristics), we aim to identify the effects of internet access separately from potential confounders that are fixed at the school level. We consider this a refinement over Hopkins (2014) — who also examines the relationship between internet access and test performance in Peru — but compares internet-connected schools to non-connected schools (including those that never become connected). We use the event study framework to examine both pre-treatment trends and dynamic effects in a non-parametric fashion for up to five periods following internet access. Standard errors are clustered at the school level to allow for arbitrary serial correlation in ε_{it}.[19]

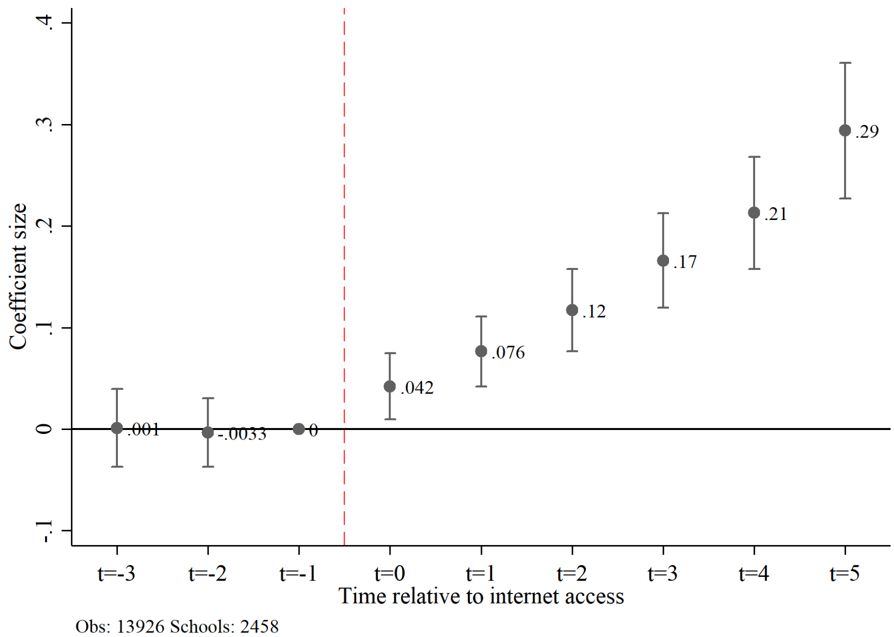
Our estimation sample excludes both schools that received internet access prior to 2007 and schools that remained unconnected as of 2014 (i.e., schools without variation in their internet access). While these observations would not aid in the identification of our coefficients of interest β_t, they could in theory identify the calendar years θ_t. However, this leads to other challenges in estimating 1. Specifically, for schools gaining access prior to 2007 or after 2014, we do not observe the first year of internet access, I_i. Thus we cannot assign the correct values for the event study indicators for these schools. To see why this is the case, consider the following simple example. A school that gained internet access in 2005 will be in $t = 2$ as of 2007 (the start of our sample period). If the causal effect of internet access is positive and grows over time (as we find), then scores will increase for this school throughout the sample period. In this situation, the causal effects of internet access will load onto the calendar year dummies, which capture changes in test scores over time. For this reason, we do not include early or late adopters in our estimation sample.

Figure 3 displays the results of estimating Equation 1 on our main sample. The full set of coefficient estimates for the event study dummy variables are reported in Appendix Table A.3. We find that, prior to internet access ($t < 0$), schools' relative math performance compared with their peers was roughly constant from year to year (Figure 3). Importantly, there is no apparent trend

[19] Our results are also robust to two-way clustering on both school and year.

in math scores prior to internet access, indicating that the timing of internet access within schools is unrelated to pre-trends in student performance. In particular, we rule out the case in which internet installation is budgeted endogenously as a reward for steadily improving test performance. While relative math performance rises in all years following initial connectivity, immediate gains are small in magnitude (0.042 standard deviations in the first partial year of access). The improvement does not surpass 0.1 standard deviations until 2 years after installation. By year 5, scores are 0.29 standard deviations higher compared with other schools than before internet installation.

Figure 3: Impact of Internet Access on Test Scores



The above figures plot the coefficients and 95% confidence intervals from estimating equation 1. Full regression results are reported in Table A.3. Scores are standardized within each calendar year to have mean zero and standard deviation of one across the universe of schools reporting scores. Coefficients capture the increase in test scores relative to the year before internet installation ($t = -1$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ and some receive it in $t = 1$. For more details, see Section 3.1. Control variables include total school enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). The sample includes all public schools with at least two observations within the sample period (2007-2014). Standard errors are clustered by school.

On the surface, this finding stands in contrast to other studies in developing countries that find limited or no impacts of ICTs on test scores. However, our short run estimates are in fact similar in magnitude to those from several studies of hardware-only ICTs[20] —which range from 0.052 to

[20]To avoid confounding the impact of internet with that of access to hardware alone, all our regressions control for the number of computers per student. Hence, the estimated effects of internet access account for differential access to computers. We discuss the relationship between the effects of internet access and computer availability more explicitly in Section 4.

0.088 standard deviations, 5 to 22 months post-intervention (Bet et al. 2014, Barrera-Osorio and Linden 2009, Cristia et al. 2017, Beuermann et al. 2015 Mo et al. 2013) — though these other studies are unable to statistically distinguish estimates from zero (based on smaller samples of schools that range from 13 to 318).[21] Results from Figure 3 suggest that though classroom internet is beneficial to learning, improvement in the initial years post-intervention is small. The majority of the studies in this literature focus on impacts within the first 18 months post intervention, an early stage in which impacts may not be statistically detectable in smaller samples. The fact that our estimates grow over time, at least through the medium-term, is also consistent with the only other two longer-term studies of ICTs in education — which have also supported the need for an adaptation period to fully utilize new technologies (Hull and Duch 2016; Mo, Zhang, Wang, Huang, Shi, Boswell and Rozelle 2015).

In the medium-run (3-5 years), the increase in math scores is sizable, though somewhat smaller than those typically found in evaluations of computer assisted learning and related interventions (0.18 to 0.59 standard deviations) (e.g., see Bando et al. 2016, Banerjee et al. 2007, Carrillo et al. 2010, He et al. 2008, Linden 2008, Muralidharan et al. 2016). Our smaller albeit statistically significant estimates may owe partly to the fact that introduction of internet into Peruvian schools was unaccompanied by any particular pedagogical software service, pre-specified uses, or complementary interventions.

3.2 Trend break specification

Though the shape of Figure 3 suggests a steadily increasing effect of internet access on math test scores over time, it does not explicitly test for a break in the trajectory of scores at the time of internet installation. To do so, we estimate a linear trend break specification as follows:

$$\begin{aligned} Y_{ir} &= \phi_1 \text{Post-internet Access}_{ir} + \phi_2 \text{Event Time}_{ir} \\ &\quad + \phi_3 \text{Post-internet Access}_{ir} \times \text{Event Time}_{ir} + \gamma X_{ir} + \alpha_i + \theta_r + \varepsilon_{ir} \end{aligned} \tag{2}$$

Here, Post-internet Access is a dummy variable that is equal to one in all periods after internet installation ($t \geq 0$). Event Time is a linear term for time relative to the year prior to access, $t = -1$. The control set (X_{ir}) is otherwise identical to that described in Section 3.1. In this specification, ϕ_1 captures the level shift in test scores in response to internet access; ϕ_3 represents the change in the linear time trend in math scores after schools gain internet access; and ϕ_2 accounts for any pre-existing linear trend. Based on the results in Section 3.1, it is unlikely that there are any existing pre-trends. However, one benefit of this specification is that even in the presence of any linear pre-trends in test scores, ϕ_3 measures the impact of internet access on the growth in test

[21] Other studies (Angrist and Lavy 2002, Meza-Cordero 2017, Sharma 2014) find negative — though not always statistically significant — effects of hardware introduction on test scores.

scores *apart* from any such trends.

Results from estimating equation 2 are displayed in Table 1. The specification in column 1 controls only for year fixed effects. Though the estimate of the trend break in scores starting in the year of internet installation is positive and significant, we also observe a (statistically insignificant) pre-trend. This suggests that cross-sectional variation across schools may not account for selection into internet access following test score growth. In column 2, we add school fixed effects to account for time-invariant school-level unobservables. Basing identification on only the within-school variation in internet connectivity reveals a positive trend break at the time of installation, with no pre-trend. This specification also shows a larger level improvement in scores upon installation (0.029), but we are unable to statistically distinguish this effect from zero. Finally, we present our preferred specification in column 3, which includes additional controls for time-varying school resources. Based on this specification, estimates of both the level shift and trend break are positive and statistically significant, while the estimate of the pre-trend is close to zero and fairly precise.

Using our preferred specification to linearly approximate the dynamic effects of internet access, we find a level improvement of 0.036 standard deviations upon installation and an additional 0.047 standard deviation gain in each later year. We take care to note a particular limitation of our analysis that stems from evaluating only the short and medium run effects of internet installation: from Figure 3, it is unclear when exactly the positive effects of internet on math scores level off (as opposed to continuing to rise at the rate estimated in Table 1). It may therefore not be appropriate to extrapolate these results over much longer term time spans.

4 Robustness checks

In this setting, identification of the impact of school-based internet access on student performance may be confounded if the timing of internet access is non-random within schools. However, conditional on school fixed effects, year effects, and a set of school-specific time varying controls, our analysis in Section 3 offers no signs that the timing of internet access relates to prior test performance. In this section, we address several other potential challenges to identification, namely endogenous changes in sample composition in terms of both schools and students (including non-random attrition), concurrent changes in school resources, and pre-existing trends that might differ by geographic area, administrative unit, installation year, or individually by school.[22]

[22] As a placebo test, we also perform an exact randomization exercise as follows. We randomly reassign schools' initial year of internet access in our sample. We maintain the actual distribution of installation years, thus ensuring that the sample sizes of each of our "treatment cohorts" matches our data. We then use the randomly assigned installation years to generate false event study dummies, which we use to re-estimate equation 1. We repeat this process 10,000 times and plot the median, 5th percentile, and 95th percentile of the resulting coefficients in Appendix Figure A.6. For the pre-internet periods, the coefficients from the baseline specification are close to the median of the coefficients from the placebo exercise (essentially zero), whereas the post-internet coefficients fall well outside the 5th and 95th percentile of the placebo coefficients. We take this as additional evidence that our coefficients yield causal estimates of the impact of internet access, and that the inference in our baseline specification is appropriate.

Table 1: Internet Access & Test Scores: Trend Break Results

	Dependent Variable: School Average Standardized Math Score		
	Only Year Fixed Effects (1)	Adding School Fixed Effects (2)	Adding Time-varying Controls (Baseline) (3)
Post-internet Access	0.017	0.029	0.036*
	(0.021)	(0.020)	(0.020)
Post-internet Access	0.046***	0.043***	0.047***
× Event Time	(0.011)	(0.011)	(0.011)
Event Time	0.014	-0.008	-0.001
	(0.010)	(0.010)	(0.010)
Observations	31,368	31,368	31,368
Number of Schools	5,903	5,903	5,903
Year Fixed Effects	Yes	Yes	Yes
School Fixed Effects	No	Yes	Yes
Time-varying Controls	No	No	Yes

The sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Standard errors are clustered by school. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Post-internet access is a dummy variable for whether a school has gained internet access (i.e. $t >= 0$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Event time is years relative to internet access. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers).

Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.

4.1 Changes in sample composition

As we use an unbalanced panel (schools are included when they participate in the ECE) observed over a limited window of time (2007-2014), it is possible that our estimated treatment effects to some extent reflect changes in sample composition. Namely, identification of pre-trends and treatment effects might rely on largely different samples of schools. In Section 4.1.1, we find that our estimates are not contaminated by this issue or by school-level attrition. Additionally, we consider also that the composition of students *within* schools may change in response to internet access. For instance, internet access may attract a different pool of students to a school (either from other schools or from non-school activities). Section 4.1.2 presents evidence that this manner of endogenous student sorting does not occur in our sample.

