Stanford cepa | Center for Education Policy Analysis

CEPA Working Paper No. 17-08

# Effectiveness of Interactive Satellite-Transmitted Instruction: Experimental Evidence from Ghanaian Primary Schools

AUTHORS

**Jamie Johnston**
Stanford University

**Christopher Ksoll**
Mathematica Policy Research

ABSTRACT

In lower- and middle-income countries, including Ghana, students in rural areas dramatically underperform their urban peers. Rural schools struggle to attract and retain professionally trained teachers (GES 2012; World Bank 2012). We explore one potential solution to the problem of teacher recruitment: distance instruction. Through a cluster randomized controlled trial, we estimate the impact of a program that broadcasts live instruction via satellite to rural primary school students. The program equipped classrooms in 70 randomly selected Ghanaian schools with the technology required to connect to a studio in Accra. An additional 77 schools served as the control. Instructors in Accra provided math and English lessons to classrooms in the treatment group. The model is interactive, and students in satellite classes could communicate in real time with their remote teachers. We estimate significant gains ($p<.05$) in rural students' numeracy and foundational literacy skills. We find no impact on attendance and classroom time-on-task (as measured through unannounced classroom observations), suggesting that these gains may result from improved instructional quality rather than from increased instruction time.

VERSION

August 2017

**Suggested citation:** Johnston, J., & Ksoll, C. (2017). Effectiveness of Interactive Satellite-Transmitted Instruction: Experimental Evidence from Ghanaian Primary Schools (CEPA Working Paper No.17-08). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp17-08

# Effectiveness of Interactive Satellite-Transmitted Instruction: Experimental Evidence from Ghanaian Primary Schools

*Jamie Johnston*[1] *and Christopher Ksoll*[2]

*Working Paper*

*August 2017*

*Abstract:* In lower- and middle-income countries, including Ghana, students in rural areas dramatically underperform their urban peers. Rural schools struggle to attract and retain professionally trained teachers (GES 2012; World Bank 2012). We explore one potential solution to the problem of teacher recruitment: distance instruction. Through a cluster randomized controlled trial, we estimate the impact of a program that broadcasts live instruction via satellite to rural primary school students. The program equipped classrooms in 70 randomly selected Ghanaian schools with the technology required to connect to a studio in Accra. An additional 77 schools served as the control. Instructors in Accra provided math and English lessons to classrooms in the treatment group. The model is interactive, and students in satellite classes could communicate in real time with their remote teachers. We estimate significant gains (p<.05) in rural students' numeracy and foundational literacy skills. We find no impact on attendance and classroom time-on-task (as measured through unannounced classroom observations), suggesting that these gains may result from improved instructional quality rather than from increased instruction time.

## 1 Introduction

Developing countries suffer from sizeable disparities in student achievement between rural and urban areas (OECD, 2013; Orazem and Kin, 2007). Although differences in students' socioeconomic backgrounds partially explain the gap, education in rural areas is also threatened by disparities in educational resources. Rural schools have far fewer instructional materials and face chronic shortages in qualified teaching staff (OECD, 2013; Fagernas and Pelkonen, 2012; Glewwe and Kremer, 2006; Duthilleul, 2005).

The limited supply of high-quality teachers may be particularly damaging to rural students. Evidence strongly suggests that teaching quality dramatically impacts learning outcomes (Sanders and Rivers, 1996; Rockoff, 2004; Rivkin, et al., 2005; Boyd et al., 2008,

[1] Stanford University, Postdoctoral Fellow, jamie.johnston@stanford.edu
[2] Mathematica Policy Research, Researcher, CKsoll@mathematica-mpr.com

Clotfelter et al., 2007). In low- and middle-income countries, teacher capacity and content knowledge have been shown to be key determinants of students' learning (Glewwe et al., 2011; McEwan, 2015).

Dozens of experimental studies suggest that, among school-level interventions, programs that target improving quality of instruction have the greatest and most consistent effect on learning outcomes (Ganimian and Murnane, 2016; Snilstsveit et al., 2015; McEwan, 2015; Kremer et al., 2013). Indeed, pedagogical reform is not only cost-effective but also necessary for gains in learning. As highlighted by Kremer et al. (2013), programs that provide material resources are largely ineffective unless they simultaneously work to improve pedagogy.

Using a cluster-randomized design, we examine whether remote instruction in rural Ghanaian primary schools improves student outcomes. The two-year pilot program provided solar power and satellite technology to 70 randomly selected primary schools and broadcast daily mathematics and English lessons, transmitted from the capital city of Accra. Professionally trained and vetted teachers taught the broadcasted lessons in real time using video conferencing equipment that allowed them to communicate directly with remote classrooms. In-person teachers, who were largely drawn from the rural schools' pre-existing workforce, managed the classrooms. These in-person teachers were responsible for setting up the equipment, assigning grades, overseeing classroom management and discipline, and providing instruction when technology failed.

Over two years of implementation, the program measurably improved students' numeracy and literacy skills. Satellite classes improved students' math scores by 0.23 standard deviations. The learning gains in math appear to be sustained; a cohort of students assessed one year after the program ended maintained the overall improvement observed immediately after their participation in the program. The satellite classes did not lead to gains in overall reading fluency in English; however, the classes did improve students' foundational skills, specifically with letter and word identification.

Classroom observations and enrollment data suggest that the satellite classes impacted neither classroom attendance nor time-on-task. Because overall instructional time (as measured through attendance and time-on-task) remained constant across control and treatment groups, we conjecture that the learning gains resulted from improved instructional quality rather than from increased instruction time. Classroom observations also showed that in-person teachers spent more than half of the remote instruction time engaged in instruction or classroom management. Thus, learning gains might also have resulted in part from the increased instructor presence–distance instructor and in-person teacher–that essentially increased the classrooms' teacher-student ratio.

Although distance instruction has gained traction in many parts of the world, this particular model of live remote instruction is distinct from most forms of distance learning because of its synchronous and interactive nature.[3] At present, little is known about the effectiveness of this model. This program is not only the first to be implemented in Sub-Saharan Africa, but also, to our knowledge, the first of its kind worldwide to be rigorously through a randomized study.

Before the Internet rose to prominence, rural K-12 schools in developed countries used satellite and cable television to supplement their core curriculum with pre-recorded video instruction (Barbour, 2007; Barker and Hall, 1994). Today, these schools frequently use

3 Synchronous distance learning generally refers to instruction in which instructor and learners interact in real time using tools like webcasting, messaging chat rooms, or audio / video conferencing tools. Asynchronous learning refers to instruction that is separated in time such that instructors and learners are not communicating in real time but rather through technologies like email or online discussion boards (Means et al., 2009; Wicks, 2010; Hrastini, 2008).

online courses to expand their curricular offerings (Heppen et al., 2011; Picciano and Seaman 2009; Tucker 2007; Schwartzbeck et al., 2003). In developing countries, remote instruction has long reached remote communities through radio and television broadcasts (Jumani, 2009; Lucas 1999), and since the 1970s over two-dozen developing countries have employed interactive radio instruction (World Bank, 2005; Ho and Thukral, 2009). India first pioneered using satellite technology to broadcast primary- and secondary-level video instruction (Desai, 2009; Bhattacharya, 2008; De, 2004), and the developers of the Ghanaian pilot based their model on the system of two-way interactive videoconferencing used in India.

The videoconferencing model pioneered in India is compelling because it allows students to communicate directly and in real time with a live instructor, whereas other online interventions typically use static instructional aids. While there is evidence that pre-recorded videos (Naslund-Hadley et al., 2014) and computer-assisted learning programs (Banerjee et al, 2007; He, et al., 2007) can enhance learning, the benefits are not found across all such programs (Barrera-Osorio and Linden, 2009; Snilstveit et al., 2015). Similarly, findings are mixed with regard to the effectiveness of distance and online instruction as compared to traditional in-person instruction (Phipps and Merisotis, 1999; Means et al., 2010). The handful of causally rigorous studies of asynchronous online instruction (i.e., instruction that has been prerecorded or preprogrammed) find that such instruction has negative or null effects on learning outcomes (Bettinger et al., forthcoming; Heppen et al., 2017; Hart et al., 2016; Figlio, et al., 2013; Xu and Jaggers, 2013; Cavulluzo et al., 2012).

There is little published evidence that speaks to the effectiveness of two-way videoconferencing instruction. The majority of distance models that have been studied are asynchronous. In those studies in which the distance programs under scrutiny *are* synchronous, students interact with instructors primarily via a text-based interface (e.g., via an online discussion forum or messaging applications) rather than through video exchange. This distinction matters because the interactivity between instructors and students has been shown to influence the effectiveness of distance learning. Two studies find that the availability of avenues for student-teacher interaction correlates with both student and instructor satisfaction (Swan, 2001; Picciano, 2002), and early developers of distance-learning models recognize that interactive components that encourage active learning positively impact student engagement (Cavanaugh, 1999; World Bank, 2005).

While the Ghanaian videoconferencing program's primary objective is to improve instructional quality, the videoconferencing has the added benefit of virtually doubling the instructor presence in the classroom, essentially increasing the teacher-student ratio. While class-size reductions alone have been shown to be ineffective at improving learning (Rockoff, 2009), the simultaneous presence of distance and in-person instructors might work in other ways to improve the classroom experience. First, the videoconferencing component could serve as a mechanism of accountability, mitigating in-person teacher absenteeism. Duflo, Dupas and Kremer (2015) find that accountability and incentive structures faced by instructors are key to improving learning. Second, the distance-learning classroom benefits from a productive division of labor; the distance and in-person teachers take on complementary roles rather than functioning as substitutes for or duplicates of one another. The in-person teacher can take advantage of distance-instruction time to attend to classroom management, to complete organizational tasks, and (perhaps most importantly) to work with students one-on-one or in small groups. A growing body of literature emphasizes the importance of teaching students at their level, and thus the availability of an instructor to work with struggling students personally may be particularly key to gains in learning outcomes (Banerjee et al., 2016; Duflo and Kiessel, 2014; Duflo et al., 2011).

The remainder of the paper is organized as follows. Section 2 provides background information on the research setting and design. Section 3 describes sources of data, sample characteristics, and estimation strategy. Section 4 describes main results. Section 5 discusses potential mechanisms. Section 6 concludes.

## 2 Research Setting and Design

### *2.1 Study Location and Population*

Although access to formal education has improved in rural areas of Ghana, students in remote areas continue to underperform their urban counterparts, and rural schools struggle to attract and retain qualified teachers (MOE, 2016a; MOE, 2011; World Bank, 2012; Ajayi, 2011). The dearth of trained teachers is attributable to a number of factors including poor sanitation, inadequate transportation, and lacking local amenities (Darvas and Balwanz, 2013). In 2012, the World Bank reported that over half of primary teachers in Ghana have no professional training at all; the study placed the pupil to trained teacher ratio at roughly 45:1. Nation-wide, primary-school race high rates of teacher absenteeism, and teachers are absent for an average of 27 percent of school days, and the rate of absenteeism is even higher in rural and poor regions. The marked lack of instructional time resulting from teacher absenteeism is worsened by the prevalence of prescriptive, rote-learning teaching methods; between failures in pedagogy and failures in attendance, it is unsurprising that students spend minimal time on task (World Bank, 2012; MOE 2016b).

The pilot program we examine was implemented in six districts of the Volta and Greater Accra regions. The Ghanaian government listed each of these districts as “deprived,” indicating they are among the most under-resourced in the nation, and our choice to include them in the videoconferencing pilot was approved by the Ghana Education Service (GES). The main economies in the treatment districts are farming and fishing. Along with GES and the organization that implemented the pilot, program implementers identified 147 public primary schools eligible for inclusion in the program. They selected these sites because they were located both in GES categorized deprived districts and within 30 km of a satellite receiver.

### *2.2 Satellite Class Intervention*

Using funding from the Department for International Development (DfID)'s Girls' Education Challenge, the UK-based Varkey Foundation designed and implemented the satellite classes through the Making Ghanaian Girls Great! (MGCubed) program.[4] The initiative equipped 140 classrooms in 70 randomly selected schools with a technology package consisting of solar panels, a satellite modem, a projector, a web cam, microphones, and a computer with interactive software.[5] The MGCubed program also established two studios in Accra, Ghana's capital city, where math and English lessons were broadcast live each day to the network of classrooms. The project recruited and trained seven studio teachers in distance learning and student-centered teaching techniques.[6] Studio teachers received continuous coaching to refine their performance. The lessons they offered were mapped to the Ghanaian primary school curriculum.

At each school, roughly 40 boys and 40 girls enrolled in grades 2-5 and ranging from 6 to 15 years of age were chosen for inclusion in the program. The program worked to

---

[4] While the program places special emphasis on girls through a girls-only after-school program, the math and English satellite classes were provided to girls and boys in the same way.
[5] The program was originally implemented in 72 schools, but we excluded two schools that lacked sufficient infrastructural support at the start of the program.
[6] Studio teachers were selected through a competitive process. Applicants were required to speak one of the major languages of the region (Dangme, Ewe or Twi) and the interview process included a practice teaching session and lesson-planning exercise. Selected teachers had 2-5 years of teaching experience, and all received training prior to the start of the classes.

prioritize each school's most marginalized students and developed a set of eligibility criteria for inclusion in the program to help ensure that priority.[7]

In the first year of the program, at each school, selected students were divided into two classes of roughly 40 students each (grades 2-3 in one class and grades 4-5 in the second class). On all five days of the school week, each class received one hour of basic math and one hour of basic English via satellite. Lessons were interactive and delivered in real time; distance teachers and their students could see and communicate with one another directly via webcams and microphones. The live lessons were broadcast to an average of 12 classrooms at a time. A single studio teacher was assigned to work with each class to ensure consistency and to personalize the relationship between the remote parties. (There were occasional instances of substitute coverage for instructors when standard personal and health conflicts arose.)

In the second year of the program, the Varkey Foundation reduced the intensity of lessons and doubled the number of students in the program, such that at 80 percent of schools, all grade 2-5 students attended MGCubed classes. Two 40-student classes of grade 2-3 students met for 1 hour of English and 1 hour of Math on Mondays through Wednesdays, and two classes of grade 4-5 students met for 1 hour of English and 1 hour of Math on Thursday-Fridays.

[7] Students were eligible to participate in this program if they met any of the following marginalization criteria: (1) being above average age for their grade level; (2) having more than five siblings or half-siblings; (3) living more than 30 minutes' walking distance from the school; and (4) having a history of truancy (missing an average of at least one class per week). Students were ranked by how many marginalization criteria they met. Each criterion was weighted equally, such that students who met only three criteria, for example, were prioritized over those meeting only two criteria, regardless of which were met. Roughly 98 percent of students in grades 2-5 met at least one of the eligibility requirements, and thus the ranking system came into play only for schools serving more than 80 2nd- to 5th-grade students. About a third of the schools served just over 80 students in the relevant grade range; in those cases, all students in the 2-5 grade range were included in the satellite classes. We are still working to obtain precise inclusion and class size data from the Varkey Foundation to pinpoint the proportion of 2nd- to 5th-grade students who were included in the program at each school.

The program trained the in-person facilitators to manage classrooms, assign grades, and operate the distance-learning equipment. Facilitators also contributed considerably to instruction, particularly in instances of equipment malfunction. These individuals were largely math and English teachers who were already employed by the treatment schools.[8] Each facilitator underwent a five-day residential training in classroom management, pedagogy, and student-centered learning at the outset of the program and attended a four-day refresher training at the start of the subsequent academic year. The facilitators also received two days of training before the start of each term. School heads or their representatives in each of the treatment schools likewise received training in the operation of the distance-learning equipment to ensure that administrators could support facilitators when necessary.

The program also included daily satellite-transmitted after-school lessons for girls, some of whom had previously dropped out of school. The after-school lessons were non-academic and focused on empowerment and health. Because after-school supplementary lessons were provided at all treatment schools, we cannot disentangle the effects of the girls' after-school lessons from those of the math and English lessons; however, we do examine treatment effect differences between girls and boys. We do not include girls who attended the after-school lessons but did not attend treatment schools in our analysis, as they did not participate in math and English classes. In the second year, an after-school program modelled on the girls' program was started for boys, although it was held monthly rather than daily.

### *2.3 Experimental Design*

Of the 147 primary schools selected for study from the six targeted geographic districts, we randomly selected 70 schools to receive treatment, leaving 77 schools to serve as controls.

[8] Roughly 80 percent of facilitators taught standard classes in treatment schools in addition to their work with the distance-learning classes.

We stratified randomization at the district level. At each participating school, we followed cohorts of students who were in grades 2-4 at baseline for two full academic years. Because students progressing to grade 6 were no longer eligible to attend the satellite classes, the cohort in grade 4 at baseline received only one academic year of treatment. As illustrated in Figure 1, our original intention was to follow grades 2-5 student for three academic years. However, due to delays in implementation in the first year ("Year 0" in Figure 1), students received the intervention beginning only in the third term (May–July) of the academic year. Thus, students in grade 5 at baseline only received three months of treatment. Given the brevity of their exposure to the program, we dropped the grade 5 cohort from our analysis and focused on the cohorts that received one to two full academic years of instruction.

[Figure 1]

As noted above, roughly 80 students at each of the 147 schools participated the program; however, due to budgetary constraints, we were able to follow only around 40 students per school for the duration of the program. These students were randomly selected and are representative of the population of eligible students.

## 3 Data and Estimation Strategy

### *3.1 Data*

We use three rounds of math and English assessments to compare the outcomes of treatment and control students. We rely on student and teacher survey data and on attendance and classroom observation data to analyse potential underlying mechanisms of the program. We also draw upon qualitative data collected through one-on-one interviews and focus groups discussions to develop hypotheses about these mechanisms. As shown in Figure 1, we collected assessment and survey data prior to the randomization of schools (baseline), after the end of the first full year (midline), and at the end of the second year (endline).

Attendance and qualitative data were collected at three monitoring points over the course of the two years and once more after endline.

*A. Skills Assessments*

We used the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) to measure the students' math and English skills. The EGRA and EGMA are oral assessments developed by USAID and RTI to measure foundational skills in math and literacy and to compare those skills across countries (Reubens and Kline, 2009; RTI International, 2009). We adapted the EGRA and EGMA to the national Ghanaian syllabus for second grade. Our field team administered these tests in the three local languages widely spoken in Ghana; rural primary school students were assessed either in Dangme, Twi, or Ewe, depending on the primary language spoken in the community. The subject tests, which take roughly 35 minutes to complete, were administered one-on-one during school hours.

The EGMA numeracy assessment includes the following subtasks: number identification, quantity discrimination, missing number, and addition / subtraction and word problems. We use the total score across subtasks as our primary numeracy outcome variable. Subtasks were weighted such that number identification, quantity discrimination, and missing number subtasks are each weighted as 20 percent of the total and addition / subtraction and word problems as 40 percent of the total.[9]

The EGRA reading assessment includes the following subtasks: letter naming fluency (letter-per-minute identified), unfamiliar word-naming fluency (invented words-per-minute identified), word-naming fluency (oral vocabulary), oral reading fluency (words-per-minute read aloud), and reading comprehension. Due to budget constraints, at midline, we were only

[9] This weighting scheme was determined by the funder, who used the outcomes in a payment-by-results contractual arrangement. We decided to retain this weighting scheme to avoid any post-intervention weighting decisions.

able to administer the EGRA the oral fluency (words-per-minute) and the reading comprehension subtasks. We privilege the words-per-minute measure as it is a commonly-used measure to reflect overall literacy and was the preferred measure of the donor for the purposes of comparing across projects. At midline, we observed a floor effect in the comprehension measure and concluded the measure was too advanced for students to capture any learning in the first year. At endline we administered the full set of EGRA subtasks and examine outcomes across all subtasks.

*B. Student and Teacher Data*

Baseline student surveys captured detailed information on household demographics and socioeconomic status (including information household size, assets and parental education). Student attendance was collected through spot checks conducted by the field team during monitoring rounds and through school records provided by the schools.

To understand whether the project results in any differences in teacher attributes and behaviors, we conducted surveys with head teachers and math and English teachers in treatment and control schools in the second year of implementation. In treatment schools, surveys were administered to all facilitators of the satellite classes, as well as standard math and English teachers. In total, 507 teachers and facilitators were interviewed. The surveys included questions about classroom time and instructional approaches. We also collected measures of teacher attendance at the same time.

*C. Classroom Observations*

We conducted class observations using a modified version of the Stallings Classroom Snapshot protocol (Stallings, 1977) in grades 4-6 math and English classes in all schools. The Stallings Classroom Snapshot, first developed as the Stanford Research Institute Classroom Observation System in 1977, generates robust quantitative data on classroom activities and instructional practice.  The instrument has been used widely by the World Bank and has been shown to have a high degree of inter-rater reliability (0.80 or higher) (World Bank, 2015).

Over a 60-minute observation period, enumerators used a modified instrument (adapted to accommodate the satellite class design) to code ten “snapshots” every six minutes recording the activities and interactions of teachers, facilitators and students in the classroom. Classroom activities are grouped into four broader categories:  1) active instruction, 2) passive instruction, 3) organization and management, and 4) off-task activities (including times when no instructor is present in the classroom, when the class is engaged in non-instructional social interaction or when students are being disciplined).

*D. Qualitative Data*

During the three monitoring rounds, we carried out open-ended qualitative interviews with a randomly sampled set of students, teachers, and school administrators at fifteen randomly selected treatment schools to better understand the changes brought about by the intervention.  During focus groups and interviews, respondents were asked to share in detail their experiences as part of the program.

### *3.2 Sample Balance and Attrition*

At baseline, prior to randomization, we surveyed and administered EGRA and EGMA assessments to 5,485 grade 2-4 students.  As shown in Table 1, our sample is roughly 60 percent female with an average age of nearly 11 years.[10]  Cohorts are equally represented in the satellite classes.  Average household size at baseline was just under eight household members, and roughly 16 percent of households are headed by mothers. Just over 50 percent

[10] As discussed in Section 2.3, for budgetary reasons, we were unable to follow all of the students receiving treatment and thus randomly sampled roughly 40 students to track for the duration of the study, oversampling girls given the focus of the project on girls’ learning.

of students reported that their mothers attended some school, and just over 60 percent reported that their fathers had attended some school.

Table 1 also shows differences in pre-program characteristics at baseline by treatment and control status, as well as joint significance tests. The treatment and control groups appear well balanced on observed traits, with differences in characteristics being small and mainly insignificant. We observe no significant differences between treatment and control students on baseline EGRA and EGMA subtasks. Furthermore, the joint tests are not significant, suggesting the randomization successfully created comparable treatment and control groups. Appendix Table A3a additionally shows balance across a greater set of variables, (i.e., parent occupation and education level) that were excluded from the main analysis and joint significance tests due to the high rate of nonresponse.

[Table 1]

The overall attrition rate (including drop outs and absenteeism at the time of follow-up data collection) was 13 percent at midline and 16 percent at endline. As shown in appendix table A1, the attrition rates did not significantly differ between treatment and control groups, and as shown in appendix table A2, there is no differential attrition by baseline scores. Furthermore, balance checks on baseline characteristics in the midline and endline samples reveal no additional differences between treatment and control groups (appendix tables A3b-A3c).

### *3.3 Estimation Strategy*

Our primary estimation strategy is a difference-in-differences (DD) specification that allows us to estimate the average program effect over the two years. Specifically, we pool across survey rounds to estimate the following specification:

$$(1) \qquad Y_{ist} = \beta_0 + \beta_1\, treat_s + \beta_2\, post_t + \beta_3\, treat_s{*}post_t + \mathbf{X}'_{is} + \delta_c + \theta_d + \varepsilon_{ist}$$

where $Y_{ist}$ represents the standardized literacy or numeracy score of student $i$ in school $s$ during year $t$. $treat_s$ indicates whether a school is assigned to treatment (treat=1) or control (treat=0). $post_t$ indicates the year of the test score (taking on a value of one for midline and endline scores and zero for baseline score). $treat_s{*}post_t$ is the interaction of being assigned to treatment and the post indicator, and $\beta_3$ is the coefficient of interest, reflecting the impact of the satellite class intervention as compared to standard control classes.

We also estimate the DD specification including cohort ($\delta_c$) and randomization strata ($\theta_d$) fixed effects, as well as a vector of student-level baseline covariates ($\mathbf{X}'_{is}$), and $\varepsilon_{ist}$ is an error term that captures any unobserved student ability or idiosyncratic shocks. For all specifications, we estimate models using ordinary least squares (OLS) cluster errors at the school level. We standardized literacy and numeracy test scores based on the contemporaneous control distribution at each test administration.

The pooled DD estimation strategy in equation (1) is our preferred specification as it controls for any potential baseline differences between treatment and control schools and also allows for the inclusion of school-level or student-level fixed effects. As a robustness check on this pooled estimations strategy, we also examine impacts using simple difference and accounting for baseline outcomes (see Appendix Table A4).

Additionally, we test whether effects differ for one versus two years in the program. Specifically, we estimate the following specification:

$$(2) \qquad Y_{ist} = \beta_0 + \beta_1\, treat_s + \beta_2\, Y1_t + \beta_3\, treat_s{*}Y1_t + \beta_4\, Y2_t + \beta_5\, treat_s{*}Y2_t + \mathbf{X}'_{is} + \delta_c + \theta_d + \varepsilon_{ist}$$

where $Y1_t$ indicates having one year of the program (taking on a value of one for midline and zero for baseline and endline scores) and $Y2_t$ indicates having two years of the program (taking on a value of one for endline and zero for baseline and midline scores). $treat_s*Y1_t$ and $treat_s*Y2_t$ are the interactions with treatment and the post indicator, and thus $\beta_3$ and $\beta_5$ are the coefficients of interest, reflecting the impact of experiencing the satellite class intervention for one year and two years.

For both specifications, we excluded the endline scores of the grade 4 cohort because they aged out of the program and only received one year of the program. In other words, we estimate immediate impact only on the years that students received treatment. We are able to use the endline scores of this cohort to examine their retention of skills a year after completion of the program.

## 4 Results

### *4.1 Average Effects of the Program*

Table 2 presents the results of our main specification, equation (1). On average, we do not observe any effect on literacy, as measured by the EGRA words-per-minute oral fluency measure (panel A). We do however find that the program has a significant effect ($p<.01$) on numeracy skills (panel B). Our results suggest that the satellite program raises students' the combined total EGMA score between 0.23-0.24 standard deviations. Across specifications, the effect is robust to the inclusion of baseline covariates, district, cohort, school and student fixed effects. Results are also robust to alternative specifications (shown in appendix table A4) including a simple difference specification, a specification controlling for baseline score, alternative standardization (to the entire contemporaneous distribution) and using raw scores as the outcome.

[Table 2]

Table 3 presents the results of the one-versus-two-year specification, equation (2). For literacy, we find that the two-year effect is significantly larger than the one-year effect (p<.05, Table 3).[11] This difference may reflect that skills build upon one another and that two years of instruction is needed to achieve a greater level of fluency. The difference may also reflect adjustments to the treatment curriculum after the first year. Because no impact was observed after the first year, implementers made adjustments to English classes after the first year (most notably of which was grouping students by skill level and differentiating instruction within the class).[12]

For numeracy, we see that the program is immediately effective in the first year, with gains in total EGMA score of 0.22 standard deviations (table 3, panel B). The full two-year treatment effect is slightly higher in magnitude 0.23-0.26 standard deviations, but the treatment effects for having one year versus two years of the program are not significantly different from one another. This suggests the large gains of the program are taking place in the first year, with fewer gains seen in the second year (although without having an exact measure of depreciation of learning gains over the summer holiday, we cannot definitively say that gains in the second year are lower). The absence of additional gains may also reflect the reduction in intensity of lessons in the second year.

---

[11] We also checked that the 1-year results are also robust to omitting the Grade 4 Cohort and are confident that differences between year 1 and year 2 are not due to differences in the sample.

[12] Appendix table A5, columns (1) and (2) show value-added specifications separately for year 1 and year 2. For literacy, the year 2 impact (as measured as gains between midline and endline) of the program is 0.10 standard deviations and significant at the 1 percent level, suggesting the intervention may have been more successful in the second year; however, given that the year 2 effect cannot be separated from students having also received the 1-year treatment, we cannot definitively compare the separate first versus second year effects. The overall 2-year treatment is robust to Table 3, but differs slightly because of the difference in the analytical sample used in the preferred DD specification, which includes endline attriters who were present at midline. Removing the endline attriters from the analytical sample results in a nearly identical effect using the DD specification.

[Table 3]

An examination of impact on subtask sheds some light on the difference in one versus two year effects. As shown in Table 4, for numeracy, we see that the full two-year effect is greater than the one-year effect for the Number Identification and Missing Number subtasks. This finding could reflect differences in the curricular focus between years one and two. It could also reflect that numeracy skills build upon one another. If so, we would expect greater gains after two years for more advanced skills.

As noted previously, due to budget constraints, we did not assess students in most of the literacy subtasks at midline. After two years, we observe large and significant gains in letter naming fluency (letters-per-minute identified) of 0.82 standard deviations. We also observe significant gains in invented word-naming fluency (invented words-per-minute identified) and word-naming fluency (oral vocabulary), both of 0.17 standard deviations each. These subtask findings suggest that the program was successful in improving more basic foundational literacy skills. However, it falls short of helping students to attain overall oral reading fluency (words-per-minute read).

[Table 4]

### *4.2 Persistence of Effects*

Because we have endline scores for the grade 4 cohort of students (who only experienced the first year of the program), we can examine whether the skills gained in the first year persist a year after the program. As we did not observe any impact in the first year for words-per-minute, we cannot speak to the persistence of literacy skills; however, we do find that one year after the program, the gain for the grade 4 cohort of students in terms of math skills is roughly the same one year after the program (0.21-0.22 standard deviations) as measured immediately after the program (0.23 standard deviations). This suggests that students do not lose the skills that they learn as part of the satellite program. It also suggests there may be little value for money for a second year of mathematics classes.

[Table 5]

### *4.3 Heterogeneous Effects*

Because we might not expect the treatment effects to be the same for different subpopulations of students, we examine whether there are heterogeneous treatment effects. We whether there are differences in effects by gender particularly because the daily after-school health and empowerment lessons were open only to girls. We also examine whether there are heterogeneous impacts by performance on the baseline tests, as the satellite classes may be differentially effective for initially higher or lower performing students. We also examine differences by grade at baseline and by a measure of poverty (PPI) constructed at baseline.[13] As shown in Table 6, we find little evidence of differential effects for the subgroups examined. One exception is that the program is marginally significantly better at improving numeracy skills for students below median PPI at baseline. However, given that we assessed 14 possible heterogeneous effects and that the heterogeneous effect on EGRA for these students is of the opposite sign, we do not think this effect should not be over-interpreted is meaningful.