4.1.1 Unbalanced panel and attrition

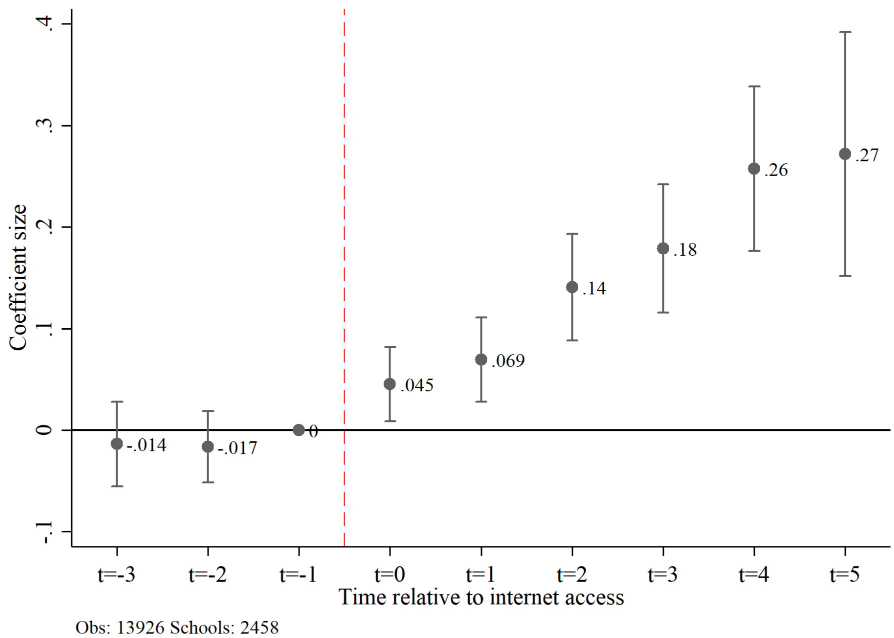
Since some schools in our main sample are observed only prior to internet installation while others are observed only after (our fixed effects estimation only excludes schools without at least two years of data), it is possible that the pre-internet coefficients and trends are identified from a different set of schools than those identifying the post-internet treatment effects. Unfortunately, the likelihood of observing a school's pre-internet years versus post-internet years is furthermore naturally influenced by the date of internet installation (in the extreme case, we of course do not observe the pre-period of any school that installed internet in 2007). If these schools in actuality experienced non-zero pre-trends, our estimates will not take these into account. On the other hand, schools that installed internet in 2014 are not observed post-internet and can only be used to identify pre-internet coefficients/trends. Even though these schools may show no pre-trends, it is possible that they also go on to experience zero (unobserved) effects of internet access.

Figure 4 and column 2 of Appendix Table A.5 suggest that our main findings are not driven by these school-level sample composition issues. Specifically, we restrict the sample to schools that appear at least twice prior to and twice following internet installation (i.e., schools for which we observe both pre-trends and treatment effects).[23] In this sample, which we refer to as the "2 Pre, 2 Post" sample, we find no statistically significant trends in performance prior to internet access, and the estimated effects are similar in magnitude and show similar dynamics as those using the full sample.

Attrition from the panel may pose another compositional issue. Overall attrition in our sample is 18.2% (missing school-year observations). Attrition can happen for several reasons. First, as mentioned in Section 2.2.1, only schools with at least 5 second grade students and in which the

[23]This limits our sample to schools that gained access to the internet during a span of four (rather than the full eight) calendar years; i.e., the earliest (latest) initial year of internet access would need to be 2009 (2012) to observe two pre-internet (post-internet) periods. Our restricted sample includes only 3,670 schools versus the 5,903 in the main sample, but is highly comparable along many observable dimensions, including student achievement (see Appendix Table A.4).

Figure 4: Effect of Internet Access in the Sample of Schools Observed in at least Two Periods before and Two Periods after Internet Access: Event Study Results



The sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice prior to and twice after internet access. Coefficients capture the increase in test scores relative to the year prior to a school receiving internet access ($t = -1$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). Standard errors are clustered by school.

language of instruction is Spanish are required to administer the ECE exam. Therefore, observations will be missing when schools fall below the threshold of 5 students (or which switch to an instructional language other than Spanish). Some schools might have experienced permanent reductions in their second grade enrollment (and drop from the sample at some point), while others might alternately meet and fall below the ECE enrollment threshold from year to year (e.g., a school might have five second graders during a year and only four during the next year). Appendix Table A.6 shows that about half of overall attrition is likely due to a school dropping below the enrollment threshold.[24] The remaining attrition is either due to missing ECE scores for another reason or missing CE (covariate) information. Only a very small portion of attrition is due to school closures.

All in all, a considerable proportion of attrited observations come from missing scores in the ECE. Because the ECE is administered and graded by the Ministry of Education, schools cannot selectively decide whether to report their ECE scores or not. However, it is possible that — if changes in enrollment are negatively correlated with changes in school quality[25] — certain schools might fall below the ECE enrollment threshold in years in which they might have potentially performed worse. As suggested by our results so far, if internet does improve school performance, then we would be more likely to observe ECE scores for schools after they gain internet connection. To formally rule out this possibility, we estimate equations 1 and 2 on the restricted sample of schools with ECE scores in *every* calendar year 2007-2014. The results are displayed in Figure 5 and column 3 of Appendix Table A.5. Even among schools tested in every year of period, we find that the estimated effects of internet access are sizable and grow over time. Though the event study coefficients are slightly larger in this restricted sample, the trend break results in column 3 of Appendix Table A.5 indicate that, net of the slight pre-trend, the estimated yearly gain in test scores due to internet access is very similar to the baseline (column 1). It is worth noting that the sample with a full set of ECE scores is similar to the baseline sample along many observable dimensions, albeit higher achieving and (somewhat naturally) larger in terms of enrollment (Appendix Table A.4).[26]

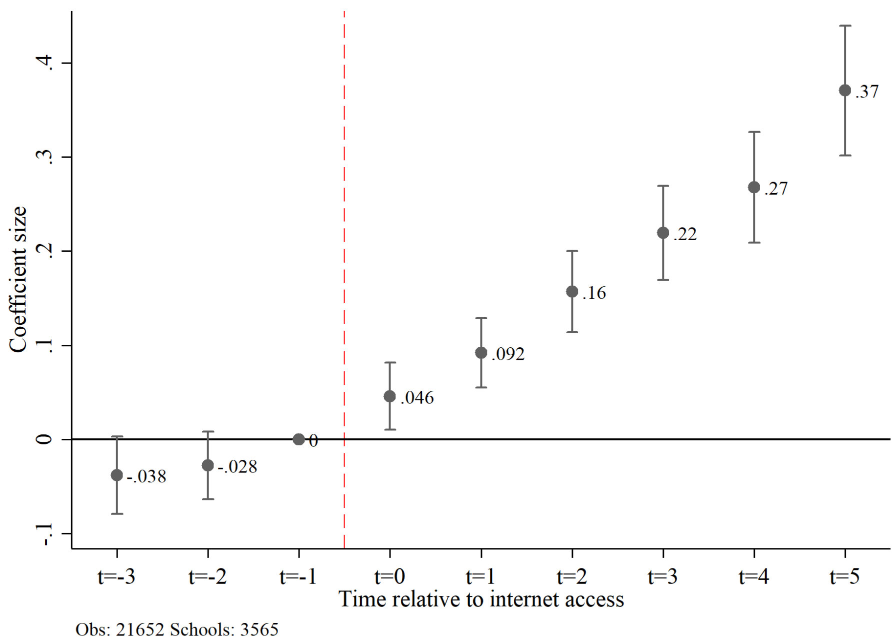
Finally, we show the robustness of our results to a very restricted sample of schools —those that appear throughout the entire event window ("complete event window" sample). This helps us rule out *both* compositional issues described above (identification of pre- and post-trends coming from potentially different schools, and attrition due to missing years in the ECE), because it requires that schools appear in all years 2007-2014 *and* that they are used to identify all seven of the

[24] A third of attrition is due to schools having fewer than 5 second grade students, and an additional sixth is explained by having enrollment "near" the threshold (defined as having 5-7 second grade students).

[25] In Section 4.1.2, we analyze the possibility that internet access increases enrollment.

[26] The data are not well suited to other methods of accounting for non-random attrition. For example, Lee bounds are not appropriate in this context, because it is not clear whether internet access would affect attrition monotonically. We do not consider inverse probability weighting because around 25% of attrited observations are missing covariate information.

Figure 5: Effect of Internet Access in the Sample of Schools with ECE Scores in All Calendar Years 2007-2014: Event Study Results



The sample includes includes all public schools that gained internet access between 2007 and 2014 and are observed for the entire sample period, i.e. for each year 2007-2014. Coefficients capture the increase in test scores relative to the year prior to a school receiving internet access ($t = -1$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). Standard errors are clustered by school.

event study dummies.[27] However, there are several important caveats to using the complete event window sample. First, for this sample we are only able to identify effects for an evaluation window that spans seven years, $t = -3$ to $t = 3$, because there are only 8 years in our sample period. This considerably limits our ability to study the dynamic path of effects, compared with our main results; we cannot estimate the $t = 4$ and $t = 5$ coefficients with this sample. Second, this limits the identifying variation in installation timing to only schools that gained internet access in 2009 and 2010. Third, the complete event window restriction shrinks the sample of schools considerably, from 5,903 to 1,043. Schools observed over a complete event window appear to be higher achieving and larger compared to the full sample overall (see Appendix Table A.4).

The results are displayed in Figure 6 and Appendix Table A.5. It is clear that using the complete event window sample makes the estimates much less precise overall, and that the short-run estimates are somewhat smaller than in the baseline sample. However, the pattern of effects is otherwise very similar. In fact, the 2- and 3-year post installation effects appear even larger in this sample (Figure 6). In Appendix Table A.5 we first reproduce the results for the full sample using the restricted evaluation window (column 4) and then using the complete event window sample (column 5). In the balanced sample, the trend break is large but imprecisely estimated. In line with the event study results in Figure 6, the immediate impact (level shift) in this sample is small (the point estimate is actually negative) and not statistically significant.

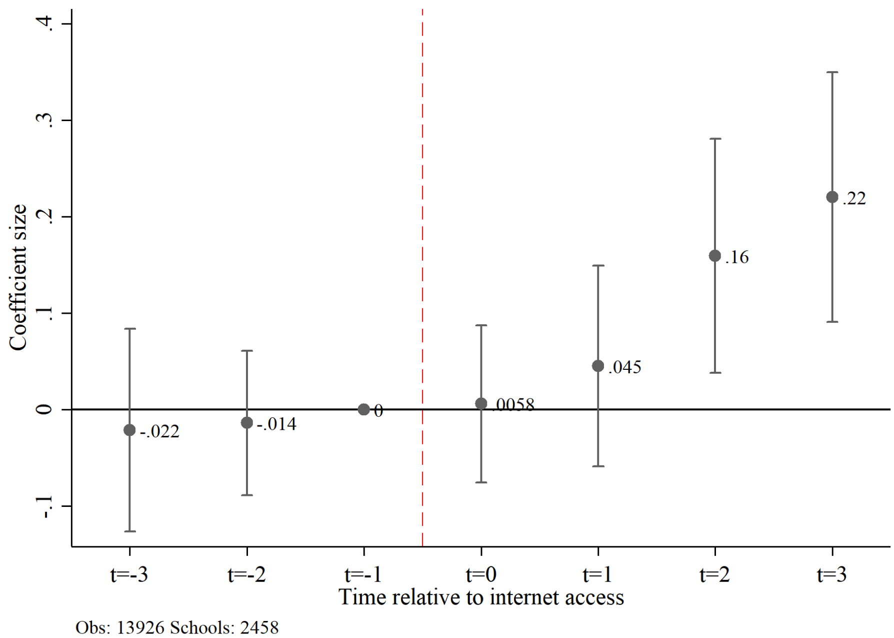
4.1.2 Student composition and endogenous sorting

Another related issue is that the composition of students *within* schools may change in response to internet access. A priori, it is hard to tell the direction of the bias that this would entail. For instance, internet access as a signal of increased school quality may imply countervailing possibilities with regard to student composition. Parents who would otherwise not have sent their kids to school might decide to enroll their children in a school connected to internet. If internet connectivity attracts students that would have otherwise performed poorly, then our estimates of treatment effects are likely conservative. Alternatively, motivated parents seeking learning opportunities for their children may decide to transfer students from schools without internet to schools that gained connectivity. If these new students are better achievers on average, then our findings of positive treatment effects may owe to upward bias from changes in student composition.

Overall, it does not appear plausible that an influx of high-achieving transfers or re-entrants explains the performance gains in our main results. To rule out this possibility, we first analyze the response of grade 2 transfers and re-entry to internet access in columns 1-2 of Table 2. Transfers are

[27] Note this is different from the results in Figure 5 and column 3 of Table A.5. For example, consider a school that gained internet access in 2007 and was tested in the ECE throughout every year 2007-2014. This school has not attrited from the sample due to ECE reporting, but only identifies the event study coefficient for $t \geq 0$ (and none of the estimates for $t < 0$). Our more restricted sample would exclude this school, as it requires that each school identifies every event study period $-3 \leq t \leq 3$.

Figure 6: Effect of Internet Access in the Sample of Schools Observed for the Entire Event Window and in All Periods (Complete Event Window Sample): Event Study Results



The sample includes includes all public schools that gained internet access between 2007 and 2014 and are observed for the entire event window and thus for the entire sample period. Coefficients capture the increase in test scores relative to the year prior to a school receiving internet access ($t = -1$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). Standard errors are clustered by school.

students enrolled in the current year who were enrolled in a different school in the previous year. Re-entrants are students that are currently enrolled but who were not enrolled in any school during the previous year (i.e., dropouts who come back to school). It appears that schools gain about 0.293 second grade transfers in the year that internet is introduced and that transfers increase by about 0.167 students in every subsequent year (column 1). However, these increases are small relative to total grade 2 enrollment (enrollment in grade 2 was 31.5, on average, prior to internet). For example, the results in column 1 predict that 5 years after internet, there will be in total about 1.1 additional transfer students. Given that 31.5 students on average take the ECE each year, it is unlikely that one additional student can substantially contribute to the observed increase in average test scores.[28] There are no apparent effects of internet access on grade 2 re-entry, though there are very few re-entrants to begin with (column 2).