[Table 6]

## 5 Potential Mechanisms

We explore several mechanisms through which program changes the nature of time on task that may bring about gains in students' skills: (1) increases to the proportion of effective

[13] As a measure of poverty, we use the Progress out of Poverty Index (PPI), developed by the Grameen Foundation (Desiere, et al., 2015). PPI is constructed based on responses on a series of questions about household resources and the occupation and education levels of heads of households.

time on task by the primary instructor; (2) evidence of improved quality / productivity of time on task by the primary instructor; (3) reduction of effective student-teacher ratio by combined efforts of studio teacher and in-person facilitator.

First, the program could increase effective time on task (i.e., attendance and the amount of time instruction takes place). This is particularly important given schools struggle with high teacher and student absenteeism, and the program may have function to improve overall student attendance. As shown in Table 7, across all measures of student attendance, we observe no differences between attendance in treatment and control schools.

[Table 7]

We also examined measures of student beliefs and engagement as we conjecture that the quality of time-on-task (and not just quantity) may matter for student learning. We examined agreement with statements about aspirations, confidence, and self-esteem, as well as asked students to report the amount of time spent on school work. Across all of these measures, we find no differences between treatment and control. We do acknowledge that it is difficult to capture these measures through self-reported surveys. Social desirability bias (i.e., respondents' desire to provide the "correct' response to enumerators) might have led both treatment and control students to respond similarly.

The Stallings classroom observations allow us to examine the proportion of time devoted to various classroom activities by the primary instructor. In satellite classes, the primary teacher is defined as the studio teacher when video feed is actively being viewed and facilitators when video feed is not being viewed). Table 8 shows the proportion of time spent in active instruction, passive instruction, organization and management, and time off task. A comparison of the amount of time in active and passive instruction with standard control classrooms reveals no significant differences. Overall more time is spent on classroom management in satellite classes. The absence of differences in the proportion of time spent on instruction suggests that there is something qualitatively different about the time on task in satellite classes that leads to gains in student learning.

[Table 8]

Second, the quality of the actual instructional time may be improved through superior pedagogical methods or improvements to the primary instructor's subject-specific content knowledge and/or teaching ability. Student engagement, as assessed by observers, is not significantly different between treatment and control classrooms. Observer assessments of facilitator engagement and knowledge are marginally significantly lower in treatment classrooms, but this may reflect that they are not able to showcase their skills in the same way as in a standard classroom.

Our observation data are unfortunately not granular enough to capture improvements in pedagogies employed.[14] To assess whether there are differences in the pedagogical methods used in the classrooms, we compared facilitator and teacher self-reports of methods used when teaching standard (non-satellite) classes (see appendix table 6). We find no significant differences between facilitators and standard teachers in control schools (with the exception that facilitators are more likely to use technology in classrooms, which is most likely a function of the availability of technology due to the treatment.[15] We recommend caution when interpreting these results as the response categories of methods were course and susceptible to response bias given responses were self-reported.

---

[14] We do have some anecdotal evidence from qualitative interviews suggesting that new teaching methods and approaches were introduced into classroom instruction, not just through the presence of the studio teacher, but also through learning on the part of the facilitators through their interactions and observations of studio teachers.

[15] We recommend caution when interpreting these results as the response categories for the pedagogical methods were coarse and susceptible to response bias given responses were self-reported.

Third, the presence of the studio teacher (when satellite feed is actively being viewed) and the facilitator essentially reduces the student to teacher ratio, allowing the facilitator to engage in other meaningful classroom activities. The examination of the total time-on-task of the primary instructor in classrooms, does not capture the fact that there are essentially now two instructors in a classroom for the same number of students. The presence of the studio teacher can allow the in-person facilitator to engage in other meaningful classroom activities such as working in small groups or one-on-one with struggling students or engage in classroom management while students are engaged in active instruction with the studio teacher.

Table 9 details facilitators' activities when the video feed is being viewed compared to when the video feed is not being viewed. We find that facilitators are off task less than half of the time that the video feed is being viewed, meaning that facilitators are engaged in meaningful classroom activities a substantial proportion of the time in which the studio teacher is an instructional presence in the classroom. Roughly 10 percent of the time the satellite feed is being viewed, facilitators are engaged in instruction of some sort with one or more students.[16]

[Table 9]

## 6 Discussion and Cost Effectiveness[17]

Our findings suggest that live, interactive distance learning via videoconferencing can overcome poor local instruction in remote environments. The videoconferencing model addresses the shortcomings of non-interactive distance-learning models by creating face-to-

[16] In subsequent analyses, we will also use Stallings data to measure classroom disruption and the proportion of time one or more students are off-task to understand how the presence of dual teachers affects the overall classroom dynamic.
[17] Cost-effectiveness estimates are preliminary in this iteration of the paper as we are finalizing cost estimates and number of students served with the Varkey Foundation.

face interactions virtually. The interactive model we examined also essentially reduces the student-teacher ratio and can allow for in-person teachers to allocate their time more efficiently; rather than leading the class in general instruction, they can help with classroom management and engage with students who learn more slowly or more quickly than their peers.

Although the model improved learning, the pilot program was relatively costly. Upfront fixed costs included the establishment of a centralized teaching studio and computing and satellite infrastructure at each school. According to estimates provided by the Varkey Foundation, these fixed costs comprised 43 percent of the total costs of the program over the two years of its operation. Ongoing costs included the maintenance of technology, continued satellite connectivity, teachers and facilitator salaries, and other administrative costs.

To gauge the cost effectiveness of the program, we map the pilot's costs to the one-year learning gains observed in numeracy (0.22 standard deviations). Specifically, we calculate the EGMA standard deviation gain per $100 USD spent. Because fixed costs comprise a large proportion of total costs, the overall cost per pupil decreases as the program's longevity lengthens.[18] Program implementers predict that the technological infrastructure can last eight to ten years if it is properly maintained. To demonstrate the cost effectiveness of a hypothetical in which technology lasts for up to ten years, Figure 2 displays the EGMA standard deviation gains per $100 USD for up to ten years of programming when fixed costs remain constant. As fixed costs are distributed across longer treatment terms (and more

[18] We applied costs to the capacity that the pilot program was intended to accommodate (80 students per school), although the actual number per school varied. In some smaller schools, there were insufficient students in grades two to five to fill all 80 slots. In other schools, there were more than 80 students, but not so many more that it made sense to run additional (non-satellite) math and English classes; in these cases, more than 80 students were enrolled in the satellite classes.

cohorts of beneficiaries), the projected standard deviations in numeracy gains per $100 USD spent increases.

Due to changes in the intensity of lessons and number of students served, we are unable to compare the cost effectiveness of one year of program participation to two years' worth.[19] That said, we note that the substantial gains in numeracy after participating in their first year were not observed after a second year. This could be due to the reduced frequency of classes or could simply reflect that the gains to be made in numeracy are accomplished after one year. Either way, the additional year of satellite math classes may be far less cost-effective than the first given no additional gains were observed.

However, this second-year drop-off in efficiency may not pertain to literacy programs, as students demonstrated greater (though still modest) gains in fluency in their second year as compared to their first. Students' overall gains in basic literacy skills (e.g., EGRA letters-per-minute) were far greater than were their gains in fluency, but because we did not measure basic subtasks at midline, we cannot compare the cost effectiveness of one-year gains in basic language skills to that of two-year cost-effectiveness.

[Figure 2]

Additionally, we show cost effectiveness when certain costs are excluded. First, we gauge cost effectiveness when the costs of snacks provided to participants in after-school lessons are excluded; we deem it is reasonable to imagine that the non-academic after-school lessons for girls had minimal impact on academic outcomes. We make this assumption not

[19] We are unable to conducted a cost-effectiveness analysis of the second year for the following two reasons: (1) Because students in the RCT received MGCubed for two subsequent years, we cannot estimate the impact of the 2nd year separate from that of the 1st year. Any effect observed in the 2nd year is a product of also having received the first year of the program. (2) Because the number of students served doubles in the second year, we have no solely 2nd year effect of the program to map to the greater number of students served that only received that second year's set of lessons

only because there were no disparities in learning outcomes between boys and girls, but also because after-school lessons reached fewer students than did the math and English classes.

Second, we exclude satellite connectivity costs, which comprise 14 percent of total costs. Large technology companies are working to expand the reach of the Internet to digitally isolated parts of the world (e.g., through Facebook's Internet.org initiative and Google's Project Loon); in this climate, it becomes increasingly possible that rural schools in Ghana will soon gain free or heavily subsidized Internet access.[20] While the technological infrastructure in rural Ghana today is incapable of hosting this type of intervention, the costs of taking rural areas online will likely lower considerably as technology improves. As shown in Figure 2, reduction in technological connectivity costs increases the cost effectiveness of distance learning, particularly when fewer years of participation are assumed (i.e., when fixed costs are not distributed across five or more years).

Nevertheless, the cost effectiveness of the pilot is still considerably lower than that of less technologically involved interventions in similar contexts (Kremer et al., 2013). We optimistically note that the cost effectiveness could improve in subsequent scale up and through other modifications. As with most pilot programs, the program's costs were probably not optimized and therefore were probably not representative of the cost per student in a scaled-up context. Consultancy costs, for example, comprised nearly 10 percent of start-up fixed costs, and centralized staffing costs comprised roughly 16 percent of each year's budget. While we acknowledge that the extra efforts made during pilot stages tend to make pilots more successful than pared-down and scaled-up iterations, we speculate there are ways to optimize program administration costs without sacrificing learning gains.

---

[20] Facebook and Google are already establishing Internet connectivity by laying fiber-optic cables and circulating unmanned solar-powered drones (Prinsloo, 2017).

To conclude, this pilot represents a useful first step in understanding how distance learning can be used to improve learning outcomes in remote areas.  While costly, the pilot successfully increased basic literacy and numeracy skills. An examination of possible mechanisms suggests that improved quality of instruction (rather than increased quantity of instruction time) drove these gains. In order to understand how best to leverage distance instruction, future studies should explore the complementary roles of the distance and in-person teachers and examine ways to optimize costs at larger scales of implementation.

## 7 References


Ajayi, K. F. (2011). School choice and educational mobility: Lessons from secondary school applications in Ghana. University of California-Berkeley Working Paper.

Barrera-Osorio, F., and Linden, L. L. (2009). *The use and misuse of computers in education: evidence from a randomized experiment in Colombia.* World Bank Policy Research Working Paper Series.

Barbour, M. K. (2007). Portrait of Rural Virtual Schooling. *Canadian Journal of Educational Administration and Policy, 59*, 1-21.

Barker, B. O., & Hall, R. F. (1994). Distance education in rural schools: Technologies and practice. *Journal of Research in Rural Education,* 10(2), 126-128.

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

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. & Walton, M. (2016). Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of “Teaching at the Right Level” in India (No. w22746). National Bureau of Economic Research.

Bettinger, E., Fox, L., Loeb, S., & Taylor, E. S. (Forthcoming). Virtual Classrooms: How Online College Courses Affect Student Success. *The American Economic Review*.

Bhattacharya, B. (2008). Engineering education in India–the role of ICT. *Innovations in education and teaching International, 45*(2), 93-101.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., and Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, 27(4), 793-818.

Cavalluzzo, L., Lowther, D. L., Mokher, C., and Fan, X. (2012). Effects of the Kentucky Virtual Schools' Hybrid Program for Algebra I on Grade 9 Student Math Achievement. Final Report. NCEE 2012-4020. National Center for Education Evaluation and Regional Assistance.

Cavanaugh, C. S. (1999). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. Retrived from https://eric.ed.gov/?id=ED430547 (20 March 2017)

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673-682.

Darvas, P., & Balwanz, D. (2013). *Basic education beyond the Millennium Development Goals in Ghana: How equity in service delivery affects educational and learning outcomes.* World Bank Publications.

De, M. (2004). EDUSAT- the Indian satellite for education. *Current Science*, 87(8), 1034.

Desai, V., Murthy, J., Saksena, R., & Gupta, A. (2009). Enhancement of primary education using EDUSAT: Rajiv Gandhi project for EDUSAT supported elementary education network (RGPEEE) overview. *Acta Astronautica*, 65(9), 1440-1445.

Desiere, S., Vellema, W., & D'Haese, M. (2015). A validity assessment of the Progress out of Poverty Index (PPI)™. *Evaluation and program planning, 49*, 10-18.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *The American Economic Review, 101*(5), 1739-1774.

Duflo, E., Dupas, P., & Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics*, 123, 92-110.

Duflo, A., & Kiessel, J. (2014). Every Child Counts: Adapting and Evaluating Research Results on Remedial Education across Contexts. Society for Research on Educational Effectiveness 2014 Conference.

Duthilleul, Y. (2005). *Lessons Learnt in the Use of. International Institute for Educational Planning* (IIEP) UNESCO. 7-9 rue Eugene-Delacroix, 75116 Paris, France.

Fagernäs, S., & Pelkonen, P. (2012). Preferences and skills of Indian public sector teachers. *IZA Journal of Labor & Development*, 1(1), 3.

Figlio, D, Rish, M., & Yin, L. 2013. "Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning." *Journal of Labor Economics* 31(4): 763-784.

Ganimian, A. J., & Murnane, R. J. (2016). Improving Education in Developing Countries Lessons From Rigorous Impact Evaluations. *Review of Educational Research*, 86(3), 719-755.

Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education, 2*, 945-1017.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010* (No. w17554). National Bureau of Economic Research.

Hart, C. M., Friedmann, E., & Hill, M. (2016). Online course-taking and student outcomes in California community colleges. *Education Finance and Policy.*

Heppen, J., Walters, K., Pareja, A. S., Clements, M., Faria, A., & Sorensen, N., Culp, K. (2011). *Access to Algebra I: The Effects of Online Mathematics for Grade 8 Students.* Institute for Education Sciences REL Northeast & Islands Report. NCEE 2012-4021.

Heppen, J., Walters, K., Sorensen, N., Allensworth, E., Walters., K., Rickles, J., Taylor, S., & Michelman, V. (2017). The Struggle to Pass Algebra I in Urban High Schools: Online vs. Face-to-Face Credit Recovery for At-Risk Students. *Journal of Research on Educational Effectiveness, 10*(2), 272-296, DOI: 10.1080/19345747.2016.1168500.

He, F., Linden, L., & MacLeod, M. (2008). *How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program.* Columbia University: Working Paper.

Ho, J., & Thukral, H. (2009). Tuned in to student success: Assessing the impact of interactive radio instruction for the hardest-to-reach. *Journal of Education for International Development*, 4(2), 34-51.

Jumani, N. B. (2009). Study on role of radio for rural education in Pakistan. *Turkish Online Journal of Distance Education*, 10(4).

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130), 297-300.

Lucas, F. F. B. (1999). A radio broadcasting model for rural women and farm households. *RAP Publication*, 8. FAO Regional Office for Asia and the Pacific, Bangkok.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research, 85*(3), 353-394.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* US Department of Education.

Ministry of Education, Ghana Education Service. (2011). National Education Assessment: 2011 Findings Report. Education Data for Decision Making (EdData II) project, Task Order Number AID-641-BC-11-00001.

Ministry of Education, Ghana Education Service. (2016a). Education Sector Performance Report 2016.

Ministry of Education, Ghana Education Service. (2016b). 2016 National Education Assessment: Report of Findings. Education Data for Decision Making (EdData II) project, Task Order Number AID-641-BC-13-00001.

Naslund-Hadley, E., Parker, S. W., & Hernandez-Agramonte, J. M. (2014). Fostering early math comprehension: Experimental evidence from Paraguay. *Global Education Review*, 1(4).

OECD (2013). What makes urban schools different? *PISA In Focus, 28*: 2013/05. PISA, OECD Publishing.

Orazem, P. F., & King, E. M. (2007). Schooling in developing countries: The roles of supply, demand and government policy. *Handbook of development economics, 4*, 3475-3559.

Phipps, R., & Merisotis, J. (1999). *What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education.* American Federation of Teachers, Washington, DC.; National Education Association, Washington, DC.

Picciano, A. G. (2002). Beyond student perceptions: Issues of interaction, presence, and performance in an online course. *Journal of Asynchronous learning networks*, 6(1), 21-40.

Picciano, A. G., & Seaman, J. (2009). K-12 Online Learning: A 2008 Follow-Up of the Survey of US School District Administrators. Sloan Consortium. PO Box 1238, Newburyport, MA 01950.

Prinsloo, L. (2017, April 5). Facebook Adds Wifi Hotspots to Sustain Africa Growth. Bloomberg Technology. https://www.bloomberg.com/news/articles/2017-04-05/facebook-adds-wifi-hotspots-to-sustain-africa-customer-growth (accessed April 12 2017).

Reubens A. & Kline T. (2009) Pilot of the Early Grade Mathematics Assessment: final report. Prepared for the USAID Bureau for Economic Growth, Agriculture and Trade (EGAT/ED) under the Education Data for Decision Making (EdData II) Project. Research Triangle Park, NC: RTI International.

Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73: 417–458. doi: 10.1111/j.1468-0262.2005.00584.x

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94, 247–252.

Rockoff, Jonah E., 2009. Field experiments in class size from the early twentieth century. *Journal of Economic Perspectives*, 23 (4), 211–230.

RTI International (2009) Early Grade Reading Assessment Toolkit. Prepared with funding from the World Bank and the USAID EdData II Project.

Sanders, W. and J. Rivers (1996). *Cumulative and Residual Effects of Teachers on Future Academic Achievement.* Technical report, University of Tennessee Value-Added Research and Assessment Center.

Schwartzbeck, T. D., Prince, C. D., Redfield, D., Morris, H., & Hammer, P. (2003). How are rural districts meeting the teacher quality requirements of No Child Left Behind. Charleston, VA*: Appalachia Educational Laboratory*.

Snilstveit, B, Stevenson, J, Phillips, D, Vojtkova, M, Gallagher, E, Schmidt, T, Jobse, H, Geelen, M , Pastorello, M, and Eyers, J. (2015). *Interventions for improving learning outcomes and access to education in low - and middle - income countries: a systematic review,* 3ie Final Review. London: International Initiative for Impact Evaluation (3ie)

Stallings, J. (1977). *Learning to look: A handbook on classroom observation and teaching models.* Wadsworth Publishing Company.

Swan, K. (2002). Building learning communities in online courses: The importance of interaction. Education, *Communication & Information, 2(*1), 23-49.

Tucker, B. (2007). Laboratories of reform: Virtual high schools and innovation in public education. *Education Sector Reports,* 1-15.

Xu, D., & Jaggars, S.S. (2011). The effectiveness of distance education across Virginia's Community Colleges: Evidence from introductory college-level math and English courses. *Educational Evaluation and Policy Analysis, 33*(3), 360−377.

World Bank. (2005). Improving Educational Quality through Interactive Radio Instruction: A Toolkit for Policymakers and Planners. Africa Region Human Development Working Paper Series. Development Research Group, The World Bank.

World Bank. (2012). Project Appraisal Document on a Global Partnership for Education Fund Grant to the Republic of Ghana. Washington DC: World Bank, Central and West Africa Education Unit, Africa Region. Report No: 72849-GH.

World Bank. (2015). *Conducting Classroom Observations: Using the Stallings "Classroom Snapshot" Observation System*. User Guide. World Bank Group, 20150.

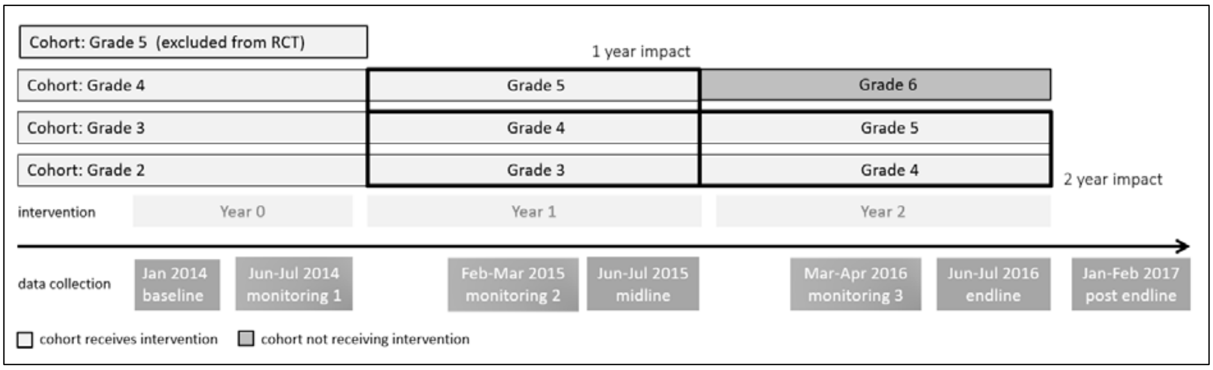


**Figure 1. Timeline of satellite class intervention and data collection, by cohort**

*Notes:* This figure shows the timeline of satellite class participation and data collection for each treatment cohort. Baseline, midline, and endline data were comprised of student surveys and assessments of math and English skills. Monitoring data was gathered by way of attendance checks and through one-on-one interviews of and focus groups with students, teachers, and school administrators.

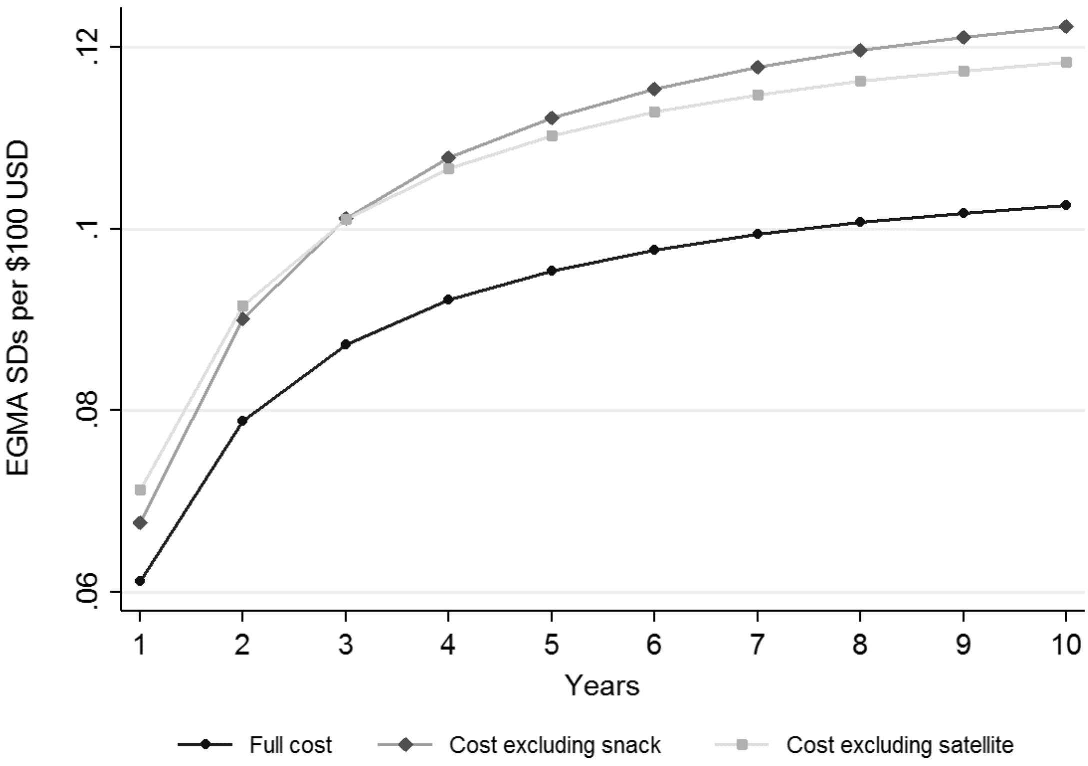


**Figure 2. EGMA Standard deviations per $100 USD over ten years**

*Notes:* This figure maps average standard deviation gains on the Early Grade Mathematics Assessment (EGMA) per $100 USD spent. Fixed costs are distributed across the number of years (represented on the x axis) to show cost effectiveness when initial infrastructure lasts for up to 10 years of programming. The figure compares gains in math performance among students who participated for one year to gains among students who participated for two. The figure also shows cost effectiveness when certain costs are excluded. These non-essential costs include 1) after-school snacks provided to girls as part of the girls' after-school health and empowerment lessons and 2) satellite connectivity costs that could be eliminated if free wifi were to become available.

**Table 1. Baseline covariate balance at baseline**

| | All students | | | Treatment | | | Control | | | Difference (T-C) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | n | mean | sd | n | mean | sd | n | coef | se | p-value |
| Panel A: Student & Household Characteristics | | | | | | | | | | | | |
| Identifies as female | 0.61 | (0.49) | 5485 | 0.61 | (0.49) | 2607 | 0.61 | (0.49) | 2878 | 0.01 | (0.01) | 0.550 |
| Age at baseline | 10.94 | (2.10) | 5379 | 10.98 | (2.10) | 2563 | 10.91 | (2.10) | 2816 | 0.11 | (0.12) | 0.346 |
| Number of school days absent in last week | 0.80 | (1.24) | 5242 | 0.79 | (1.26) | 2504 | 0.80 | (1.22) | 2738 | -0.01 | (0.05) | 0.876 |
| In grade 2 at baseline | 0.33 | (0.47) | 5485 | 0.32 | (0.47) | 2607 | 0.34 | (0.47) | 2878 | -0.02 | (0.01) | 0.139 |
| In grade 3 at baseline | 0.34 | (0.47) | 5485 | 0.34 | (0.48) | 2607 | 0.33 | (0.47) | 2878 | 0.02 | (0.01) | 0.218 |
| In grade 4 at baseline | 0.34 | (0.47) | 5485 | 0.34 | (0.47) | 2607 | 0.33 | (0.47) | 2878 | 0.01 | (0.01) | 0.596 |
| Travels 60 minutes or more to school | 0.14 | (0.34) | 5092 | 0.12 | (0.32) | 2429 | 0.15 | (0.36) | 2663 | -0.04 | (0.02) | 0.024 ** |
| Total permanent household members | 7.84 | (3.98) | 5098 | 7.93 | (4.09) | 2414 | 7.76 | (3.89) | 2684 | 0.18 | (0.18) | 0.305 |
| Single mother (father not present in) household | 0.16 | (0.37) | 5407 | 0.16 | (0.37) | 2561 | 0.16 | (0.37) | 2846 | 0.00 | (0.01) | 0.999 |
| All children 5-12 years in household attend school | 0.89 | (0.32) | 5417 | 0.88 | (0.32) | 2565 | 0.89 | (0.32) | 2852 | 0.00 | (0.01) | 0.789 |
| Household has permanent roof (e.g. iron, cement, tile) | 0.90 | (0.30) | 5375 | 0.90 | (0.30) | 2545 | 0.90 | (0.30) | 2830 | -0.01 | (0.02) | 0.771 |
| Household has wired electricity | 0.62 | (0.49) | 5395 | 0.63 | (0.48) | 2557 | 0.61 | (0.49) | 2838 | 0.03 | (0.04) | 0.497 |
| Household has indoor plumbing | 0.52 | (0.50) | 5393 | 0.53 | (0.50) | 2558 | 0.52 | (0.50) | 2835 | -0.01 | (0.03) | 0.716 |
| Household has working stove | 0.17 | (0.38) | 5302 | 0.18 | (0.39) | 2505 | 0.16 | (0.37) | 2797 | 0.02 | (0.02) | 0.424 |
| Household has working iron | 0.60 | (0.49) | 5389 | 0.62 | (0.49) | 2553 | 0.59 | (0.49) | 2836 | 0.03 | (0.02) | 0.208 |
| Household has working radio | 0.26 | (0.44) | 5391 | 0.26 | (0.44) | 2552 | 0.25 | (0.43) | 2839 | 0.01 | (0.02) | 0.743 |
| Mother attended some school | 0.56 | (0.50) | 5399 | 0.56 | (0.50) | 2557 | 0.55 | (0.50) | 2842 | 0.01 | (0.03) | 0.771 |
| Father attended some school | 0.63 | (0.48) | 5387 | 0.63 | (0.48) | 2550 | 0.63 | (0.48) | 2837 | -0.01 | (0.02) | 0.846 |
| Language spoken at home | | | | | | | | | | | | |
| Twi | 0.10 | (0.30) | 5475 | 0.10 | (0.29) | 2601 | 0.10 | (0.30) | 2874 | -0.01 | (0.02) | 0.693 |
| Ewe | 0.19 | (0.40) | 5475 | 0.21 | (0.40) | 2601 | 0.18 | (0.39) | 2874 | 0.02 | (0.03) | 0.537 |
| Ga / Dangme | 0.46 | (0.50) | 5475 | 0.45 | (0.50) | 2601 | 0.47 | (0.50) | 2874 | -0.01 | (0.02) | 0.446 |
| English | 0.02 | (0.15) | 5475 | 0.02 | (0.15) | 2601 | 0.02 | (0.15) | 2874 | 0.00 | (0.01) | 0.861 |
| Other | 0.31 | (0.46) | 5475 | 0.32 | (0.47) | 2601 | 0.30 | (0.46) | 2874 | 0.01 | (0.03) | 0.850 |
| Religion | | | | | | | | | | | | |
| Catholic | 0.10 | (0.30) | 5402 | 0.10 | (0.30) | 2559 | 0.10 | (0.31) | 2843 | -0.01 | (0.02) | 0.784 |
| Presbyterian | 0.10 | (0.30) | 5402 | 0.09 | (0.29) | 2559 | 0.11 | (0.31) | 2843 | -0.02 | (0.02) | 0.322 |
| Pentacostal | 0.28 | (0.45) | 5402 | 0.26 | (0.44) | 2559 | 0.29 | (0.45) | 2843 | -0.04 | (0.02) | 0.070 * |
| Charismatic | 0.23 | (0.42) | 5402 | 0.24 | (0.43) | 2559 | 0.22 | (0.41) | 2843 | 0.03 | (0.02) | 0.054 * |
| Other Christian | 0.09 | (0.28) | 5402 | 0.09 | (0.28) | 2559 | 0.09 | (0.28) | 2843 | 0.00 | (0.01) | 0.893 |
| Muslim | 0.11 | (0.31) | 5402 | 0.11 | (0.32) | 2559 | 0.11 | (0.31) | 2843 | 0.01 | (0.03) | 0.684 |
| Other | 0.10 | (0.30) | 5402 | 0.11 | (0.31) | 2559 | 0.09 | (0.29) | 2843 | 0.02 | (0.02) | 0.321 |
| Panel B: Baseline Test Scores | | | | | | | | | | | | |
| EGRA Subtasks | | | | | | | | | | | | |
| Words per minute | 11.53 | (21.21) | 5485 | 11.65 | (22.23) | 2607 | 11.41 | (20.26) | 2878 | 0.29 | (1.03) | 0.779 |
| Letters per minute | 7.94 | (13.78) | 5485 | 8.04 | (14.36) | 2607 | 7.85 | (13.22) | 2878 | 0.16 | (0.72) | 0.822 |
| Invented words per minute | 2.38 | (7.00) | 5485 | 2.55 | (7.57) | 2607 | 2.22 | (6.44) | 2878 | 0.36 | (0.28) | 0.201 |
| Oral vocabulary score (out of 8) | 3.72 | (1.29) | 5485 | 3.71 | (1.26) | 2607 | 3.73 | (1.32) | 2878 | -0.01 | (0.05) | 0.887 |
| Listening comprehension score (out of 3) | 0.51 | (0.86) | 5485 | 0.53 | (0.88) | 2607 | 0.48 | (0.84) | 2878 | 0.03 | (0.05) | 0.462 |
| Reading comprehension score (out of 2) | 0.12 | (0.40) | 5485 | 0.12 | (0.41) | 2607 | 0.12 | (0.39) | 2878 | 0.01 | (0.02) | 0.731 |
| EGMA Subtasks | | | | | | | | | | | | |
| Number identification score (out of 20) | 14.30 | (4.37) | 5485 | 14.30 | (4.34) | 2607 | 14.29 | (4.40) | 2878 | 0.06 | (0.22) | 0.780 |
| Quantity discrimination score (out of 10) | 6.49 | (2.71) | 5485 | 6.49 | (2.65) | 2607 | 6.48 | (2.76) | 2878 | 0.05 | (0.13) | 0.690 |
| Missing number score (out of 10) | 2.95 | (1.79) | 5485 | 2.94 | (1.76) | 2607 | 2.97 | (1.82) | 2878 | -0.02 | (0.08) | 0.773 |
| Addition / subtraction score (out of 56) | 19.72 | (13.00) | 5485 | 19.58 | (12.89) | 2607 | 19.85 | (13.10) | 2878 | -0.05 | (0.63) | 0.941 |
| EGMA total (out of 100) | 46.02 | (20.13) | 5485 | 45.84 | (19.77) | 2607 | 46.18 | (20.45) | 2878 | 0.00 | (1.01) | 0.996 |
| Joint test (p-value) - All variables | | | | | | | | | | | | 0.150 |
| Joint test (p-value) - Student & household characteristics | | | | | | | | | | | | 0.351 |
| Joint test (p-value) - EGRA & EGMA subtasks | | | | | | | | | | | | 0.507 |