Nevertheless, even if total enrollment remains relatively unchanged in response to internet connectivity, the makeup of the students that take the test could still change. For example, if internet availability induces attendance, then a different set of students will be present to take the test after a school gains internet access. We investigate this possibility in columns 3 and 4 of Table 2. Column 3 examines the effect of internet on the number of students scheduled to take the ECE, conditional on the total number of grade 2 students enrolled. It appears that, after internet is introduced to a school, the number of test takers actually declines by 0.334 students. Not only is this effect likely too small (and, for this scenario, in the "wrong" direction) to drive the estimated effects of internet access, but this also represents a one-time decrease in the number of test takers — which is unlikely to explain gradual performance gains that occur over time. These results are consistent with Cristia et al. (2017) and He et al. (2008), who find that neither hardware nor CAI/CAL interventions has any significant effects on attendance.[29] In column 4, we further explore whether student background changes in response to internet access. The only information on the background of students in the *Censo Escolar* is the proportion of native Spanish speakers enrolled.[30] To the extent that native language captures student background, it does not appear that internet access attracts more advantaged students. Overall, the evidence in Table 2 does not seem to indicate that endogenous student sorting drives our estimated impacts of internet access.

[28] Additionally, the number of grade 2 transfer students is unrelated to test scores (point estimate = 0.0003, p-value = .619) in a regression including school and year fixed effects and the controls listed in Section 3.1.

[29] Relatedly, Cristia et al. (2014) find no effects of computer and internet access on enrollment, grade repetition, or dropout in secondary public schools in Peru.

[30] The proportion of Spanish-speaking students is positively related to higher test scores, even after conditioning on school and year fixed effects and the controls listed in Section 3.1.

Table 2: Effect of Internet Access on Grade 2 Transfers, Re-entry, Test Taking, and Student Composition

	Grade 2 Transfers (1)	Grade 2 Re-entry (2)	Grade 2 Students Scheduled to Take Test (3)	Proportion of Native Spanish Speakers in Grade 2 (4)
Post-internet Access	0.293**	0.028	-0.334*	-0.002
	(0.120)	(0.031)	(0.179)	(0.005)
Post-internet Access X Event Time	0.167***	0.005	-0.133	0.000
	(0.061)	(0.017)	(0.104)	(0.002)
Pre-internet mean of dep. variable	2.866	0.372	30.89	0.850
Observations	31,368	31,368	31,368	31,357
Number of Schools	5,903	5,903	5,903	5,903
Year Fixed Effects	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes

Transfers are students enrolled in the current year who were enrolled in a different school in the previous year. Re-entrants are students that are currently enrolled but who were not enrolled in the previous year. The sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Standard errors are clustered by school. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Post-internet access is a dummy variable for whether a school has gained internet access (i.e. $t >= 0$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Event time is years relative to internet access. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). Columns 1-3 controls for both second grade enrollment and enrollment in other grades separately, not including transfers or re-entrants when specified as an outcome variable. Column 4 also controls for the number of second grade students scheduled to take the test.

Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.

4.2 Concurrent changes in school resources

Timing of internet access may also possibly correlate with changes in other school resources.[31] For example, it might be that internet provision is bundled with other inputs in a multifaceted approach to improve quality of schooling.[32] If this is the case, the improvement of students' performance that we observe might be due to increases in these other resources.

For the most part, we do not find that the timing of internet access is correlated with increases in other observable inputs (Table 3). Classrooms per students actually *fall* very slightly after internet access (a one-time shift, column 1). In column 2, we report small and statistically insignificant changes in the number of textbooks per student in schools that gain access to internet. We do observe small increases in both teachers and computers per student. Teachers per students rise steadily after internet access, though the magnitudes are small in both absolute terms and relative to the pre-internet mean (column 3). In column 4, we note a positive and non-negligible level increase in computing resources at the time of internet installation (on the other hand, the estimate of the trend break is near zero and statistically insignificant).

To better understand the potential for increases in these resources to confound our estimates of the impact of internet access, we perform some back-of-the envelope calculations. Note that, by definition, internet connectivity is complementary to computer access (i.e., generally students cannot use the internet *without* computers). However, students might nevertheless benefit from computers without access to internet (e.g., using preloaded software and resources installed from flash drives / DVDs, etc.). Joint increases in computer and internet availability might thus possibly imply that our estimates capture (at least partially) the utility of computers themselves apart from internet access. Similarly, while internet access may be a complement to or substitute for teachers, if access is accompanied by increases in teachers, that may exert an independent effect on student performance that is not attributed directly to internet access itself.

By our calculations (displayed in Appendix Table A.8), we find that the increases in teachers and computers alone explain very little of the observed rise in test scores (at most, 7.2% for teachers and 3.5% for computers). For these calculations we use two pieces of information: (i) the impact

[31] Though all of our specifications include school and year effects as well as time-varying school characteristics, it is still possible that our results capture in part the effect of time-varying school-level unobservables. Oster (2017) gives an alternative method of ruling out bias due to unobserved heterogeneity, based on movements in estimated treatment effects and the explanatory power (R^2) of models estimated with and without (observable) control variables. We apply her method to equation 2 to estimate bounds on ϕ_1 and ϕ_3. Following the parameters suggested in Oster (2017), we assume: (a) a coefficient of proportionality of 1 between observable and unobservable controls, and (b) an increase in R^2 of 30% due to the inclusion of unobservables. Under these assumptions, we estimate hypothetical coefficients adjusted for omission of unobservables for ϕ_1 and ϕ_3. We find that adjusted estimates fall within the 95% confidence intervals of our estimated coefficients and conclude that it is unlikely our estimates capture solely effects of unobservables.

[32] It could also be the case that internet access at schools is correlated with alternative sources of internet. For example, students who gain internet access at school may already have internet connections at home or via cyber-cafes. However, we find that only 23% (32%) of students with access to internet at school also use it at cyber-cafes (or at home) according to the 2014 Peruvian National Household Survey (ENAHO). Additionally, our results are unchanged if we include a control for whether the town nearest the school has a cyber cafe.

Table 3: Effect of Internet Access on School Resources

	Classrooms per Student (1)	Textbooks per Student (2)	Teachers per Student (3)	Computers per Student (4)
Post-internet Access	-0.005*	-0.132	0.003***	0.045***
	(0.003)	(0.088)	(0.001)	(0.006)
Post-internet Access X Event Time	-0.001	-0.063	0.002***	-0.001
	(0.001)	(0.046)	(0.001)	(0.003)
Pre-internet mean of dep. variable	0.0721	3.932	0.0549	0.0676
Observations	31,368	23,636	31,368	31,368
Number of Schools	5,903	5,857	5,903	5,903
Year Fixed Effects	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes

In column 2, the dependent variable is the number of 2nd grade textbooks per 2nd grade student. Textbook information is not available for 2012, so the sample size is slightly smaller for this outcome. In column 3, the number of teachers per students excludes computer teachers (discussed separately in Section 5.2). In column 4, the dependent variable is the number of instructional computers per student. The sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Standard errors are clustered by school. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Post-internet access is a dummy variable for whether a school has gained internet access (i.e. $t >= 0$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Event time is years relative to internet access. Control variables include (excluding the dependent variable) enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers).
Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.

of internet access on teachers and computer resources and (ii) the impact of teachers/computer resources alone (i.e., without connections to the internet) on student test scores. Our estimates of (i) comes from columns 3 and 4 of Table 3 above. For example, column 3 suggests that the approximate rise in teachers concurrent with internet access follows the formulas $0.003 + 0.002 \times t$, where t is years after internet access. To approximate (ii), we use the sample of schools that do not gain access to the internet during our sample period and regress math scores on teachers and computers per student (including school and year fixed effects as well as all of the same controls listed in Section 3.1). These results are displayed in Appendix Table A.7. In each successive column, we add in lags of teachers/computers per student to allow for dynamic effects of these resources on student performance. Using both (i) and (ii), we can calculate the approximate gain in math scores due to increases in teaching and computing resources (*without* internet access) that occur around the introduction of internet.

Appendix Table A.8 displays these calculations. For example, let us examine the calculations for computers in $t = 0$ (first row of bottom panel). Computers per student rise by 0.045 when internet is introduced (column 4, Table 3). Additionally, the largest estimate of contemporaneous (non-internet connected) computers' impact on test scores is 0.031 (from column 1 of Appendix Table A.7). Based on both estimates, we would expect a test score gain of $0.045 \times 0.031 = 0.001$ standard deviations at $t = 0$. From our event study specification, our overall estimated effect of internet access at $t = 0$ is 0.042 standard deviations. Therefore, we estimate that the increase in computers per student alone at $t = 0$ explains only about $0.001/0.042 = 3.4\%$ of the rise in test scores that we observe in response to internet access. The calculations for other post-internet periods are more complex when the effect of resources is allowed to be dynamic. However, our most "generous" estimates indicate that, at most, acquisition of more teachers and computers explains only a small portion of the estimated gains in math following internet access.[33] These calculations aside, we include time-varying measures of these resources in our control set to ensure that the internet effects we identify are conditional on the teachers and computers available to the students.

4.3 Differential pre-trends

Another possibility is that access to internet is correlated with pre-existing trends in test scores. For example, states with faster growing economies might be better able to finance internet expansions, increase public spending on education, or otherwise improve student learning. In Table 4 we show that our results are robust to allowing for a wide array of group-specific pre-trends and group-year fixed effects. These include groups that are defined geographically: *departamento* (state), state-sector (urban/rural areas within states), province, province-sector, district, and district-sector.[34]

[33] To get the "most generous" estimates, we take the largest individual estimated effect of teachers/computers for each lag across all specifications in Appendix Table A.7, regardless of significance level.

[34] As there are 1,889 districts (and 2,989 district-sector combinations) in Peru, we do not include district-year fixed effects; doing so would add 13,223 (20,923) additional covariates to the model.

We consider other groups defined administratively (i.e., by Local Educational Management Unit, UGEL), by timing of internet installation, by initial academic performance, or even by individual school.

The point estimates of the level shift and trend break in test scores vary across specifications, but they are all positive and generally statistically significant. Importantly, we can show that accounting for any pre-existing trends in installation year cohort (i.e., for schools that install internet in a given year) yields, if anything, slightly larger estimated impacts of internet access (column 16). This helps rule out the possibility that our estimated effects are driven by selection — in other words, that better schools exhibiting steeper score trajectories receive internet access earlier.[35] Similarly, we show that our estimates are stable even when accounting for differential trends by initial academic performance as measured by initial scoring decile (columns 17 and 18).

The trend break coefficients' p-values are 0.1 or lower for all specifications except the one that includes school-specific linear trends.[36] It is worth emphasizing that the estimation of school-specific linear trends — which add an additional 3,670 covariates to the model — likely absorbs much of any potential exogenous variation in test scores and internet access. As expected, precision drops considerably in this specification. Thus, while we are unable to reject the null hypothesis of zero effects when including the additional trends, we are also unable to reject that these estimates are equal to our baseline estimates. Overall, we take the evidence in Table 4 to indicate that pre-existing trends in test scores, or coincident shocks to scores, do not confound our estimates of the effect of internet access.

5 Explaining dynamics and identifying potential mechanisms

5.1 How Prevalent in Internet Use in Schools?

While the CE provides information about internet access, it does not gather data about how schools use internet when they gain connections. To aid the interpretation of our main results, we present suggestive evidence based on descriptive statistics from two nationally representative surveys of schools in Peru: the 2014 National Survey of Educational Institutions (ENIE) and the 2014 National Survey of Teachers (ENDO).[37] These surveys directly interview teachers and include information about their use of internet use in classrooms and their perceptions of the advantages of ICTs in education. Unfortunately, neither surveys provide school identifiers linking these surveys

[35] We cannot include cohort-year fixed effects, because these would be collinear with our covariates of interest.

[36] The inclusion of school-specific pre-trends limits our sample to those schools within which we can identify both pre- and post-internet trends: namely, our "2 Pre & 2 Post" sample.

[37] The ENIE (*Encuesta Nacional a Instituciones Educativas*) is collected by the National Statistical Institute (INEI) among principals, primary school teachers (first and second grade instructors), and secondary school teachers (seventh and eleventh grade instructors) in public schools. The ENDO (*Encuesta Nacional a Docentes*) is collected by the Ministry of Education among school teachers (any grade) in private and public primary and secondary schools in Peru.