*Notes:* Table shows the means and standard deviations of student baseline characteristics. For binary characteristics, the proportion of students with the characteristic is shown. The treatment-control difference is the coefficient from a regression of the dependent variable on an indicator variable for treatment and randomization strata (i.e. district) fixed effects. Thus, the difference shown is not exactly equal to the difference between the treatment and control means shown. Results are robust to omitting the strata fixed effects. Standard errors are clustered at the school level. The EGMA total is calculated such that number identification, quantity discrimination, and missing number subtasks are each weighted as 20 percent of the total and addition / subtract as 40 percent of the total. Joint test excludes reference categories. (N=4545). ***p<0.01, ** p<0.05, * p<0.1

**Table 2. Effect of satellite-classes on average standardized EGRA and EGMA: Pooled Years**

| | EGRA Words Per Minute | | | | | | EGMA Math Total | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Treated * Post | 0.03 | 0.03 | 0.02 | 0.02 | 0.02 | 0.03 | 0.24*** | 0.24*** | 0.23*** | 0.23*** | 0.23*** | 0.23*** |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| Treated | 0.00 | 0.01 | 0.00 | -0.01 | . | . | -0.02 | -0.02 | -0.02 | -0.03 | . | . |
| | (0.07) | (0.05) | (0.05) | (0.05) | . | . | (0.08) | (0.05) | (0.05) | (0.05) | . | . |
| Post | -0.07*** | -0.07*** | 0.02 | 0.03* | 0.03 | 0.01 | -0.07*** | -0.06*** | 0.02 | 0.03 | 0.03 | 0.01 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Grade 2 cohort | | | -0.81*** | -0.83*** | -0.90*** | | | | -0.82*** | -0.76*** | -0.82*** | |
| | | | (0.05) | (0.06) | (0.06) | | | | (0.03) | (0.03) | (0.03) | |
| Grade 3 cohort | | | -0.50*** | -0.51*** | -0.54*** | | | | -0.42*** | -0.38*** | -0.42*** | |
| | | | (0.04) | (0.04) | (0.04) | | | | (0.03) | (0.03) | (0.03) | |
| Observations | 12,849 | 12,849 | 12,849 | 12,849 | 12,849 | 12,849 | 12,848 | 12,848 | 12,848 | 12,848 | 12,848 | 12,848 |
| R-squared | 0.001 | 0.098 | 0.197 | 0.232 | 0.154 | 0.002 | 0.008 | 0.137 | 0.237 | 0.267 | 0.160 | 0.047 |
| District fixed effects | | Yes | Yes | Yes | | | | Yes | Yes | Yes | | |
| Cohort fixed effects | | | Yes | Yes | Yes | | | | Yes | Yes | Yes | |
| Baseline covariates | | | | Yes | Yes | | | | | Yes | Yes | |
| School fixed effects | | | | | Yes | | | | | | Yes | |
| Student fixed effects | | | | | | Yes | | | | | | Yes |

*Notes* : Each column represents a separate regression. In Panel A, the outcome variable of interest is the EGRA words-per-minute measure, and in Panel B, the outcome variable is the total EGMA score, with subsections weighted equally). Treated is an indicator for whether schools were assigned to the satellite-class intervention, zero otherwise. Post is an indicator variable equal to one after students had at least one year of the satellite class intervention (the Year 1 / midline and Year 2 / endline test scores). Districts are the randomization strata. Baseline characteristics are those included in Table 1; missing observations were imputed using the mean of non-missing observations. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1

**Table 3. Effect of satellite-classes on average standardized EGRA and EGMA: By Year**

| | EGRA Words Per Minute | | | | EGMA Math Total | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treated * Post (1 year treatment) | -0.00 | -0.00 | -0.01 | -0.01 | 0.22*** | 0.22*** | 0.22*** | 0.22*** |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| Treated * Post (2 year treatment) | 0.08* | 0.07* | 0.07 | 0.05 | 0.26*** | 0.26*** | 0.25*** | 0.23*** |
| | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) |
| Post (1 year treatment) | -0.00 | 0.01 | 0.01 | 0.02 | -0.01 | 0.00 | 0.00 | 0.01 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Post (2 year treatment) | -0.18*** | -0.18*** | 0.05* | 0.06** | -0.17*** | -0.17*** | 0.05** | 0.06** |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.02) | (0.02) | (0.02) | (0.02) |
| Treated | 0.00 | 0.01 | 0.00 | -0.01 | -0.02 | -0.02 | -0.02 | -0.03 |
| | (0.07) | (0.05) | (0.05) | (0.05) | (0.08) | (0.05) | (0.05) | (0.05) |
| Grade 2 cohort | | | -0.84*** | -0.85*** | | | -0.83*** | -0.77*** |
| | | | (0.05) | (0.06) | | | (0.03) | (0.03) |
| Grade 3 cohort | | | -0.52*** | -0.53*** | | | -0.43*** | -0.40*** |
| | | | (0.04) | (0.05) | | | (0.03) | (0.03) |
| Observations | 12,849 | 12,849 | 12,849 | 12,849 | 12,848 | 12,848 | 12,848 | 12,848 |
| R-Squared | 0.004 | 0.101 | 0.199 | 0.233 | 0.011 | 0.140 | 0.238 | 0.267 |
| p-value (Treated * 1 year = Treated * 2 year) | 0.025** | 0.029** | 0.025** | 0.061* | 0.269 | 0.312 | 0.304 | 0.697 |
| District fixed effects | | Yes | Yes | Yes | | Yes | Yes | Yes |
| Cohort fixed effects | | | Yes | Yes | | | Yes | Yes |
| Baseline covariates | | | | Yes | | | | Yes |

*Notes:* Each column represents a separate regression. In Panel A, the outcome variable of interest is the EGRA words-per-minute measure, and in Panel B, the outcome variable is the total EGMA score, with subsections weighted equally). Treated is an indicator for whether schools were assigned to the satellite-class intervention, zero otherwise. Post Y1 is an indicator variable equal to one after students hadthe first year of the satellite class intervention (Year 1 / midline test score) and Post Y2, an indicator if student had two years of satellite instruction (Year 2 / endline test score). Districts are the randomization strata. Baseline characteristics are imputed using the mean of non-missing observations. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1

**Table 4. Effect of satellite-classes on average standardized EGRA and EGMA: By Subtask**

| | EGRA | | | | | | EGMA | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Words per minute | Letters per minute | Invented words per minute | Oral vocabulary | Listening comprehension | Reading comprehension | Number identification | Quantity discrimination | Missing number | Addition / subtraction |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Treated * Post (1 year treatment) | -0.01 | | | | | 0.01 | 0.20*** | 0.15*** | 0.17*** | 0.19*** |
| | (0.03) | | | | | (0.05) | (0.04) | (0.04) | (0.04) | (0.04) |
| Treated * Post (2 year treatment) | 0.07 | 0.82*** | 0.17*** | 0.17** | 0.04 | 0.00 | 0.34*** | 0.12** | 0.25*** | 0.21*** |
| | (0.04) | (0.08) | (0.05) | (0.06) | (0.06) | (0.05) | (0.06) | (0.05) | (0.04) | (0.04) |
| Post (1 year treatment) | 0.01 | | | | | 0.01 | 0.00 | -0.00 | 0.00 | 0.00 |
| | (0.02) | | | | | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) |
| Post (2 year treatment) | 0.05* | 0.04 | 0.05 | -0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.04* | 0.04* |
| | (0.03) | (0.04) | (0.03) | (0.05) | (0.04) | (0.03) | (0.05) | (0.03) | (0.03) | (0.02) |
| Treated | 0.00 | 0.02 | 0.05 | -0.00 | 0.03 | 0.01 | -0.01 | -0.01 | -0.02 | -0.02 |
| | (0.05) | (0.06) | (0.05) | (0.04) | (0.05) | (0.05) | (0.05) | (0.05) | (0.04) | (0.05) |
| Grade 2 cohort | -0.84*** | -0.50*** | -0.58*** | -0.03 | -0.60*** | -0.62*** | -0.62*** | -0.66*** | -0.72*** | -0.80*** |
| | (0.05) | (0.04) | (0.06) | (0.03) | (0.04) | (0.04) | (0.04) | (0.03) | (0.03) | (0.03) |
| Grade 3 cohort | -0.52*** | -0.32*** | -0.37*** | -0.00 | -0.36*** | -0.39*** | -0.28*** | -0.31*** | -0.38*** | -0.44*** |
| | (0.04) | (0.04) | (0.05) | (0.03) | (0.04) | (0.04) | | | | |
| | 0 | | | | | | | | | |
| Observations | 12,849 | 8,088 | 8,088 | 8,088 | 8,088 | 12,849 | 12,848 | 12,848 | 12,849 | 12,849 |
| R-Squared | 0.199 | 0.153 | 0.082 | 0.063 | 0.216 | 0.134 | 0.095 | 0.105 | 0.178 | 0.244 |
| | 0 | | | | | | | | | |
| p-value (Treated * 1 year = Treated * 2 year) | 0.025** | | | | | 0.852 | 0.004*** | 0.484 | 0.054* | 0.671 |

*Notes*: Each column represents a separate regression. District fixed effects included in all models. The same estimation specification was used as described for table 3, with the exception that outcome variables are EGRA and EGMA subtasks, standardized to their contemporaneous control distributions. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1

**Table 5. Effect of satellite-classes on average standardized EGRA and EGMA: Grade 4 Cohort**

| | EGRA Words Per Minute | | | EGMA Math Total | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated * Post (1 year treatment) | 0.01 | -0.00 | 0.01 | 0.23*** | 0.23*** | 0.23*** |
| | (0.05) | (0.05) | (0.05) | (0.04) | (0.04) | (0.04) |
| Treated * Post (1 year after treatment) | 0.01 | 0.01 | 0.00 | 0.22*** | 0.22*** | 0.21*** |
| | (0.07) | (0.07) | (0.07) | (0.04) | (0.04) | (0.05) |
| Post (1 year treatment) | -0.07* | -0.05 | -0.07* | -0.12*** | -0.11*** | -0.11*** |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| Post (1 year after treatment) | -0.12*** | -0.13*** | -0.11*** | -0.18*** | -0.19*** | -0.17*** |
| | (0.04) | (0.04) | (0.04) | (0.03) | (0.03) | (0.03) |
| Treated | 0.02 | 0.04 | 0.03 | -0.05 | -0.03 | -0.03 |
| | (0.14) | (0.11) | (0.10) | (0.08) | (0.06) | (0.05) |
| Observations | 4,821 | 4,821 | 4,821 | 4,821 | 4,821 | 4,821 |
| R-Squared | 0.002 | 0.135 | 0.200 | 0.008 | 0.151 | 0.216 |
| p-value (Treated * 1 year = Treated * 2 year) | 0.942 | 0.7364 | 0.871 | 0.693 | 0.884 | 0.676 |
| District fixed effects | | Yes | Yes | | Yes | Yes |
| Baseline covariates | | | Yes | | | Yes |

*Notes:* Each column represents a separate regression. District fixed effects included in all models. The same estimation specification was used as described for table 3, with the exception that specifications were run for just the cohort of students in grade 4, who did not receive a second year of the intervention and hence their Y2 endline scores reflect skills one year after receiving the intervention. Because scores were standardized on the entire sample rather than within the grade 4 cohort, the coefficients on the post dummies are not meaningful. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1