Table 4: Allowing for Differential Linear Pre-trends and Year Effects by Various Groups

	Baseline (1)	State-specific Trends (2)	State-Yr FE (3)	State-Sector Trends (4)	State-Sector-Year FE (5)
Post-internet Access	0.036*	0.039**	0.027	0.038*	0.019
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)
Post-internet X Event Time	0.047***	0.045***	0.034***	0.041***	0.023**
	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
Observations	31,368	31,368	31,368	31,368	31,368
No. of schools	5,903	5,903	5,903	5,903	5,903
No. of groups	N/A	26	182	51	357
	Prov.-specific Trends (6)	Province-Yr FE (7)	Prov.-Sector Trends (8)	Prov.-Sector Yr FE (9)	Dist.-Specific Trends (10)
Post-internet Access	0.038*	0.022	0.039*	0.017	0.031
	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)
Post-internet X Event Time	0.042***	0.027**	0.040***	0.019*	0.051***
	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)
Observations	31,368	31,368	31,368	31,368	31,368
No. of schools	5,903	5,903	5,903	5,903	5,903
No. of groups	195	1,365	388	2,716	1,889
	Dist.-Sector Trends (11)	UGEL-specific Trends (12)	UGEL-Year FE (13)	UGEL-Sector Trends (14)	UGEL-Sector Year FE (15)
Post-internet Access	0.018	0.037*	0.021	0.037*	0.034*
	(0.022)	(0.020)	(0.021)	(0.020)	(0.020)
Post-internet X Event Time	0.022*	0.044***	0.030***	0.041***	0.048***
	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)
Observations	31,368	31,368	31,368	31,368	31,368
No. of schools	5,903	5,903	5,903	5,903	5,903
No. of groups	2,989	219	1,533	427	1,327
	Installation Year (cohort)-specific Trends (16)		Initial Scoring Decile Trends (17)	Initial Scoring Decile-Year FE (18)	School-specific Trends (19)
Post-internet Access	0.036*		0.041**	0.034*	0.024
	(0.020)		(0.020)	(0.020)	(0.029)
Post-internet X Event Time	0.072***		0.054***	0.041***	0.031
	(0.012)		(0.011)	(0.011)	(0.038)
Observations	31,368		27,141	27,141	22,321
No. of schools	5,903		4,915	4,915	3,670
No. of groups	8		10	70	3,670

"No. of groups" refers to the number of groups used to create additional controls listed in each column; e.g., in col. 2 including state-specific linear trends adds an additional regressor per state (26). In cols 1-16, the sample includes all public primary schools that gained internet access between 2007 and 2014 and are observed at least twice. In cols 17-18, the sample is further restricted to schools that are observed prior to internet access. Initial scoring decile is defined as the school's scoring decile based on pre-internet scoring average; the decile is calculated based on all estimation sample schools that have pre-internet scores. In col. 19, the sample is further restricted to schools that are observed at least twice prior to and after gaining internet access. Math scores are standardized to have mean zero and standard deviation of 1 within each year. Post-internet access is a dummy variable for whether a school has gained internet access. Event time is years relative to internet access. Controls: enrollment, no. of 2nd grade students scheduled to take the ECE, computer room, library, administrative offices, and resources per student (classrooms, computers, and teachers), school and year fixed effects and the additional fixed effects and/or trends indicated in each column heading. Standard errors are clustered by school. *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

to the CE or ECE. However, they allow us to characterize teachers' approaches to internet use in Peruvian schools by the end of our period of analysis.

We use these datasets to gauge students' and teachers' use of school-based internet. The ENIE suggests a considerable degree of internet usage by 2014: 65.6% of second grade teachers at internet equipped schools report using internet in class. Among those who use it, 76% report using internet at least once per week (and, on average, for 1.85 hours per week. Information in the ENDO also supports the notion of regular internet use in classrooms. As of 2014, 31% of Peruvian public primary school teachers at internet-connected schools listed internet-connected computers among the top 3 most-used classroom tools.[38]

Additionally, we use the ENDO to characterize teachers' attitudes towards and perceived benefits of ICTs. On one hand, teachers find the internet useful for their students. Among second grade teachers in pubic schools with internet access, 83.2% believe internet increases students' access to information (that is probably otherwise unavailable to them) and 81.7% state that it improves collaborative learning among students. On the other hand, student use of internet-equipped computers is only one way in which internet access could benefit students; it also can serve as a tool for teachers themselves. For example, teachers can access "off the shelf" lesson plans, repositories of practice questions, instructional aids, etc.[39] 63% of teachers at internet-equipped primary schools considered ICTs a top 3 factor in enhancing teaching — a larger proportion than listed either reference materials (38%) or networking with colleagues (38%) (ENDO 2014). Thus, teacher access to online materials may boost student performance above and beyond the impacts of direct student use of internet-connected computers.

These additional surveys provide suggestive evidence on the extent of internet use in schools and teacher's attitudes on how it can aid learning. Unfortunately, we lack data on more specific ways in which teachers make use of internet access in our sample schools. Instead we now turn to additional analysis using data on school availability of particular resources and hetereogeneity to supplement our main results and explore two potential mechanisms by which internet access improves student learning.

5.2 Why do the effects of internet access take time to emerge?

One explanation for why we observe delayed impacts of internet access may be that schools require teachers with digital and internet skills in order to incorporate the new technology into the classroom. To investigate this possibility, we study whether schools respond to internet access by hiring teachers with expertise in "computer and information technology." This includes both teachers

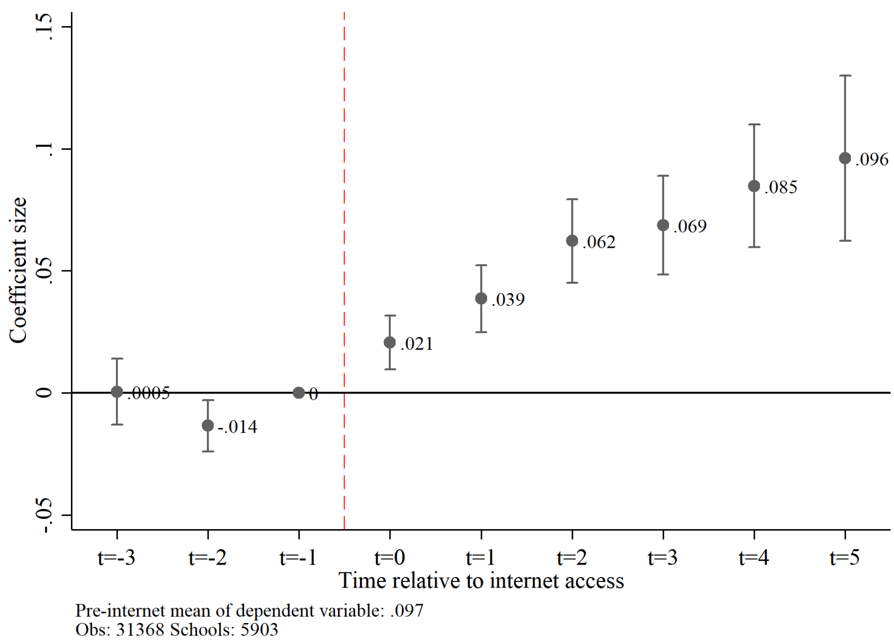
[38]Among schools with internet access, only very basic materials (such as photocopies and flip charts) score higher in this ranking, relative to internet-connected computers.

[39]For example, websites such as CarpetaPedadgogica.com provide teachers with a wealth of free educational resources, including detailed classroom activities for various learning objectives, links to educational videos, sample questions, and more.

trained to teach computer skills, as well as teachers who themselves underwent advanced education relating to computers; hereafter, these are referred to as "computer teachers." We estimate equation 1 using an indicator for the presence of a computer teacher as the outcome.

Figure 7 shows that internet access is accompanied by a slow but steady increase in the likelihood a school has a computer teacher; by year 5, this results in a doubling of the pre-internet likelihood. When taken together, the findings for computer teachers and test scores are consistent with the idea that schools may need time to make complementary investments to fully exploit new classroom technologies, such as teachers with computer training. That the timing of hiring computer teachers corresponds (with a delay) in the timing of test score gains suggests that the two are related. However, because the presence of a computer teacher is a function of internet access (and is not exogenously given), we are not fully able to show the impact of the complementarity of both inputs on school performance.[40]

Figure 7: Internet Access and Presence of a Computer Teacher



Coefficients capture the increase in the likelihood of having a computer teacher on staff relative to the year prior to a school receiving internet access ($t = -1$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). The sample includes all schools with at least two observations within the sample period (2007-2014). Standard errors are clustered by school.

[40] For example, we do not estimate equation 2 including a triple interaction between post-internet access, event time, and presence of a computer teacher because as shown in Figure 7, schools hire computer teachers *as a lagged response* to gaining internet access. Therefore, it would be unclear how to interpret the results of such triple interaction.

5.3 Increased individualized attention

Previous work suggests that ICTs may enable providing students with more individualized attention than they would otherwise receive from teachers. To explore whether increased individualized attention might explain our results, we follow Barrow et al. (2009) and examine heterogeneity along the lines of class size. In theory, teachers divide their time in classrooms between group and individualized instruction. If ICTs reduce the time teachers spend in group activities, they might be able to increase the time they allocate to individualized instruction. In particular, teachers assigned to larger classes might be more constrained in providing individualized instruction, and thus may be expected to see larger gains from ICTs.

Alternatively, internet access might be especially useful in strengthening the effectiveness of group work. For example, children may be more likely to focus on a group learning activity that involves watching a video or playing an educational game online than more traditional paper- or text-based activities. As mentioned, the overwhelming majority (81.7%) of teachers in internet-connected schools believe that internet access enhances collaborative learning among students (ENDO, 2014). This implies that internet connectivity would be very useful in classrooms with many students, where teachers rely heavily on group work. For both of these reasons, we expect internet access to matter more for schools with high versus low student to teacher ratios (STR).

Splitting schools by the pre-internet median STR, we find, as expected, that the positive effects of internet access are concentrated among schools with high STRs. We define "high STR" and "low STR" groups as follows. First, we calculate the total number of teachers per second grade student (we do not use the number of teachers exclusively dedicated to second grade, because many smaller schools assign teachers to multiple grades). Then, we calculate each school's pre-internet average STR (time-invariant). Finally, we divide the schools into high and low STR groups based on having a pre-internet average STR above or below the median.[41] In Figure 8a, the high and low STR trends in test scores prior to internet access are nearly identical, but diverge once internet is introduced. In low STR schools, the effects are much smaller, though the estimates are not very precise in either subsample. Appendix Table A.10 (Columns 1-2) confirms that the trend break in test scores is larger and only statistically significant in high STR schools, though the point estimate of the immediate effect of internet access is larger in low STR schools. Thus, these results are qualitatively consistent with increased individualized attention as a causal pathway through which internet access improves student performance.

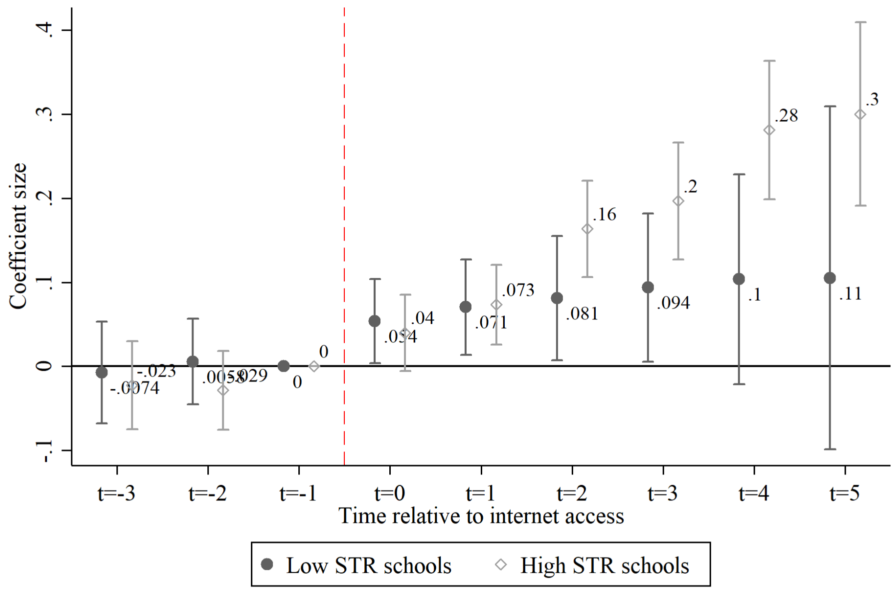
5.4 Substitution between internet access and teacher qualifications

Another possibility is that ICTs generate gains in student learning because they compensate for the lack or low quality of other inputs. Relatedly, some have found that the success of ICT interventions may depend on whether they displace traditional instruction or constitute additional learning

[41] The average 2nd grade student-teacher ratios in the high and low STR groups are 8.28 and 2.77, respectively.