**Table 6. Heterogeneous effects of satellite-classes on average standardized EGRA and EGMA**

| | EGRA Words Per Minute | | | | | | EGMA Math Total | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Treated * Post | 0.02 | 0.01 | 0.03 | -0.01 | 0.00 | 0.05 | 0.20*** | 0.24*** | 0.25*** | 0.26*** | 0.23*** | 0.20*** |
| | (0.04) | (0.05) | (0.04) | (0.05) | (0.05) | (0.04) | (0.04) | (0.03) | (0.03) | (0.04) | (0.04) | (0.04) |
| Treated * Post * Female | 0.01 | | | | | | 0.06 | | | | | |
| | (0.05) | | | | | | (0.04) | | | | | |
| Post * Female | 0.06* | | | | | | -0.00 | | | | | |
| | (0.03) | | | | | | (0.03) | | | | | |
| Treated * Female | 0.01 | | | | | | 0.04 | | | | | |
| | (0.06) | | | | | | (0.05) | | | | | |
| Treated * Post * Below median EGMA at baseline | | 0.04 | | | | | | -0.02 | | | | |
| | | (0.06) | | | | | | (0.05) | | | | |
| Post * Below median EGMA at baseline | | -0.12*** | | | | | | 0.40*** | | | | |
| | | (0.03) | | | | | | (0.03) | | | | |
| Treated * Below median EGMA at baseline | | 0.01 | | | | | | 0.07* | | | | |
| | | (0.08) | | | | | | (0.04) | | | | |
| Treated * Post * Below 25th percentile EGMA at baseline | | | -0.02 | | | | | | -0.07 | | | |
| | | | (0.04) | | | | | | (0.06) | | | |
| Post * Below 25th percentile EGMA at baseline | | | -0.18*** | | | | | | 0.38*** | | | |
| | | | (0.03) | | | | | | (0.04) | | | |
| Treated * Below 25th percentile EGMA at baseline | | | 0.02 | | | | | | 0.06 | | | |
| | | | (0.07) | | | | | | (0.05) | | | |
| Treated * Post * Zero EGRA WPM at baseline | | | | 0.09 | | | | | | -0.04 | | |
| | | | | (0.06) | | | | | | (0.05) | | |
| Post * Zero EGRA WPM at baseline | | | | 0.01 | | | | | | 0.20*** | | |
| | | | | (0.04) | | | | | | (0.03) | | |
| Treated * Zero EGRA WPM at baseline | | | | -0.03 | | | | | | 0.09 | | |
| | | | | (0.09) | | | | | | (0.06) | | |
| Treated * Post * Cohort 2 | | | | | 0.05 | | | | | | 0.05 | |
| | | | | | (0.07) | | | | | | (0.05) | |
| Post * Cohort 2 | | | | | 0.09* | | | | | | 0.29*** | |
| | | | | | (0.05) | | | | | | (0.04) | |
| Treated * Cohort 2 | | | | | -0.07 | | | | | | -0.00 | |
| | | | | | (0.12) | | | | | | (0.07) | |
| Treated * Post * Cohort 3 | | | | | 0.02 | | | | | | -0.02 | |
| | | | | | (0.06) | | | | | | (0.04) | |
| Post * Cohort 3 | | | | | 0.12*** | | | | | | 0.06** | |
| | | | | | (0.04) | | | | | | (0.03) | |
| Treated * Cohort 3 | | | | | -0.04 | | | | | | 0.02 | |
| | | | | | (0.11) | | | | | | (0.07) | |
| Treated * Post * below median PPI at baseline | | | | | | -0.04 | | | | | | 0.08* |
| | | | | | | (0.05) | | | | | | (0.05) |
| Post * Below median PPI at baseline | | | | | | -0.03 | | | | | | -0.07** |
| | | | | | | (0.03) | | | | | | (0.03) |
| Treated * Below median PPI at baseline | | | | | | -0.04 | | | | | | -0.06 |
| | | | | | | (0.07) | | | | | | (0.06) |
| Observations | 12,849 | 12,849 | 12,849 | 12,849 | 12,849 | 11,586 | 12,848 | 12,848 | 12,848 | 12,848 | 12,848 | 11,585 |
| R-squared | 0.098 | 0.256 | 0.185 | 0.285 | 0.198 | 0.098 | 0.141 | 0.509 | 0.493 | 0.317 | 0.242 | 0.131 |

*Notes* : Each column represents a separate regression. The same estimation specification was used as described for table 2, with the exception that treatment and post are included in three way interactions with indicators for student subgroups to test for heterogeneity in treatment effects. All regressions include binary variables for treatment and post and district fixed effects. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1

**Table 7. Effect of satellite-classes on student attendance**

| | Treatment | | | Control | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | n | mean | sd | n | coef | se | p-value |
| <u>Panel A. Student attendance in Year 1</u> | | | | | | | | | |
| Days attended in term 1 | 48.99 | (10.96) | 2225 | 48.29 | (12.85) | 2569 | 0.39 | (0.72) | 0.593 |
| Days attended in term 2 | 58.31 | (15.19) | 2218 | 57.22 | (16.18) | 2517 | 0.73 | (0.66) | 0.268 |
| Days attended in term 3 | 22.29 | (9.37) | 2212 | 21.67 | (9.25) | 2531 | 0.76 | (0.89) | 0.398 |
| Days attended prior week | 3.88 | (1.60) | 2217 | 3.90 | (1.54) | 2508 | -0.02 | (0.08) | 0.834 |
| Share attended prior day of survey | 0.84 | (0.37) | 2108 | 0.83 | (0.37) | 2417 | 0.00 | (0.02) | 0.812 |
| Share attended on same day of survey | 0.88 | (0.32) | 2141 | 0.87 | (0.34) | 2396 | 0.02 | (0.01) | 0.219 |
| Standardized index | 0.03 | (0.57) | 1977 | 0.00 | (0.55) | 2300 | 0.04 | (0.03) | 0.245 |
| <u>Panel B. Student attendance in Year 2</u> | | | | | | | | | |
| Days attended in term 1 | 62.56 | (9.92) | 1907 | 62.59 | (9.60) | 2122 | -0.05 | (0.47) | 0.924 |
| Days attended in term 2 | 56.09 | (10.21) | 1875 | 56.53 | (9.67) | 2082 | -0.47 | (0.50) | 0.356 |
| Days attended in term 3 | 24.24 | (7.84) | 1872 | 24.85 | (8.09) | 2080 | -0.61 | (0.97) | 0.527 |
| Days attended prior week | 4.30 | (1.15) | 1877 | 4.26 | (1.14) | 2062 | 0.02 | (0.07) | 0.770 |
| Share attended prior day of survey | 0.81 | (0.40) | 2027 | 0.82 | (0.38) | 2158 | 0.00 | (0.02) | 0.919 |
| Share attended on same day of survey | 0.86 | (0.35) | 1910 | 0.84 | (0.37) | 2083 | 0.02 | (0.02) | 0.242 |
| Standardized index | 0.17 | (0.52) | 1841 | 0.17 | (0.56) | 2020 | -0.01 | (0.03) | 0.831 |

*Notes* : Table shows the means and standard deviations of student attendance measusres in years 1 and 2 of the program. The treatment-control difference is the coefficient from a regression of the attendance measures on an indicator variable for treatment and randomization strata (i.e. district) fixed effects. Thus, the difference shown is not exactly equal to the difference between the treatment and control means shown. A standardized index was created averaging all measures in each year. Results are robust to omitting the strata fixed effects. Standard errors are clustered at the school level.

**Table 8. Teacher activities as measured by classroom observations in treatment and control schools**

| | Treatment Satellite | | Treatment Standard | | Control Standard | | Difference (Satellite - Control) | | | Difference (Standard Treatment - Control) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd | coef | se | p-value | coef | se | p-value |
| **Active Instruction** | **0.54** | **(0.25)** | **0.54** | **(0.25)** | **0.59** | **(0.24)** | **-0.05** | **(0.04)** | **0.184** | **-0.05** | **(0.03)** | **0.139** |
| Reading aloud | 0.04 | (0.11) | 0.08 | (0.16) | 0.07 | (0.13) | -0.02 | (0.02) | 0.193 | 0.01 | (0.02) | 0.624 |
| Instruction / demonstration / lecture | 0.34 | (0.24) | 0.33 | (0.22) | 0.37 | (0.25) | -0.06 | (0.04) | 0.145 | -0.05 | (0.03) | 0.138 |
| Discussion | 0.10 | (0.13) | 0.10 | (0.14) | 0.12 | (0.15) | 0.01 | (0.02) | 0.498 | -0.01 | (0.02) | 0.469 |
| Practice / drill | 0.05 | (0.09) | 0.03 | (0.08) | 0.03 | (0.06) | 0.02 | (0.01) | 0.142 | 0.00 | (0.01) | 0.720 |
| Kinesthetic / projects | 0.00 | (0.00) | 0.00 | (0.01) | 0.00 | (0.02) | 0.00 | (0.00) | 0.061 * | 0.00 | (0.00) | 0.343 |
| **Passive Instruction** | **0.07** | **(0.14)** | **0.07** | **(0.12)** | **0.08** | **(0.12)** | **-0.01** | **(0.02)** | **0.775** | **-0.01** | **(0.01)** | **0.646** |
| Working on class assignments | 0.06 | (0.13) | 0.05 | (0.1) | 0.06 | (0.10) | 0.00 | (0.02) | 0.892 | 0.00 | (0.01) | 0.728 |
| Copying | 0.01 | (0.05) | 0.02 | (0.06) | 0.02 | (0.07) | -0.01 | (0.01) | 0.298 | 0.00 | (0.01) | 0.748 |
| **Organization and Management** | **0.28** | **(0.21)** | **0.20** | **(0.17)** | **0.18** | **(0.17)** | **0.10** | **(0.03)** | **0.000 ***** | **0.02** | **(0.02)** | **0.275** |
| Verbal instruction | 0.10 | (0.13) | 0.06 | (0.09) | 0.05 | (0.07) | 0.05 | (0.01) | 0.000 *** | 0.01 | (0.01) | 0.332 |
| Discipline | 0.00 | (0.02) | 0.01 | (0.02) | 0.00 | (0.01) | 0.00 | (0.00) | 0.736 | 0.00 | (0.00) | 0.177 |
| Classroom management | 0.08 | (0.11) | 0.05 | (0.08) | 0.04 | (0.07) | 0.04 | (0.01) | 0.014 ** | 0.01 | (0.01) | 0.342 |
| Teacher management | 0.10 | (0.14) | 0.09 | (0.14) | 0.09 | (0.13) | 0.02 | (0.02) | 0.332 | 0.00 | (0.02) | 0.839 |
| **Off Task** | **0.11** | **(0.16)** | **0.18** | **(0.25)** | **0.15** | **(0.23)** | **-0.05** | **(0.03)** | **0.115** | **0.03** | **(0.03)** | **0.355** |
| Other | 0.00 | (0.03) | 0.01 | (0.08) | 0.01 | (0.04) | 0.00 | (0.01) | 0.690 | 0.00 | (0.01) | 0.722 |
| Social interactions | 0.01 | (0.03) | 0.01 | (0.04) | 0.01 | (0.03) | 0.00 | (0.00) | 0.417 | 0.00 | (0.00) | 0.842 |
| Teacher present but uninvolved | 0.05 | (0.09) | 0.05 | (0.10) | 0.05 | (0.10) | 0.00 | (0.01) | 0.908 | 0.00 | (0.01) | 0.793 |
| Teacher out of room | 0.05 | (0.13) | 0.12 | (0.24) | 0.09 | (0.20) | -0.05 | (0.03) | 0.087 * | 0.02 | (0.03) | 0.461 |
| **Total Time** | **1.00** | | **1.00** | | **1.00** | | | | | | | |
| Satellite technology functional | 0.85 | (0.31) | | | | | | | | | | |
| Satellite feed being viewed | 0.52 | (0.31) | | | | | | | | | | |
| **Observer Assessment: Likert scale 1-4 or 5 -> 4 or 5 greatest** | | | | | | | | | | | | |
| Students engaged in the class (4 pt scale) | 3.23 | (0.56) | 3.27 | (0.55) | 3.33 | (0.55) | -0.08 | (0.08) | 0.364 | -0.06 | (0.07) | 0.355 |
| Facilitator / teacher is organized (4 pt scale) | 3.16 | (0.56) | 3.26 | (0.57) | 3.34 | (0.64) | -0.13 | (0.10) | 0.210 | -0.08 | (0.08) | 0.329 |
| Facilitator / teacher is engaged (4 pt scale) | 3.25 | (0.60) | 3.37 | (0.53) | 3.42 | (0.56) | -0.15 | (0.09) | 0.089 * | -0.06 | (0.07) | 0.385 |
| Facilitator / teacher is knowledgeable (5 pt scale) | 4.25 | (0.51) | 4.39 | (0.61) | 4.40 | (0.49) | -0.14 | (0.08) | 0.083 * | -0.01 | (0.07) | 0.935 |
| Questions adequately answered by facilitator / teacher (5 pt scale) | 4.28 | (0.46) | 4.21 | (0.41) | 4.16 | (0.45) | 0.15 | (0.13) | 0.267 | 0.06 | (0.08) | 0.459 |
| Questions adequately answered by studio teacher (5 pt scale) | 4.08 | (0.76) | . | . | . . | | . . | | | | | |
| **Number of Observations** | **93** | | **198** | | **202** | | **493** | | | **493** | | |

*Notes:* Table reflects the proportion of time spent on classroom activities as measured by Stallings classroom observations. Columns (1) and (2) reflect the activities of the studio teacher and facilitator separately. We record studio teacher activities only when video feed is being viewed. Column (3) combines the studio teacher and facilitator to record the activities of the primary instructor, defined as the studio teacher when the video is being viewed and the facilitator otherwise. The combination of the studio teacher and facilitator thus reflects the main activities of satellite classes and can be compared to treatment and control classrooms. Differences shown control for districted fixed effects and are thus not exactly equal to the difference between the groups compared. The observer assessment questions reflect observer agreement with the statements listed. The first three statements - (1) students engaged in the class; (2) facilitator / teacher is organized; and (3) facilitator / teacher is knowledgable - were measured on a 4-point Likert sale, and the remaining statements on a 5-point Likert scale. On the Likert scales, 1 represents the lowest level of agreement and 4 or 5 the highest level of agreement

**Table 9. Facilitator activities when video feed active vs not active**

| | Studio Teacher (video active) | | Facilitator (video active) | | Facilitator (video not active) | |
|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd |
| **Active Instruction** | **0.39** | **(0.27)** | **0.04** | **(0.10)** | **0.15** | **(0.22)** |
| Reading aloud | 0.02 | (0.06) | 0.00 | (0.02) | 0.02 | (0.08) |
| Instruction / demonstration / lecture | 0.28 | (0.25) | 0.02 | (0.05) | 0.06 | (0.13) |
| Discussion | 0.05 | (0.09) | 0.02 | (0.06) | 0.05 | (0.09) |
| Practice / drill | 0.04 | (0.09) | 0.00 | (0.03) | 0.01 | (0.04) |
| Kinesthetic / projects | 0.00 | (0.00) | 0.00 | (0.00) | 0.00 | (0.00) |
| **Passive Instruction** | **0.05** | **(0.12)** | **0.01** | **(0.02)** | **0.02** | **(0.08)** |
| Working on class assignments | 0.04 | (0.11) | 0.01 | (0.02) | 0.02 | (0.07) |
| Copying | 0.01 | (0.05) | 0.00 | (0.00) | 0.00 | (0.03) |
| **Organization and Management** | **0.07** | **(0.13)** | **0.28** | **(0.23)** | **0.21** | **(0.19)** |
| Verbal instruction | 0.07 | (0.12) | 0.04 | (0.10) | 0.03 | (0.06) |
| Discipline | . | . | 0.00 | (0.02) | 0.00 | (0.02) |
| Classroom management | 0.01 | (0.03) | 0.07 | (0.12) | 0.07 | (0.11) |
| Teacher management | . | . | 0.17 | (0.18) | 0.10 | (0.14) |
| **Off Task** | **0.01** | **(0.03)** | **0.20** | **(0.20)** | **0.10** | **(0.16)** |
| Other | 0.00 | (0.03) | 0.00 | (0.00) | 0.00 | (0.00) |
| Social interactions | 0.00 | (0.02) | 0.01 | (0.03) | 0.01 | (0.03) |
| Teacher present but uninvolved | . | . | 0.16 | (0.17) | 0.05 | (0.09) |
| Teacher out of room | . | . | 0.03 | (0.07) | 0.05 | (0.13) |
| **Total Time** | **0.52** | | **0.52** | | **0.48** | |
| Satellite feed being viewed | yes | | yes | | no | |
| **Number of Observations** | **93** | | **93** | | **93** | |

*Notes:* Table reflects the proportion of time spent on classroom activities as measured by Stallings classroom observations as described in Table 8 notes. Facilitator time is shown when satellite video feed is actively being viewed and when video feed is not being viewed.