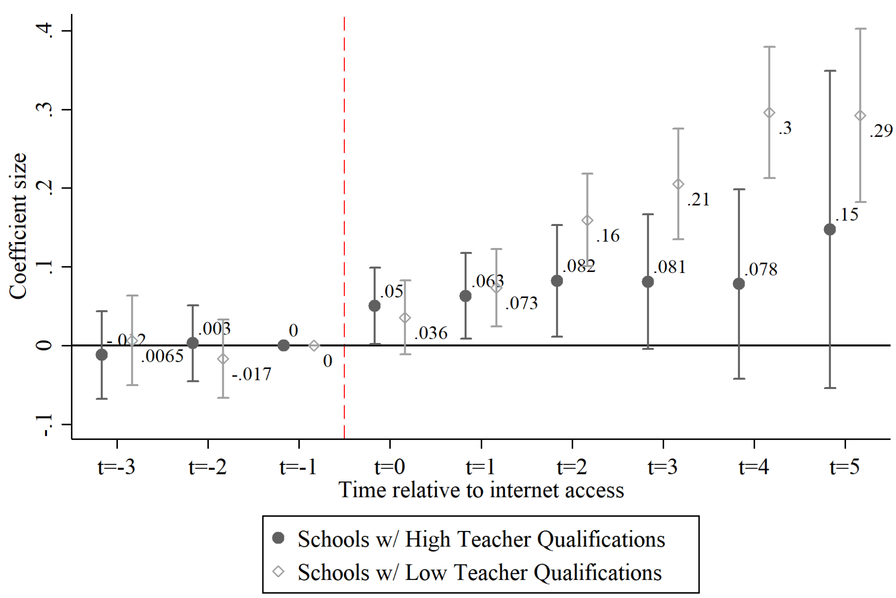
Figure 8: Heterogeneity in the Impact of Internet Access on Test Scores: Event Study Results

(a) Heterogeneity by Student to Teacher Ratios



Obs (Low STR Schools): 12834 Number of Low STR Schools: 2457
Obs (High STR Schools): 14307 Number of High STR Schools: 2458

(b) Heterogeneity by Teacher Qualifications



Obs (Schools w/ High Teacher Qual.): 13215 Number of Schools w/ High Teacher Qual.: 2457
Obs (Schools w/ Low Teacher Qual.): 13926 Number of Schools w/ Low Teacher Qual.: 2458

Figure (a): The sample is split based on each school's pre-internet average ratio of second graders to total teachers (STR) — high (low) STR schools fall above (below) the median pre-internet STR. Figure (b): The sample is split based on each school's pre-internet average number of teachers with a pedagogical or higher education degree per student over the sample period relative to the median of all schools' sample averages. Coefficients capture the increase in test scores relative to the year prior to a school receiving internet access ($t = -1$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). The sample includes all schools with at least two observations within the sample period (2007-2014). Standard errors are clustered by school.

activities outside of traditional classroom hours (as part of an after school tutoring program, for example, as in Linden 2008). In cases where ICTs substitute for traditional instruction, impacts may depend on the quality of instruction that the new technology is displacing. Such is hypothesized in Bulman and Fairlie (2016, p. 20), "[...] Interestingly, evidence of positive effects appears to be the strongest in developing countries. This could be due to the fact that the instruction that is being substituted for is not as of high quality in these countries."

To shed some light on whether internet access may substitute for low teacher quality in the context of our study, we examine heterogeneity in results by the level of qualification that a school's teachers have obtained. In Figure 8b, we see that schools with low teacher qualifications experience relatively larger gains in test scores over time, though all schools see a moderate rise in scores immediately following access. The trend break among low teacher qualification schools is large (0.063 standard deviations per year after internet access) and statistically different from the trend break among high qualification schools (columns 3 and 4 in Appendix Table A.10). Here, we measure teacher qualification as the per student number of teachers with a pedagogical or university degree. We estimate the average ratio of qualified teachers-to-students by school using the pre-internet period, and split the sample in two groups based on the sample median across schools. Those with ratios above (below) the sample median are classified as schools with "high" ("low") teacher qualifications.[42]

These results align with findings from other contexts, such as Jackson and Makarin (2016), who determine that the benefits — in terms of math achievement — of providing teachers with online access to "off the shelf" lesson plans were larger among students with weaker teachers. Importantly, if technological tools can substitute for teacher quality, ICT interventions such as school-based internet may help poor schools narrow or close achievement gaps vis-à-vis wealthier schools.[43]

6 Conclusions

We find evidence that the introduction of internet to Peruvian primary schools produces economically meaningful improvements in student performance (as measured by standardized test scores

[42]Schools with high and low teacher qualifications are distinct from schools with high and low STR; about a quarter of high qualification schools are high STR schools and around 18% of low qualification schools are low STR schools. The analysis in Figure 8b and Appendix Table A.10 also controls for the overall STR.

[43]Another possible interpretation of the results in Figure 8b is that less qualified teachers are simply more likely to use or are better at using internet-based materials. In the 2014 ENDO, there is very little gradient in the usage of internet-connected computers with respect to teacher qualifications; if anything, more qualified teachers (e.g. those holding a college degree or higher) are more likely to report internet-connected PCs as one of the three most often used classroom tools (34% versus 28%). Perceptions of the usefulness of internet connections are also similar across teacher qualifications; around 76% (86%) of teachers with and without college degrees expect internet connectivity to help teachers in the classroom (increase students' access to information). The results in Figure 8b are hence *not* consistent with less qualified teachers being less likely to adapt to new technologies: schools with relatively high proportions of less qualified teachers are the ones for which the test score gains are largest.

for grade 2). Gains increase over time, growing from 0.042 standard deviations in the year of installation to 0.29 standard deviations 5 years after installation. Importantly, there are no apparent pre-existing trends in math test scores prior to internet access, suggesting little role for reverse causality. Using a trend break specification, we confirm that there is a level shift and (linear) trend break in test scores that occurs at the time of internet access. In the medium term, the yearly gain in test scores is about 0.047 standard deviations. These results, which are representative of about one quarter of all public primary schools in Peru, are robust to a number of potential confounding factors, including changes in sample composition with respect to either schools or students, changes in school resources, and endogenous timing of installation with respect to prior trends in test performance. In our setting, the nationwide scale of roll-out, large sample of schools, and extended time frame uniquely enable the analysis of this technology's application at the farthest-reaching level of policy.

On the one hand, previous research on ICTs has found that providing hardware with few or no complementary learning tools has little immediate impact on student performance (Bet et al. 2014; Barrera-Osorio and Linden 2009; Cristia et al. 2017; etc.). Our short run results (based on up to 1 year after internet installation) confirm that any effects are small in magnitude — and thus perhaps impossible to detect in the small samples used by this literature's many RCT studies. On the other hand, medium run gains are sizable, pointing towards the necessity of a longer evaluation window for understanding the effectiveness of ICT interventions. Ultimately, our estimated effects of internet access for years 2-5 still fall below prior estimates of the impact of computer assisted learning and instruction. Even so, while school-based internet does not fully confer the benefits of individualized pedagogical tools, it may provide access to learning resources that are otherwise unavailable to many students in developing countries.

We provide supporting evidence that achievement gains are slow to emerge because schools need time to adapt to new technologies. Specifically, after installing internet public schools require time to augment their staff with teachers experienced in computers and information technology. We thus concur with several prior studies finding that student achievement begins to increase only as teachers learn to integrate new technology into their curricula (Hull and Duch 2016, Mo et al. 2013, Sprietsma 2007).

Our data also yield evidence suggestive of two channels through which school-based internet access facilitates human capital accumulation: allowance of greater individualized instruction and substitution for low teacher qualifications. Gains in math scores are concentrated among schools that have high student-teacher ratios and in which relatively few teachers hold pedagogical or university degrees. Hence, school-based internet may generate important gains in learning particularly when individualized instruction and teacher quality are constrained below the optimum.

However, interpretation of the results presented is subject to a number of limitations. Perhaps most notably, school-level analysis may mask important individual-level dynamics. We are largely unable to explore heterogeneity in the effectiveness of school-based internet based on student characteristics. Indeed, previous work suggests that individual heterogeneity - especially with regard to initial achievement - significantly determines how technology affects the learning process (Bai et al. 2016; Barrow et al. 2009; Linden 2008; He et al. 2008; Muralidharan et al. 2016). Future research on heterogeneous impacts of internet in education could bear broad implications for inequality within and across learning environments.

References

Angrist, J. and Lavy, V. (2002), 'New Evidence on Classroom Computers and Pupil Learning', *The Economic Journal* **112**(482), 735–765. 15

Bai, Y., Mo, D., Zhang, L., Boswell, M. and Rozelle, S. (2016), 'The impact of integrating ICT with teaching: Evidence from a randomized controlled trial in rural schools in China', *Computers & Education* **96**, 1–14. 1, 37

Bando, R., Gallego, F., Gertler, P. and Romero, D. (2016), Books or Laptops? The Cost-Effectiveness of Shifting from Printed to Digital Delivery of Educational Content, NBER Working Paper 22928, National Bureau of Economic Research, Cambridge, MA. 1, 15

Banerjee, A. V., Cole, S., Duflo, E. and Linden, L. (2007), 'Remedying Education: Evidence from Two Randomized Experiments in India', *Quarterly Journal of Economics* **122**(3), 1235–1264. 1, 12, 15

Barrera-Osorio, F. and Linden, L. L. (2009), The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia, Policy Research Working Paper 4836, The World Bank, Washington, DC. 1, 4, 15, 36

Barrow, L., Markman, L. and Rouse, C. (2009), 'Technology's Edge: The Educational Benefits of Computer-Aided Instruction', *American Economic Journal: Economic Policy* **1**(1), 52–74. 33, 37

Barseghian, T. (2011), Teaching Strategies: Are Online Math Programs Better Than Literacy?
URL: *http://tinyurl.com/ybzvtv8f* 12

Belo, R., Ferreira, P. and Telang, R. (2014), 'Broadband in School: Impact on Student Performance', *Management Science* **60**(2), 265–282. 1

Bet, G., Ibarraran, P. and Cristia, J. (2014), The Effects of Shared SchoolTechnology Access on Students' Digital Skills in Peru, IDB Working Paper Series 476, Inter-American Development Bank, Washington, DC. 2, 4, 15, 36

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O. and Cruz-Aguayo, Y. (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. 1, 4, 6, 15

Bulman, G. and Fairlie, R. W. (2016), Technology and Education: Computers, Software, and the Internet, NBER Working Paper 22237, National Bureau of Economic Research, Cambridge, MA. 1, 35

Carrillo, P., Onofa, M. and Ponce, J. (2010), Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador, IDB Working Paper Series 223, Inter-American Development Bank, Washington, DC. 1, 12, 15

Cristia, J., Cueto, S., Ibarraran, P., Santiago, A. and Severin, E. (2017), 'Technology and Child Development: Evidence from the One Laptop per Child Program', *American Economic Journal: Applied Economics* **9**(3), 295–320. 1, 4, 6, 15, 24, 36

Cristia, J., Czerwonko, A. and Garofalo, P. (2014), 'Does Technology in Schools Affect Repetition, Dropout and Enrollment? Evidence from Peru', *Journal of Applied Economics* **17**(11), 89–111. 2, 24

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. and Sundararaman, V. (2013), 'School Inputs, Household Substitution, and Test Scores', *American Economic Journal: Applied Economics* **5**(2), 29–57. 1

de Melo, G., Machado, A., Miranda, A. and Viera, M. (2013), Impacto del Plan Ceibal en el Aprendizaje. Evidencia de la Mayor Experiencia OLPC, Serie Documentos de Trabajo DT 13, Instituto de Economía, Facultad de Ciencias Económicas y de Administración, Unversidad de la República, Montevideo, Uruguay. 1

Escueta, M., Quan, V., Nickow, A. J. and Oreopoulos, P. (2017), 'Education Technology: An Evidence-based Review', *NBER Working Paper* (23744). 1

Faber, B., Sanchis-Guarner, R. and Weinhardt, F. (2015), 'ICT and Education: Evidence from Student Home Addresses', *NBER Working Paper* (21306). 1

Gibson, S. and Oberg, D. (2004), 'Visions and Realities of Internet Use in Schools: Canadian Perspectives', *British Journal of Educational Technology* **35**(5), 569–585. 1

Glewwe, P. and Kremer, M. (2006), Schools, Teachers, and Education Outcomes in Developing Countries, *in* E. A. Hanushek and F. Welch, eds, 'Handbook of the Economics of Education', Vol. 2, North Holland, chapter 16, pp. 945–1017. 1

Goolsbee, A. and Guryan, J. (2006), 'The Impact of Internet Subsidies in Public Schools', *The Review of Economics and Statistics* **88**(2), 336–347. 1

He, F., Linden, L. L. and MacLeod, M. (2008), How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program. 1, 15, 24, 37

Hjort, J. and Poulsen, J. (2017), 'The Arrival of Fast Internet and Employment in Africa', *NBER Working Paper* (23582). 4