**Table A1. Attrition after Year 1 and Year 2**

| | All students | | | Treatment | | | Control | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | n | mean | sd | n | mean | sd | n | coef | se | p-value |
| Attrition after Year 1 | 0.13 | (0.34) | 5485 | 0.14 | (0.35) | 2607 | 0.12 | (0.33) | 2878 | 0.02 | (0.01) | 0.102 |
| Attrition after Year 2 | 0.16 | (0.37) | 5485 | 0.17 | (0.38) | 2607 | 0.15 | (0.36) | 2878 | 0.02 | (0.01) | 0.191 |

*Notes:* This table shows mean attrition after Year 1 and Year 2 of the study. The treatment-control difference reported is the coefficient from a regression of the dependent variable on an indicator variable for treatment and randomization strata (i.e. district) fixed effects. Thus, the difference shown is not exactly equal to the difference between treatment and control means shown. Results are robust to omitting the strata fixed effects. Standard errors are clustered at the school level. ***p<0.01, ** p<0.05, * p<0.1

**Table A2. Differential attrition by baseline EGRA and EGMA score**

| | Attrit at Year 1 | | | Attrit at Year 2 | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.02 | 0.02 | 0.02 | 0.02 | -0.00 | 0.01 |
| | (0.011) | (0.027) | (0.012) | (0.013) | (0.027) | (0.013) |
| EGMA total | | -0.00 | | | -0.00 | |
| | | (0.000) | | | (0.000) | |
| Treatment * EGMA total | | -0.00 | | | 0.00 | |
| | | (0.001) | | | (0.001) | |
| EGRA words-per-minute | | | -0.00 | | | -0.00 |
| | | | (0.000) | | | (0.000) |
| Treatment * EGMA words-per-minute | | | 0.00 | | | 0.00 |
| | | | (0.000) | | | (0.000) |
| Constant | 0.14*** | 0.17*** | 0.14*** | 0.12*** | 0.14*** | 0.12*** |
| | (0.018) | (0.028) | (0.019) | (0.014) | (0.027) | (0.016) |
| Observations | 5,485 | 5,485 | 5,485 | 5,485 | 5,485 | 5,485 |
| R-squared | 0.006 | 0.007 | 0.006 | 0.005 | 0.005 | 0.005 |

*Notes:* Each column represents a separate regression, interacting treatment with baseline EGRA / EGMA score to test whether attrition is differential by baseline performance. Standard errors are clustered at the school level. ***p<0.01, ** p<0.05, * p<0.1

**Table A3a. Baseline covariate balance at baseline**

| | All students | | | Treatment | | | Control | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | n | mean | sd | n | mean | sd | n | coef | se | p-value |
| Student Characteristics | | | | | | | | | | | | |
| Identifies as female | 0.61 | (0.49) | 5485 | 0.61 | (0.49) | 2607 | 0.61 | (0.49) | 2878 | 0.01 | (0.01) | 0.550 |
| Age at baseline | 10.94 | (2.10) | 5379 | 10.98 | (2.10) | 2563 | 10.91 | (2.10) | 2816 | 0.11 | (0.12) | 0.346 |
| In grade 2 at baseline | 0.33 | (0.47) | 5485 | 0.32 | (0.47) | 2607 | 0.34 | (0.47) | 2878 | -0.02 | (0.01) | 0.139 |
| In grade 3 at baseline | 0.34 | (0.47) | 5485 | 0.34 | (0.48) | 2607 | 0.33 | (0.47) | 2878 | 0.02 | (0.01) | 0.218 |
| In grade 4 at baseline | 0.34 | (0.47) | 5485 | 0.34 | (0.47) | 2607 | 0.33 | (0.47) | 2878 | 0.01 | (0.01) | 0.596 |
| Number of school days absent in last week | 0.80 | (1.24) | 5242 | 0.79 | (1.26) | 2504 | 0.80 | (1.22) | 2738 | -0.01 | (0.05) | 0.876 |
| Travels 60 minutes or more to school | 0.14 | (0.34) | 5092 | 0.12 | (0.32) | 2429 | 0.15 | (0.36) | 2663 | -0.04 | (0.02) | 0.024 ** |
| Household Characteristics | | | | | | | | | | | | |
| Total permanent household members | 7.84 | (3.98) | 5098 | 7.93 | (4.09) | 2414 | 7.76 | (3.89) | 2684 | 0.18 | (0.18) | 0.305 |
| Single mother (father not present in) household | 0.16 | (0.37) | 5407 | 0.16 | (0.37) | 2561 | 0.16 | (0.37) | 2846 | 0.00 | (0.01) | 0.999 |
| All children 5-12 years in household attend school | 0.89 | (0.32) | 5417 | 0.88 | (0.32) | 2565 | 0.89 | (0.32) | 2852 | 0.00 | (0.01) | 0.789 |
| Household has permanent roof (e.g. iron, cement, tile) | 0.90 | (0.30) | 5375 | 0.90 | (0.30) | 2545 | 0.90 | (0.30) | 2830 | -0.01 | (0.02) | 0.771 |
| Household has wired electricity | 0.62 | (0.49) | 5395 | 0.63 | (0.48) | 2557 | 0.61 | (0.49) | 2838 | 0.03 | (0.04) | 0.497 |
| Household has indoor plumbing | 0.52 | (0.50) | 5393 | 0.53 | (0.50) | 2558 | 0.52 | (0.50) | 2835 | -0.01 | (0.03) | 0.716 |
| Household has working stove | 0.17 | (0.38) | 5302 | 0.18 | (0.39) | 2505 | 0.16 | (0.37) | 2797 | 0.02 | (0.02) | 0.424 |
| Household has working iron | 0.60 | (0.49) | 5389 | 0.62 | (0.49) | 2553 | 0.59 | (0.49) | 2836 | 0.03 | (0.02) | 0.208 |
| Household has working radio | 0.26 | (0.44) | 5391 | 0.26 | (0.44) | 2552 | 0.25 | (0.43) | 2839 | 0.01 | (0.02) | 0.743 |
| Progress-out-of-Poverty Index (PPI) score | 34.91 | (13.48) | 4970 | 35.15 | (13.68) | 2350 | 34.69 | (13.29) | 2620 | 0.23 | (0.99) | 0.817 |
| Probability household living on $1.25 USD a day | 33.68 | (27.23) | 4970 | 33.43 | (27.26) | 2350 | 33.90 | (27.20) | 2620 | -0.08 | (2.00) | 0.969 |
| Parent Schooling | | | | | | | | | | | | |
| Mother attended some school | 0.56 | (0.50) | 5399 | 0.56 | (0.50) | 2557 | 0.55 | (0.50) | 2842 | 0.01 | (0.03) | 0.771 |
| Highest primary grade attended by mother | 1.62 | (1.96) | 4226 | 1.66 | (1.97) | 1999 | 1.59 | (1.95) | 2227 | 0.05 | (0.12) | 0.675 |
| Father attended some school | 0.63 | (0.48) | 5387 | 0.63 | (0.48) | 2550 | 0.63 | (0.48) | 2837 | -0.01 | (0.02) | 0.846 |
| Highest primary grade attended by father | 1.97 | (2.13) | 3908 | 1.95 | (2.13) | 1842 | 1.99 | (2.13) | 2066 | -0.02 | (0.12) | 0.858 |
| Mother Occupation | | | | | | | | | | | | |
| Mother's occupation not reported | 0.46 | (0.50) | 5485 | 0.46 | (0.50) | 2607 | 0.46 | (0.50) | 2878 | -0.01 | (0.03) | 0.781 |
| Professional / Technical / Managerial / Clerical | 0.03 | (0.18) | 2966 | 0.03 | (0.16) | 1420 | 0.04 | (0.19) | 1546 | -0.01 | (0.01) | 0.234 |
| Government / Public Sector | 0.02 | (0.15) | 2966 | 0.03 | (0.16) | 1420 | 0.02 | (0.15) | 1546 | 0.00 | (0.01) | 0.774 |
| Agriculture / Livestock / Processing / Fishing | 0.24 | (0.42) | 2966 | 0.24 | (0.43) | 1420 | 0.23 | (0.42) | 1546 | 0.01 | (0.03) | 0.841 |
| Production / Building / Mining / Waged Labor | 0.02 | (0.13) | 2966 | 0.02 | (0.14) | 1420 | 0.02 | (0.12) | 1546 | 0.00 | (0.01) | 0.479 |
| Sales / Food Provision | 0.48 | (0.50) | 2966 | 0.48 | (0.50) | 1420 | 0.48 | (0.50) | 1546 | 0.02 | (0.03) | 0.538 |
| Services | 0.08 | (0.27) | 2966 | 0.07 | (0.26) | 1420 | 0.08 | (0.27) | 1546 | -0.01 | (0.01) | 0.311 |
| Unemployed | 0.07 | (0.26) | 2966 | 0.07 | (0.26) | 1420 | 0.07 | (0.26) | 1546 | 0.00 | (0.01) | 0.782 |
| Other | 0.06 | (0.24) | 2966 | 0.06 | (0.24) | 1420 | 0.06 | (0.24) | 1546 | 0.00 | (0.01) | 0.746 |

**Table A3a. Baseline covariate balance at baseline - continued**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Father Occupation | | | | | | | | | | | | |
| Father's occupation not reported | 0.41 | (0.49) | 5485 | 0.42 | (0.49) | 2607 | 0.40 | (0.49) | 2878 | 0.01 | (0.02) | 0.667 |
| Professional / Technical / Managerial / Clerical | 0.06 | (0.23) | 3242 | 0.06 | (0.24) | 1523 | 0.06 | (0.23) | 1719 | 0.00 | (0.01) | 0.744 |
| Government / Public Sector | 0.04 | (0.20) | 3242 | 0.04 | (0.21) | 1523 | 0.04 | (0.20) | 1719 | 0.00 | (0.01) | 0.724 |
| Agriculture / Livestock / Processing / Fishing | 0.48 | (0.50) | 3242 | 0.49 | (0.50) | 1523 | 0.48 | (0.50) | 1719 | 0.02 | (0.03) | 0.575 |
| Production / Building / Mining / Waged Labor | 0.12 | (0.32) | 3242 | 0.12 | (0.32) | 1523 | 0.12 | (0.32) | 1719 | 0.00 | (0.01) | 0.800 |
| Sales / Food Provision | 0.05 | (0.21) | 3242 | 0.05 | (0.21) | 1523 | 0.05 | (0.21) | 1719 | 0.00 | (0.01) | 0.756 |
| Services | 0.12 | (0.32) | 3242 | 0.11 | (0.31) | 1523 | 0.12 | (0.33) | 1719 | -0.01 | (0.01) | 0.501 |
| Unemployed | 0.04 | (0.18) | 3242 | 0.04 | (0.19) | 1523 | 0.03 | (0.18) | 1719 | 0.01 | (0.01) | 0.308 |
| Other | 0.10 | (0.30) | 3242 | 0.09 | (0.29) | 1523 | 0.11 | (0.31) | 1719 | -0.02 | (0.01) | 0.294 |
| Language Spoken at Home | | | | | | | | | | | | |
| Twi | 0.10 | (0.30) | 5475 | 0.10 | (0.29) | 2601 | 0.10 | (0.30) | 2874 | -0.01 | (0.02) | 0.693 |
| Ewe | 0.19 | (0.40) | 5475 | 0.21 | (0.40) | 2601 | 0.18 | (0.39) | 2874 | 0.02 | (0.03) | 0.537 |
| Ga / Dangme | 0.46 | (0.50) | 5475 | 0.45 | (0.50) | 2601 | 0.47 | (0.50) | 2874 | -0.01 | (0.02) | 0.446 |
| English | 0.02 | (0.15) | 5475 | 0.02 | (0.15) | 2601 | 0.02 | (0.15) | 2874 | 0.00 | (0.01) | 0.861 |
| Other | 0.31 | (0.46) | 5475 | 0.32 | (0.47) | 2601 | 0.30 | (0.46) | 2874 | 0.01 | (0.03) | 0.850 |
| Religion | | | | | | | | | | | | |
| Catholic | 0.10 | (0.30) | 5402 | 0.10 | (0.30) | 2559 | 0.10 | (0.31) | 2843 | -0.01 | (0.02) | 0.784 |
| Presbyterian | 0.10 | (0.30) | 5402 | 0.09 | (0.29) | 2559 | 0.11 | (0.31) | 2843 | -0.02 | (0.02) | 0.322 |
| Pentacostal | 0.28 | (0.45) | 5402 | 0.26 | (0.44) | 2559 | 0.29 | (0.45) | 2843 | -0.04 | (0.02) | 0.070 * |
| Charismatic | 0.23 | (0.42) | 5402 | 0.24 | (0.43) | 2559 | 0.22 | (0.41) | 2843 | 0.03 | (0.02) | 0.054 * |
| Other Christian | 0.09 | (0.28) | 5402 | 0.09 | (0.28) | 2559 | 0.09 | (0.28) | 2843 | 0.00 | (0.01) | 0.893 |
| Muslim | 0.11 | (0.31) | 5402 | 0.11 | (0.32) | 2559 | 0.11 | (0.31) | 2843 | 0.01 | (0.03) | 0.684 |
| Other | 0.10 | (0.30) | 5402 | 0.11 | (0.31) | 2559 | 0.09 | (0.29) | 2843 | 0.02 | (0.02) | 0.321 |
| EGRA Subtasks | | | | | | | | | | | | |
| Words per minute | 11.53 | (21.21) | 5485 | 11.65 | (22.23) | 2607 | 11.41 | (20.26) | 2878 | 0.29 | (1.03) | 0.779 |
| Letters per minute | 7.94 | (13.78) | 5485 | 8.04 | (14.36) | 2607 | 7.85 | (13.22) | 2878 | 0.16 | (0.72) | 0.822 |
| Invented words per minute | 2.38 | (7.00) | 5485 | 2.55 | (7.57) | 2607 | 2.22 | (6.44) | 2878 | 0.36 | (0.28) | 0.201 |
| Oral vocabulary score (out of 8) | 3.72 | (1.29) | 5485 | 3.71 | (1.26) | 2607 | 3.73 | (1.32) | 2878 | -0.01 | (0.05) | 0.887 |
| Listening comprehension score (out of 3) | 0.51 | (0.86) | 5485 | 0.53 | (0.88) | 2607 | 0.48 | (0.84) | 2878 | 0.03 | (0.05) | 0.462 |
| Reading comprehension score (out of 2) | 0.12 | (0.40) | 5485 | 0.12 | (0.41) | 2607 | 0.12 | (0.39) | 2878 | 0.01 | (0.02) | 0.731 |
| EGMA subtasks | | | | | | | | | | | | |
| Number identification score (out of 20) | 14.30 | (4.37) | 5485 | 14.30 | (4.34) | 2607 | 14.29 | (4.40) | 2878 | 0.06 | (0.22) | 0.780 |
| Quantity discrimination score (out of 10) | 6.49 | (2.71) | 5485 | 6.49 | (2.65) | 2607 | 6.48 | (2.76) | 2878 | 0.05 | (0.13) | 0.690 |
| Missing number score (out of 10) | 2.95 | (1.79) | 5485 | 2.94 | (1.76) | 2607 | 2.97 | (1.82) | 2878 | -0.02 | (0.08) | 0.773 |
| Addition / subtraction score (out of 56) | 19.72 | (13.00) | 5485 | 19.58 | (12.89) | 2607 | 19.85 | (13.10) | 2878 | -0.05 | (0.63) | 0.941 |
| EGMA total (out of 100) | 46.02 | (20.13) | 5485 | 45.84 | (19.77) | 2607 | 46.18 | (20.45) | 2878 | 0.00 | (1.01) | 0.996 |
| Joint test (p-value) - All variables | | | | | | | | | | | | 0.150 |
| Joint test (p-value) - Student & household characteristics | | | | | | | | | | | | 0.351 |
| Joint test (p-value) - EGRA & EGMA subtasks | | | | | | | | | | | | 0.507 |