Hopkins, A. (2014), Internet in Schools. The effect on educational performance in Peru: 2007-2011. 13

Hull, M. and Duch, K. (2016), One-to-one Technology and Student Outcomes, Unpublished manuscript, Department of Economics, University of North Carolina at Greensboro, Greensboro, NC. 3, 15, 36

International Telecommunication Union (2014), Final WSIS Target Review: Achievements, Challenges and the Way Forward, Technical report, Partnership on Measuring ICT for Development, Geneva, Switzerland. 1

Jackson, C. K. and Makarin, A. (2016), 'Can Online Off-The-Shelf Lessons Improve Student Outcomes? Evidence from A Field Experiment', *NBER Working Paper* (22398). 1, 3, 35

Kremer, M., Brannen, C. and Glennerster, R. (2013), 'The Challenge of Education and Learning in the Developing World', *Science* **340**(6130), 297–300. 1

Levin, D. and Arafeh, S. (2002), The Digital Disconnect: The Widening Gap between Internet-Savvy Students and their Schools, Technical report, American Institutes for Research, Washington, DC. Report prepared for the Pew Internet and American Life Project. 1

Linden, L. L. (2008), Complement or substitute?: The effect of technology on student achievement in India, InfoDev Working Paper 22928, World Bank, Washington, DC. 1, 15, 35, 37

Machin, S., McNally, S. and Silva, O. (2007), 'New Technology in Schools: Is There a Payoff?', *The Economic Journal* **117**(522), 1145–1167. 1

Malamud, O., Cueto, S., Cristia, J. and Beuermann, D. W. (2018), Do Children Benefit from Internet Access? Experimental Evidence from a Developing Country. Mimeo. Inter-American Development Bank. 1

Meza-Cordero, J. A. (2017), 'Learn to Play and Play to Learn: Evaluation of the One Laptop per Child Program in Costa Rica', *Journal of International Development* **29**(1), 3–31. 1, 15

Mo, D., Huang, W., Shi, Y., Zhang, L., Boswell, M. and Rozelle, S. (2015), 'Computer technology in education: Evidence from a pooled study of computer assisted learning programs among rural students in China', *China Economic Review* **36**, 131–145. 12

Mo, D., Swinnen, J., Zhang, L., Yi, H., Qu, Q., Boswell, M. and Rozelle, S. (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. 1, 4, 12, 15, 36

Mo, D., Zhang, L., Wang, J., Huang, W., Shi, Y., Boswell, M. and Rozelle, S. (2015), 'Persistence of learning gains from computer assisted learning: Experimental evidence from China', *Journal of Computer Assisted Learning* **31**(6), 562–581. 3, 15

Muralidharan, K., Singh, A. and Ganimian, A. J. (2016), 'Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India', *NBER Working Paper* (22923). 1, 12, 15, 37

Muralidharan, K. and Sundararaman, V. (2013), 'Contract Teachers: Experimental Evidence from India', *NBER Working Paper* (19440). 1

National Public Radio (2012), 'One Child, One Laptop ... And Mixed Results In Peru', Article for Weekend Edition Saturday on October 11, 2012. Last accessed March 2016.
URL: *http://www.npr.org/2012/10/13/162719126/one-child-one-laptop-and-mixed-results-in-peru* 4

OECD (2015), *Students, Computers and Learning: Making the Connection*, OECD Publishing. 1

One Laptop per Child (2011), 'Internet Connectivity: The Achilles Heel of the Peru Deployment'. Last accessed March 2016.
URL: *http://www.olpcnews.com/countries/peru* 4

One Laptop per Child (2016), '2016 Annual Report: A Year of Growth'.
URL: *http://one.laptop.org/magazine/annualreport/index.html* 1

Oster, E. (2017), 'Unobservable Selection and Coefficient Stability: Theory and Evidence', *Journal of Business and Economic Statistics, forthcoming* . 26

Parr, J. M. and Fung, I. (2000), A Review of the Literature on Computer-Assisted Learning, particularly Integrated Learning Systems, and Outcomes with respect to Literacy and Numeracy, Report prepared for the ministry of education of new zealand, Auckland Uniservices Limited, The University of Auckland, Auckland, Australia. 12

Purcell, K., Heaps, A., Buchanan, J. and Friedrich, L. (2013), How Teachers are using Technology at Home and in their Classrooms, Technical report, PEW Research Center, Washington, DC. 1

Saavedra, J. and Suarez, P. (2002), El Financiamiento de la Educación en el Perú: El Rol de las Familias, Documento de Trabajo 38, Grupo de Análisis para el Desarrollo (GRADE), Lima, Peru. 5

Sharma, U. (2014), Can Computers Increase Human Capital in Developing Countries? An Evaluation of Nepal's One Laptop per Child Program. Paper presented at the 2014 Annual Meeting of the Agricultural and Applied Economics Association. 1, 15

Sprietsma, M. (2007), Computers as Pedagogical Tools in Brazil: A Pseudo-Panel Analysis, ZEW Discussion Paper 07-040, Centre for European Economic Research, Mannheim, Germany. 2, 36

Texas Center for Educational Research (2009), Evaluation of the Texas Technology Immersion Pilot, Report prepared for the texas education agency, Texas Center for Educational Research, Austin, TX. 12

The World Bank (2016), 'World Development Indicators'. Accessed August 2017.
URL: *https://openknowledge.worldbank.org/handle/10986/23969* 5

Toyama, K. (2015), *Geek Heresy: Rescuing Social Change from the Cult of Technology*, first edn, PublicAffairs, New York, NY. 1

Trucano, M. (2012), Evaluating One Laptop Per Child (OLPC) in Peru. Post in the EduTech blog.
URL: *http://blogs.worldbank.org/edutech/olpc-peru2* 6

Trucano, M. (2016), 'Technologies in Education Across the Americas: The promise and the peril and some potential ways forward', *World Bank Education, Technology & Innovation: SABER-ICT Technical Paper* (12). 1

UNESCO (2012), ICT IN EDUCATION IN LATIN AMERICA AND THE CARIBBEAN: A regional analysis of ICT integration and e-readiness, Technical Report 22923, UNESCO Institute for Statistics, Montreal, Canada. 1

Vigdor, J. L., Ladd, H. F. and Martinez, E. (2014), 'Scaling the Digital Divide: Home Computer Technology and Student Achievement', *Economic Inquiry* **52**(3), 1103–1119. 1

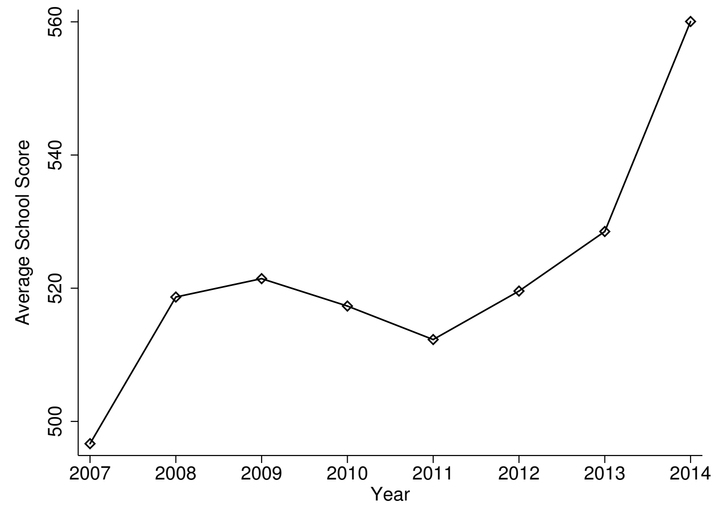
World Bank (2012), Public Expenditure Review for Peru:Spending for Results. Report No. 62586 - PE. 5

World Bank (2018), World Development Report 2018: Learning to Realize Education's Promise, Technical report, The World Bank, Washington, DC. 1

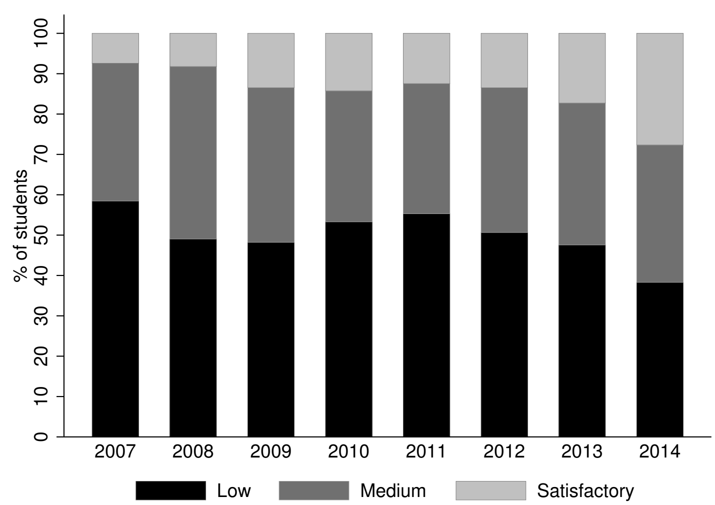
Appendix Figures and Tables

Figure A.1: Performance of Grade 2 Students in Public Schools on the ECE (2007-2014)

(a) Average Test Scores in Math

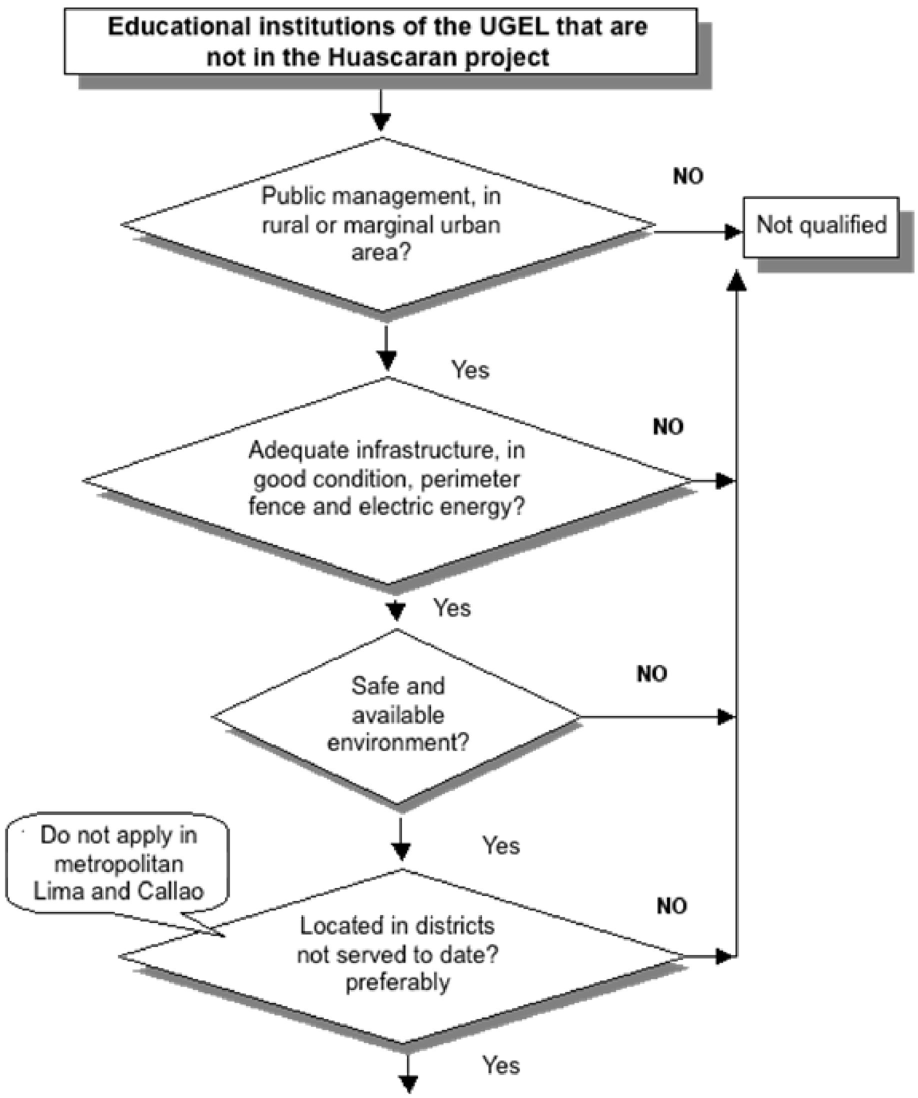


(b) Percentage of Students with Low, Medium, and Satisfactory Performance in Public Schools



Source: Authors' calculations based on the Evaluación Censal de Estudiantes, 2007-2014

Figure A.2: Translated Excerpt of Flow Chart for Prioritization under *Plan Huascarán*



Authors' translation. Original document in Spanish can be found here: http://www.minedu.gob.pe/normatividad/directivas/Dir083VMGP2003.php. <Accessed October 4, 2017>

Figure A.3: Translated Excerpt of Prioritization Directive for *Plan Huascarán*

VI. Specific rules

6.1 The prioritization criteria are as follows:

a. Educational institutions of public management, preferably located in areas with higher poverty, rural or marginal urban areas.
b. Primary, secondary and integrated educational institutions.
c. Establishment of a quota of educational institutions by UGEL (See Annex No. 1).
d. For each level of education (only primary, secondary only and both levels) a percentage will be applied that will allow to set the maximum number of educational institutions for each level in the UGEL and that in total will be equal to the quota set for each UGEL. The percentages are as follows:
    - Primary: 50%
    - Secondary: 30%
    - Integrated: 20%
e. It will benefit educational institutions of larger school populations.
f. Infrastructure of the school premises in good condition and with electricity service.
g. Available environment for the Innovation Classroom and with appropriate security measures to prevent theft.
h. Educational institutions located in districts not served to date, preferably.
i. The Ministry of Education, through the Huascarán Project, will establish the criteria and procedures for attending educational entities that do not have electricity.

Authors' translation. Original and complete directive in Spanish can be found here: http://www.minedu.gob.pe/normatividad/directivas/Dir083VMGP2003.php. <Accessed October 4, 2017>

Figure A.4: Translated School Data Sheet for *Plan Huascarán*

ANNEX N ° 2

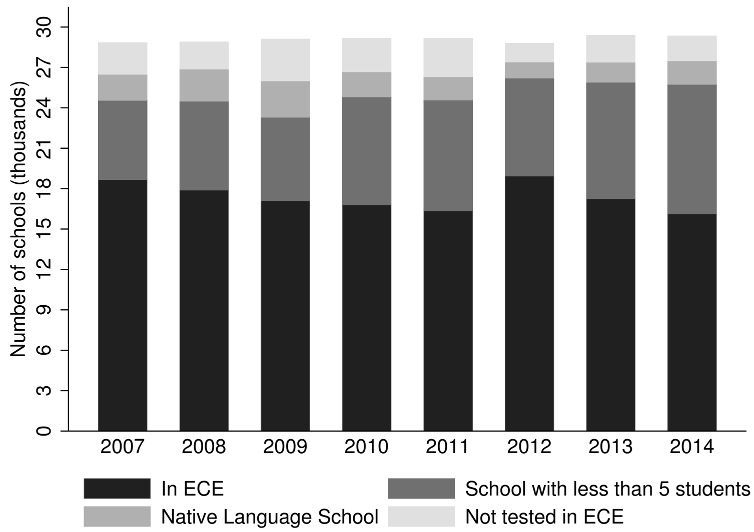
DATA SHEET OF THE EDUCATIONAL INSTITUTION		
Name of Educational Institution		
School Site Code		
Address		
Department		
Province		
District		
Town Center		
Phone		
Principal's name		
Direct intermediate organ		
Geographical area (urban, rural)		
Type of Management (State, Parish, Cooperative, Supervised, etc.)		
Number of computers for school use (only Pentium I or more)		
Number of computers for administrative use (only Pentium I or more)		
Do you have electricity?		
Number of hours of electricity		
Number of students and teachers per level	Students	Teachers
Initial		
Primary		
High school		
Number of students and teachers per shift	Sections	
Morning		
Late		
Night		
Number of sections per shift	Sections	
Morning		
Late		
Night		
Number of sections per level	Students	Teachers
Initial		
Primary		
High school		
Is there home-based telephone in the locality?		
Native language of students		
Distance to the nearest Huascarán Program educational institution		

Source: Authors' translation of Annex No. 2 available in Spanish here:
http://www.minedu.gob.pe/normatividad/directivas/Dir083VMGP2003.php

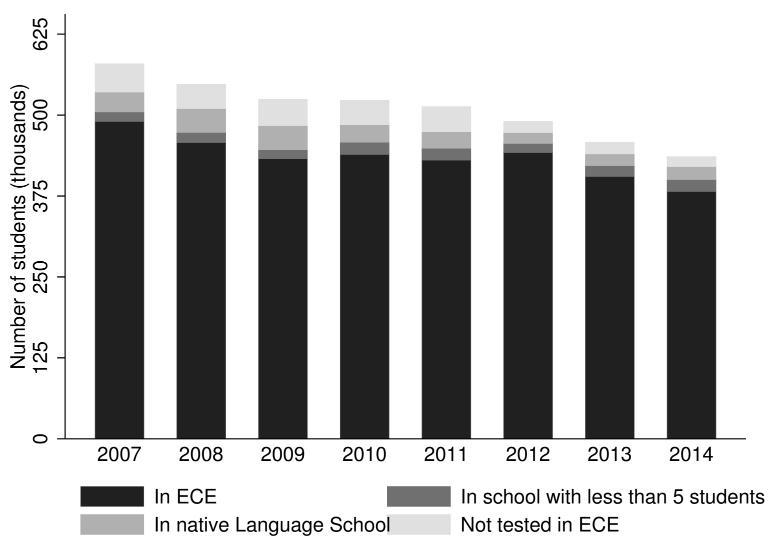
Authors' translation. Original document in Spanish can be found here: http://www.minedu.gob.pe/normatividad/directivas/Dir083VMGP2003.php. <Accessed October 4, 2017>

Figure A.5: Schools and Students in ECE, 2007-2014

(a) Primary Schools tested in ECE

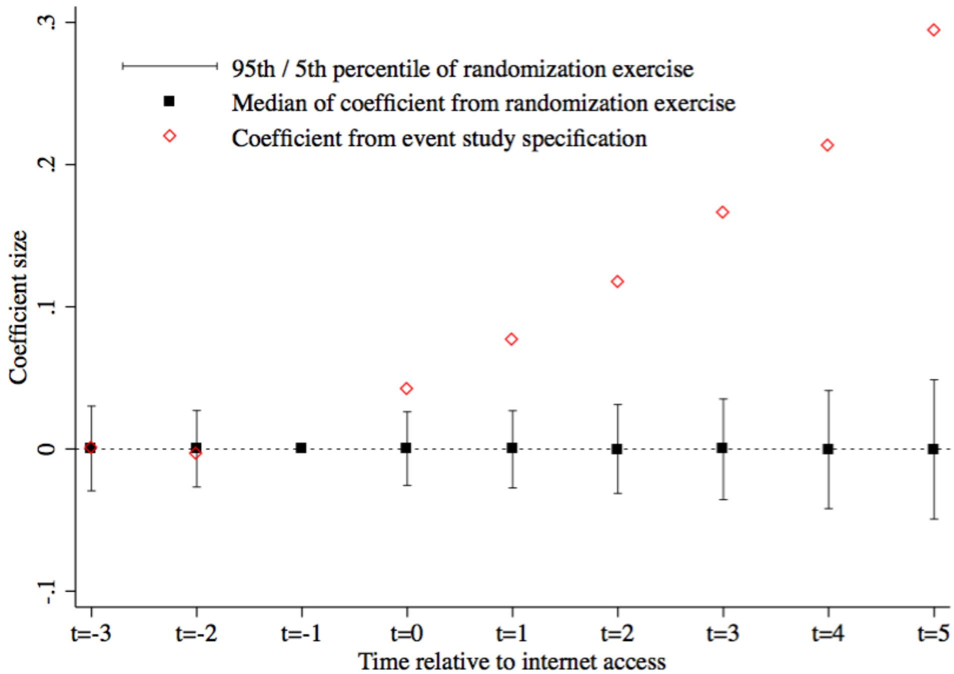


(b) Primary Students tested in ECE



Source: Authors' calculations based on the Peruvian *Censo Escolar* (CE) and *Evaluacion Censsal de Estudiantes* (ECE), 2007-2014.
* Note: Some schools both have fewer than five second graders and teach primarily in native languages. For simplicity, the graph includes these under "Fewer than five students."

Figure A.6: Randomization Exercise



This figure plots the mean as well as the 5th and 95th percentile of the coefficients from the following placebo test: We randomly reassign internet installation years to schools in our sample, maintaining the actual distribution of installation years and thus ensuring that the sample sizes of each of our "treatment cohorts" matches our baseline specification. We then use the randomly assigned installation years to generate the false event study variables and estimate equation 1. We repeat this process 10,000 times. For reference, we also plot the estimates from our baseline specification (equation 1). Control variables include total school enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). The sample includes all public schools with at least two observations within the sample period (2007-2014). Standard errors are clustered by school.

Table A.1: Predictors of Internet Access

	Dependent Variable: School Has Gained Internet Access			
	Including Characteristics Prioritized by *Plan Huascarán*	Adding School FE	Adding a Control for for Fencing (2010 and Later)	Adding Characteristics in School Data Sheets
	(1)	(2)	(3)	(4)
School Has a Computer Room	0.092***	0.033***	0.025***	0.034***
	(0.006)	(0.005)	(0.006)	(0.005)
School Has a Library	0.039***	0.014***	0.019***	0.015***
	(0.005)	(0.005)	(0.007)	(0.005)
School Has Administrative Office(s)	0.014**	0.014***	0.008	0.013***
	(0.006)	(0.005)	(0.006)	(0.005)
Total Enrollment (in 100s of students)	0.033***	0.028***	0.030***	0.027***
	(0.002)	(0.007)	(0.011)	(0.007)
School Is in an Urban Area	0.008			
	(0.010)			
School Has a Full Perimeter Fence			0.014	
			(0.009)	
Number of Computers for Instruction				-0.001***
				(0.000)
Number of Computers for Administration				0.000
				(0.000)
Total Number of Teachers				-0.003***
				(0.001)
Observations	31,368	31,368	22,669	31,368
Number of Schools	5,903	5,903	5,766	5,903
UGEL and District FE	Yes	No	No	No
Year Fixed Effects	Yes	Yes	Yes	Yes
School Fixed Effects	No	Yes	Yes	Yes

The sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Standard errors are clustered at the school level. Prioritized characteristics include enrollment and facilities (computer room, electricity, library, administrative offices), Local Educational Management Unit (UGEL) fixed effects (to capture UGEL-level quotas), district fixed effects (to capture poverty status) and year fixed effects. Information on the existence of a perimeter fence is only available for 2010 and later. Additional characteristics on school data sheets include number of computers used for instruction, number of computers used for administrative purposes, and number of teachers. Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.2: Summary Statistics (2007 or Earliest Available Year) of Public Primary Schools

	A. All Public Schools			B. Public Schools in ECE 2007-2014			C. Estimation Sample
	Internet before 2007	Internet 2007-2014	No Internet by 2014	Internet before 2007	Internet 2007-2014	No Internet by 2014	Internet 2007-2014, Matched to ECE, Observed at least Twice
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
School Average Standardized				0.23	-0.04	-0.27	-0.02
Math Score				(0.62)	(0.91)	(1.16)	(0.89)
School Is in an Urban Area	0.93	0.46	0.10	0.94	0.49	0.11	0.53
	(0.25)	(0.50)	(0.30)	(0.24)	(0.50)	(0.32)	(0.50)
One Teacher per Grade	0.94	0.50	0.11	0.94	0.54	0.13	0.58
	(0.24)	(0.50)	(0.31)	(0.24)	(0.50)	(0.34)	(0.49)
Total Enrollment	551.30	186.54	55.78	553.69	199.56	62.62	214.4
(Grades 1-6)	(348.57)	(229.59)	(67.32)	(347.83)	(233.92)	(69.83)	(239.18)
Enrollment in 2nd Grade	89.32	32.37	10.90	89.73	34.62	12.23	37.08
	(58.75)	(39.11)	(12.30)	(58.62)	(39.82)	(12.62)	(40.77)
School Is Connected to	0.87	0.55	0.30	0.87	0.57	0.32	0.59
Public Water Network	(0.34)	(0.50)	(0.46)	(0.34)	(0.50)	(0.47)	(0.49)
School has Library	0.70	0.37	0.25	0.70	0.37	0.25	0.38
	(0.46)	(0.48)	(0.43)	(0.46)	(0.48)	(0.44)	(0.49)
School Has Administrative	0.80	0.45	0.25	0.80	0.47	0.27	0.48
Office(s)	(0.40)	(0.50)	(0.43)	(0.40)	(0.50)	(0.44)	(0.50)
Number of Teachers	6.12	5.56	5.22	6.14	5.57	5.27	5.63
	(7.87)	(7.89)	(7.85)	(7.88)	(7.95)	(7.78)	(8.08)
Number of Classrooms	20.47	7.76	3.08	20.53	8.18	3.36	8.67
	(11.62)	(6.84)	(2.77)	(11.60)	(6.83)	(2.92)	(6.89)
Computers in School	17.89	3.07	0.45	17.92	3.20	0.49	3.42
	(19.58)	(7.82)	(2.55)	(19.57)	(7.26)	(2.59)	(7.30)
Second Grade Textbooks	344.37	122.89	39.80	345.76	131.22	44.30	140.4
	(279.80)	(176.53)	(52.61)	(279.61)	(181.14)	(54.97)	(186.85)
Number of Schools	1366	7083	21833	1359	6527	17738	5903

Panel A includes all public schools. Panel B restricts the analysis to schools that participated in the ECE at least once between 2007 and 2014. Panel C further restricts the sample to schools that appear at least twice (conditional on ECE participation). We present data for 2007 (or the earliest available year with information in the sample, when data for 2007 is not available). Average school test scores have been standardized to mean zero and standard deviation one across all tested schools within each year. In the Peruvian school system, schools might be *unidocente* (only one teacher in the school teaches all grades), *multigrado* (more than one teacher, but each might teach more than one grade in the same classroom), or *polidocente completo* (there is one teacher per grade in the school). Standard deviations in parentheses.

Table A.3: Effect of Internet Access on Standardized Test Scores

	School Average Standardized Math Score (1)
t = -3	0.001
	(0.020)
t = -2	-0.003
	(0.017)
t = 0	0.042**
	(0.017)
t = 1	0.076***
	(0.018)
t = 2	0.117***
	(0.021)
t = 3	0.166***
	(0.024)
t = 4	0.213***
	(0.028)
t = 5	0.294***
	(0.034)
Joint Test of Significance for All $t < 0$:	0.968
Joint Test of Significance for All $t \geq 0$:	0.000
Year Fixed Effects	Yes
School Fixed Effects	Yes
Time-varying controls	
Observations	31,368
Number of Schools	5,903

The sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Standard errors are clustered by school. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers).
Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.

[htbp]

Table A.4: Summary Statistics for Alternate Estimation Samples

	Baseline Sample (1)	2 Pre & 2 Post Sample (2)	ECE Scores in All Years Sample (3)	Complete Event Window Sample (4)
School Average Standardized Math Score	0.069	0.076	0.153	0.110
	(0.924)	(0.917)	(0.856)	(0.878)
Years of internet access	1.119	0.851	1.187	0.857
	(1.448)	(1.227)	(1.492)	(1.125)
Number of Students Scheduled for the Test	35.689	33.475	43.949	38.410
	(38.955)	(34.180)	(41.640)	(34.447)
School has library	0.446	0.433	0.487	0.460
	(0.497)	(0.496)	(0.500)	(0.498)
School has administrative office(s)	0.480	0.458	0.518	0.540
	(0.500)	(0.498)	(0.500)	(0.498)
Ratio of Classrooms to Students	0.076	0.075	0.068	0.068
	(0.076)	(0.079)	(0.077)	(0.047)
Ratio of Computers to Students	0.192	0.189	0.156	0.148
	(0.329)	(0.319)	(0.267)	(0.252)
Ratio of Teachers to Students	0.057	0.056	0.052	0.053
	(0.040)	(0.033)	(0.027)	(0.029)
Total School Enrollment	209.530	195.701	258.249	224.555
	(227.649)	(197.098)	(241.775)	(199.243)
Observations	31,368	22,321	21,652	7,322
Number of Schools	5,903	3,670	3,565	1,046

Column 1: The baseline sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Column 2: The "2 Pre & 2 Post" sample includes schools observed at least twice prior to and twice after internet access. Column 3: The non-attritor sample are schools that are observed for each year 2007-2014. Column 4: The complete event window sample includes schools that are observed throughout the entire sample period and restricted event window $t = -3$ to $t = 3$. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year.

Table A.5: Understanding the Role of School-level Compositional Changes: Trend Break Results

	Dependent Variable: School Average Standardized Math Score				
	Full Event Window: $t = -3$ to $t = 5$			Restricted Event Window: $t = -3$ to $t = 3$	
	Baseline Sample	2 Pre & 2 Post Sample	ECE Scores in All Years Sample	Baseline Sample	Complete Event Window Sample
	(1)	(2)	(3)	(4)	(5)
Post-internet Access	0.036*	0.031	0.021	0.041**	-0.005
	(0.020)	(0.022)	(0.022)	(0.021)	(0.053)
Post-internet Access X Event Time	0.047***	0.043***	0.042***	0.044***	0.061
	(0.011)	(0.013)	(0.011)	(0.012)	(0.042)
Event Time	-0.001	0.007	0.018*	-0.002	0.011
	(0.010)	(0.011)	(0.010)	(0.010)	(0.027)
Observations	31,368	22,321	21,652	28,392	7,322
R-squared	0.013	0.015	0.023	0.009	0.020
Number of Schools	5,903	3,670	3,565	5,836	1,046
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes	Yes

Column 1 reproduces the baseline results using all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Column 2 presents the results using the restricted sample of schools observed at least twice prior to and twice after internet access. Column 3 presents the results using the sample are schools that are observed for each year 2007-2014 (i.e. non-attritors). Column 4 presents the results using the baseline sample for the restricted event window $t = -3$ to $t = 3$. Column 5 presents the results using the sample of schools that are observed throughout the entire sample period and restricted event window. Standard errors are clustered by school. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Post-internet access is a dummy variable for whether a school has gained internet access (i.e. $t >= 0$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Event time is years relative to internet access. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers). Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.6: Attrition (Number of Observations)

	N	%
All Information Observed (Not Attrited)	31,368	81.78
Attrited	6,989	18.22
Missing ECE Scores		
Second Grade Enrollment less than 5	2,340	6.10
Second Grade Enrollment between 5 and 7	1,022	2.66
Other	1,915	4.99
Missing Census (CE) Information		
Missing Information on Infrastructure	1,467	3.82
Missing Information on Resources	218	0.57
School Permanently Closed	27	0.07

Based on each school's initial year of internet connection and our event study window ($t = -3$ to $t = 5$), we determine all the periods that should be included in our panel dataset. Enrollment is measured from the CE (reported at the beginning of each year). Only schools with five or more second graders by the end of each year are tested in the ECE. Schools that have 5-7 students at the beginning of the year might not have been included in the ECE if they fell below the 5-student threshold by the end of the year. CE information is used to calculate infrastructure (and, importantly, internet access) and school resources (control variables in our regressions).

Table A.7: Effects of Teachers and Computers in Non-internet Schools

	School Average Standardized Math Score					
	(1)	(2)	(3)	(4)	(5)	(6)
Teachers per Student	0.026	0.779***	0.912***	1.059***	0.954***	0.872**
	(0.055)	(0.279)	(0.296)	(0.328)	(0.361)	(0.445)
Teachers per Student (1 Year Lag)		0.022	0.206	0.297	0.201	0.193
		(0.051)	(0.316)	(0.330)	(0.381)	(0.459)
Teachers per Student (2 Year Lag)			-0.004	-0.159	-0.437	-0.573
			(0.058)	(0.352)	(0.396)	(0.492)
Teachers per Student (3 Year Lag)				0.059	-0.558	-0.265
				(0.072)	(0.469)	(0.574)
Teachers per Student (4 Year Lag)					0.062	0.736
					(0.065)	(0.657)
Teachers per Student (5 Year Lag)						0.050
						(0.059)
Observations	81,415	69,135	57,617	47,364	38,666	29,411
Number of schools	17,557	16,498	15,536	14,215	13,448	12,645
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes
	School Average Standardized Math Score					
	(6)	(7)	(8)	(9)	(10)	
Computers per Student	0.031**	0.030*	0.026	0.019	0.016	0.024
	(0.015)	(0.018)	(0.019)	(0.022)	(0.027)	(0.037)
Computers per Student (1 Year Lag)		0.022	0.015	0.011	0.023	0.006
		(0.018)	(0.020)	(0.021)	(0.026)	(0.037)
Computers per Student (2 Year Lag)			0.031	0.037**	0.028	0.001
			(0.023)	(0.019)	(0.021)	(0.032)
Computers per Student (3 Year Lag)				-0.021	-0.034	-0.017
				(0.023)	(0.024)	(0.042)
Computers per Student (4 Year Lag)					-0.036	-0.051
					(0.036)	(0.060)
Computers per Student (5 Year Lag)						-0.049
						(0.056)
Observations	80,288	65,596	51,450	39,351	29,727	21,852
Number of Schools	17,464	16,198	14,903	12,732	10,509	9,516
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes

The sample includes all schools that report ECE scores during 2007-2014 and that had not gained internet access by 2014. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Control variables in all columns include school facilities (piped water, library, administrative offices) and resources per student (computers, classrooms, and teachers) except when specified as an outcome variable. Standard errors are clustered by school. Significance levels denoted by: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Table A.8: Effects of Concurrent Resource Investments: Back of the Envelope Calculations

Time Relative to Internet Access (1)	Total Predicted Rise in Teachers per Student (2)	Total Dynamic Effect of Teachers on Scores (3)	Estimated Effect of Internet Access on Scores (4)	Share of Total Internet Effect Explained by Teachers (%) (5)
$t = 0$	0.003	0.003	0.042	6.58%
$t = 1$	0.004	0.005	0.076	7.15%
$t = 2$	0.006	0.008	0.117	6.72%
$t = 3$	0.008	0.010	0.166	6.29%
$t = 4$	0.010	0.015	0.213	6.99%
$t = 5$	0.012	0.019	0.294	6.42%
Time Relative to Internet Access (6)	**Total Predicted Rise in Computers per Student (7)**	**Total Dynamic Effect of Computers on Scores (8)**	**Estimated Effect of Internet Access on Scores (9)**	**Share of Total Internet Effect Explained by Computers (%) (10)**
t=0	0.045	0.001	0.042	3.35%
t=1	0.045	0.002	0.076	3.22%
t=2	0.045	0.004	0.117	3.53%
t=3	0.045	0.003	0.166	2.02%
t=4	0.045	0.002	0.213	0.81%
t=5	0.045	0.000	0.294	-0.17%

Column 2 (7) gives the total predicted rise in teachers (computers) in each period using the parameter estimates from the trend break regression of resources per student on Post-internet access, event time, and the interaction between the two (from Table 3). Column 3 (8) calculates the total dynamic effect of teachers (computers) as of time t using (i) the largest parameter values from regressing ECE scores on teachers (computers) per students and lags using the sample of non-internet connected schools (from Appendix Table A.7), without regard to significance level and (ii) the total predicted rise in teachers (computers) from column 2 (7). Columns 4 and 9 display estimated effects of internet access on test scores from the baseline event study specification (Appendix Table A.3). Column 5 (10) expresses the total effect of teachers (computers) as a percent of the total effect of internet access (column 3 (8) divided by column 4 (9)).

Table A.9: Internet Access and Computer Teachers: Trend Break Results

	Dependent Variable: Presence of a Teacher with Formal Computer Training
Post-internet Access	0.027***
	(0.007)
Post-internet Access X Event Time	0.015***
	(0.004)
Event Time	0.000
	(0.003)
Observations	31,368
Number of Schools	5,903
R-squared	0.035
Year Fixed Effects	Yes
School Fixed Effects	Yes
Time-varying controls	Yes

The sample includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Standard errors are clustered by school. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Post-internet access is a dummy variable for whether a school has gained internet access (i.e. $t >= 0$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Event time is years relative to internet access. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers).
Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.10: Heterogeneity in the Impact of Internet Access across Schools by Student to Teacher Ratios and Teacher Qualifications: Trend Break Results

	Dependent Variable: School Average Standardized Math Score			
	High STR (1)	Low STR (2)	High Teacher Qualifications (3)	Low Teacher Qualifications (4)
Post-internet Access	0.025	0.050	0.036	0.043
	(0.027)	(0.031)	(0.029)	(0.030)
Post-internet Access	0.046***	0.009	0.063***	0.006
× Event Time	(0.015)	(0.020)	(0.016)	(0.018)
Event Time	0.012	0.004	-0.002	0.006
	(0.013)	(0.015)	(0.014)	(0.014)
	p-value for test that High STR = Low STR		p-value for test that High Qual. = Low Qual.	
...for Post-internet Access	0.552		0.866	
...for Post-internet Access × Event Time	0.139		0.022	
Observations	14,307	12,834	13,926	13,215
Number of Schools	2,458	2,457	2,458	2,457
Year Fixed Effects	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes

High and Low STR schools are defined based on each school's pre-internet average ratio of grade 2 students to total teachers in the school (STR) over the sample period relative to the median of all schools' sample averages. High and Low Teacher Quality schools are defined based on each school's pre-internet average number of teachers with a pedagogical or higher education degree per student over the sample period relative to the median of all schools' sample averages. Note that degree information is not available for all schools. Samples includes all public schools that gained internet access between 2007 and 2014 and are observed at least twice. Standard errors are clustered by school. Math scores are standardized to have mean zero and standard deviation of 1 across the universe of schools reporting scores within each calendar year. Post-internet access is a dummy variable for whether a school has gained internet access (i.e. $t >= 0$). Note that due to the timing of the *Censo Escolar* relative to the ECE exam, some schools receive internet access in $t = 0$ while some receive it in $t = 1$. For more details, see Section 3.1. Event time is years relative to internet access. Control variables include enrollment, number of second grade students scheduled to take the ECE, facilities (computer room, library, administrative offices), and resources per student (classrooms, computers, and teachers).
Significance levels denoted by: *** p< 0.01, ** p< 0.05, * p< 0.1.