*Notes:* Table shows the means and standard deviations of student baseline characteristics. For binary characteristics, the proportion of students with the characteristic is shown. The treatment-control difference is the coefficient from a regression of the dependent variable on an indicator variable for treatment and randomization strata (i.e. district) fixed effects. Thus, the difference shown is not exactly equal to the difference between the treatment and control means shown. Results are robust to omitting the strata fixed effects. Standard errors are clustered at the school level. The EGMA total is calculated such that numer identification, quantity discrimination, and missing number subtasks are each weighted as 20 percent of the total and addition / subtract as 40 percent of the total. Joint test excludes parent occupation and highest grade variables due to high number of missing observations (N=4545). ***p<0.01, ** p<0.05, * p<0.1

**Table A3b. Baseline covariate balance at midline (after Year 1)**

| | All students | | | Treatment | | | Control | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | n | mean | sd | n | mean | sd | n | coef | se | p-value |
| Student Characteristics | | | | | | | | | | | | |
| Identifies as female | 0.60 | (0.49) | 4761 | 0.60 | (0.49) | 2239 | 0.60 | (0.49) | 2522 | 0.01 | (0.01) | 0.607 |
| Age at baseline | 10.88 | (2.08) | 4677 | 10.91 | (2.08) | 2208 | 10.84 | (2.09) | 2469 | 0.09 | (0.12) | 0.462 |
| In grade 2 at baseline | 0.33 | (0.47) | 4761 | 0.32 | (0.47) | 2239 | 0.34 | (0.47) | 2522 | -0.02 | (0.02) | 0.259 |
| In grade 3 at baseline | 0.34 | (0.47) | 4761 | 0.34 | (0.48) | 2239 | 0.33 | (0.47) | 2522 | 0.01 | (0.01) | 0.324 |
| In grade 4 at baseline | 0.34 | (0.47) | 4761 | 0.34 | (0.47) | 2239 | 0.33 | (0.47) | 2522 | 0.00 | (0.01) | 0.759 |
| Number of school days absent in last week | 0.79 | (1.22) | 4552 | 0.77 | (1.24) | 2153 | 0.80 | (1.21) | 2399 | -0.03 | (0.06) | 0.605 |
| Travels 60 minutes or more to school | 0.13 | (0.34) | 4438 | 0.11 | (0.32) | 2095 | 0.15 | (0.36) | 2343 | -0.04 | (0.02) | 0.026 ** |
| Household Characteristics | | | | | | | | | | | | |
| Total permanent household members | 7.87 | (3.99) | 4424 | 7.96 | (4.09) | 2068 | 7.80 | (3.90) | 2356 | 0.17 | (0.18) | 0.350 |
| Single mother (father not present in) household | 0.16 | (0.37) | 4698 | 0.16 | (0.37) | 2201 | 0.16 | (0.37) | 2497 | 0.00 | (0.01) | 0.799 |
| All children 5-12 years in household attend school | 0.88 | (0.32) | 4707 | 0.88 | (0.33) | 2204 | 0.89 | (0.32) | 2503 | -0.01 | (0.02) | 0.693 |
| Household has permanent roof (e.g. iron, cement, tile) | 0.90 | (0.30) | 4668 | 0.90 | (0.30) | 2185 | 0.90 | (0.30) | 2483 | -0.01 | (0.02) | 0.787 |
| Household has wired electricity | 0.62 | (0.49) | 4686 | 0.63 | (0.48) | 2197 | 0.60 | (0.49) | 2489 | 0.03 | (0.04) | 0.521 |
| Household has indoor plumbing | 0.52 | (0.50) | 4687 | 0.52 | (0.50) | 2198 | 0.52 | (0.50) | 2489 | -0.01 | (0.03) | 0.828 |
| Household has working stove | 0.16 | (0.37) | 4604 | 0.17 | (0.37) | 2150 | 0.16 | (0.37) | 2454 | 0.00 | (0.02) | 0.877 |
| Household has working iron | 0.60 | (0.49) | 4683 | 0.62 | (0.49) | 2194 | 0.59 | (0.49) | 2489 | 0.02 | (0.02) | 0.326 |
| Household has working radio | 0.26 | (0.44) | 4684 | 0.26 | (0.44) | 2193 | 0.25 | (0.43) | 2491 | 0.01 | (0.02) | 0.670 |
| Progress-out-of-Poverty Index (PPI) score | 34.72 | (13.35) | 4307 | 34.74 | (13.41) | 2007 | 34.70 | (13.31) | 2300 | -0.17 | (1.02) | 0.868 |
| Probability household living on $1.25 USD a day | 33.94 | (27.24) | 4307 | 34.05 | (27.20) | 2007 | 33.86 | (27.29) | 2300 | 0.54 | (2.09) | 0.799 |
| Parent Schooling | | | | | | | | | | | | |
| Mother attended some school | 0.56 | (0.50) | 4691 | 0.56 | (0.50) | 2198 | 0.55 | (0.50) | 2493 | 0.01 | (0.03) | 0.766 |
| Highest primary grade attended by mother | 1.62 | (1.96) | 3677 | 1.65 | (1.97) | 1720 | 1.60 | (1.95) | 1957 | 0.04 | (0.13) | 0.730 |
| Father attended some school | 0.63 | (0.48) | 4681 | 0.62 | (0.49) | 2192 | 0.64 | (0.48) | 2489 | -0.02 | (0.03) | 0.547 |
| Highest primary grade attended by father | 1.96 | (2.12) | 3389 | 1.90 | (2.12) | 1577 | 2.00 | (2.12) | 1812 | -0.09 | (0.12) | 0.470 |
| Mother Occupation | | | | | | | | | | | | |
| Mother's occupation not reported | 0.46 | (0.50) | 4761 | 0.46 | (0.50) | 2239 | 0.46 | (0.50) | 2522 | -0.01 | (0.03) | 0.764 |
| Professional / Technical / Managerial / Clerical | 0.03 | (0.18) | 2574 | 0.03 | (0.16) | 1218 | 0.04 | (0.19) | 1356 | -0.01 | (0.01) | 0.291 |
| Government / Public Sector | 0.02 | (0.15) | 2574 | 0.03 | (0.16) | 1218 | 0.02 | (0.15) | 1356 | 0.00 | (0.01) | 0.814 |
| Agriculture / Livestock / Processing / Fishing | 0.25 | (0.43) | 2574 | 0.25 | (0.43) | 1218 | 0.24 | (0.43) | 1356 | 0.01 | (0.03) | 0.780 |
| Production / Building / Mining / Waged Labor | 0.02 | (0.13) | 2574 | 0.02 | (0.15) | 1218 | 0.02 | (0.12) | 1356 | 0.00 | (0.01) | 0.488 |
| Sales / Food Provision | 0.47 | (0.50) | 2574 | 0.48 | (0.50) | 1218 | 0.47 | (0.50) | 1356 | 0.02 | (0.03) | 0.450 |
| Services | 0.08 | (0.27) | 2574 | 0.07 | (0.25) | 1218 | 0.08 | (0.28) | 1356 | -0.02 | (0.01) | 0.141 |
| Unemployed | 0.07 | (0.26) | 2574 | 0.07 | (0.26) | 1218 | 0.07 | (0.26) | 1356 | 0.00 | (0.01) | 0.810 |
| Other | 0.06 | (0.24) | 2574 | 0.06 | (0.24) | 1218 | 0.06 | (0.24) | 1356 | -0.01 | (0.01) | 0.581 |

**Table A3b. Baseline covariate balance at midline (after Year 1) - continued**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Father Occupation | | | | | | | | | | | | |
| Father's occupation not reported | 0.41 | (0.49) | 4761 | 0.42 | (0.49) | 2239 | 0.40 | (0.49) | 2522 | 0.02 | (0.03) | 0.395 |
| Professional / Technical / Managerial / Clerical | 0.06 | (0.23) | 2812 | 0.06 | (0.24) | 1293 | 0.05 | (0.22) | 1519 | 0.01 | (0.01) | 0.333 |
| Government / Public Sector | 0.04 | (0.20) | 2812 | 0.04 | (0.20) | 1293 | 0.04 | (0.20) | 1519 | 0.00 | (0.01) | 0.946 |
| Agriculture / Livestock / Processing / Fishing | 0.49 | (0.50) | 2812 | 0.50 | (0.50) | 1293 | 0.48 | (0.50) | 1519 | 0.02 | (0.03) | 0.530 |
| Production / Building / Mining / Waged Labor | 0.11 | (0.32) | 2812 | 0.11 | (0.32) | 1293 | 0.11 | (0.32) | 1519 | 0.00 | (0.01) | 0.807 |
| Sales / Food Provision | 0.05 | (0.21) | 2812 | 0.05 | (0.21) | 1293 | 0.05 | (0.21) | 1519 | 0.00 | (0.01) | 0.729 |
| Services | 0.12 | (0.32) | 2812 | 0.11 | (0.31) | 1293 | 0.12 | (0.33) | 1519 | -0.01 | (0.01) | 0.353 |
| Unemployed | 0.03 | (0.18) | 2812 | 0.04 | (0.19) | 1293 | 0.03 | (0.17) | 1519 | 0.01 | (0.01) | 0.264 |
| Other | 0.10 | (0.30) | 2812 | 0.09 | (0.29) | 1293 | 0.11 | (0.31) | 1519 | -0.02 | (0.01) | 0.271 |
| Language Spoken at Home | | | | | | | | | | | | |
| Twi | 0.09 | (0.29) | 4752 | 0.09 | (0.29) | 2233 | 0.10 | (0.30) | 2519 | -0.01 | (0.02) | 0.568 |
| Ewe | 0.18 | (0.39) | 4752 | 0.19 | (0.39) | 2233 | 0.18 | (0.38) | 2519 | 0.01 | (0.03) | 0.651 |
| Ga / Dangme | 0.46 | (0.50) | 4752 | 0.46 | (0.50) | 2233 | 0.46 | (0.50) | 2519 | -0.01 | (0.01) | 0.607 |
| English | 0.02 | (0.15) | 4752 | 0.02 | (0.14) | 2233 | 0.03 | (0.16) | 2519 | 0.00 | (0.01) | 0.480 |
| Other | 0.32 | (0.47) | 4752 | 0.33 | (0.47) | 2233 | 0.31 | (0.46) | 2519 | 0.01 | (0.03) | 0.756 |
| Religion | | | | | | | | | | | | |
| Catholic | 0.10 | (0.31) | 4694 | 0.11 | (0.31) | 2200 | 0.10 | (0.31) | 2494 | 0.00 | (0.02) | 0.961 |
| Presbyterian | 0.10 | (0.30) | 4694 | 0.09 | (0.28) | 2200 | 0.11 | (0.31) | 2494 | -0.02 | (0.02) | 0.315 |
| Pentacostal | 0.27 | (0.45) | 4694 | 0.26 | (0.44) | 2200 | 0.29 | (0.45) | 2494 | -0.04 | (0.02) | 0.077 * |
| Charismatic | 0.23 | (0.42) | 4694 | 0.24 | (0.43) | 2200 | 0.22 | (0.41) | 2494 | 0.03 | (0.02) | 0.145 |
| Other Christian | 0.08 | (0.28) | 4694 | 0.08 | (0.28) | 2200 | 0.08 | (0.27) | 2494 | 0.00 | (0.01) | 0.981 |
| Muslim | 0.11 | (0.32) | 4694 | 0.12 | (0.32) | 2200 | 0.11 | (0.31) | 2494 | 0.01 | (0.03) | 0.603 |
| Other | 0.10 | (0.30) | 4694 | 0.11 | (0.32) | 2200 | 0.10 | (0.29) | 2494 | 0.02 | (0.02) | 0.302 |
| EGRA Subtasks | | | | | | | | | | | | |
| Words per minute | 11.46 | (20.85) | 4761 | 11.62 | (21.92) | 2239 | 11.32 | (19.86) | 2522 | 0.27 | (1.07) | 0.803 |
| Letters per minute | 7.88 | (13.51) | 4761 | 8.00 | (13.98) | 2239 | 7.78 | (13.07) | 2522 | 0.19 | (0.72) | 0.792 |
| Invented words per minute | 2.38 | (6.98) | 4761 | 2.57 | (7.57) | 2239 | 2.21 | (6.41) | 2522 | 0.37 | (0.30) | 0.219 |
| Oral vocabulary score (out of 8) | 3.74 | (1.29) | 4761 | 3.74 | (1.26) | 2239 | 3.74 | (1.32) | 2522 | 0.01 | (0.05) | 0.910 |
| Listening comprehension score (out of 3) | 0.50 | (0.85) | 4761 | 0.52 | (0.87) | 2239 | 0.48 | (0.84) | 2522 | 0.03 | (0.05) | 0.583 |
| Reading comprehension score (out of 2) | 0.12 | (0.40) | 4761 | 0.12 | (0.40) | 2239 | 0.12 | (0.40) | 2522 | 0.01 | (0.02) | 0.794 |
| EGMA subtasks | | | | | | | | | | | | |
| Number identification score (out of 20) | 14.29 | (4.37) | 4761 | 14.28 | (4.35) | 2239 | 14.29 | (4.38) | 2522 | 0.03 | (0.23) | 0.910 |
| Quantity discrimination score (out of 10) | 6.49 | (2.70) | 4761 | 6.48 | (2.64) | 2239 | 6.49 | (2.75) | 2522 | 0.02 | (0.13) | 0.895 |
| Missing number score (out of 10) | 2.97 | (1.80) | 4761 | 2.96 | (1.78) | 2239 | 2.97 | (1.81) | 2522 | -0.01 | (0.08) | 0.939 |
| Addition / subtraction score (out of 56) | 19.80 | (13.02) | 4761 | 19.70 | (12.94) | 2239 | 19.88 | (13.09) | 2522 | -0.03 | (0.65) | 0.969 |
| EGMA total (out of 100) | 46.10 | (20.11) | 4761 | 45.95 | (19.80) | 2239 | 46.24 | (20.38) | 2522 | -0.07 | (1.05) | 0.945 |
| Joint test (p-value) - All variables | | | | | | | | | | | | 0.365 |
| Joint test (p-value) - Student & household characteristics | | | | | | | | | | | | 0.465 |
| Joint test (p-value) - EGRA & EGMA subtasks | | | | | | | | | | | | 0.654 |

*Notes:* Table shows the means and standard deviations of student baseline characteristics. For binary characteristics, the proportion of students with the characteristic is shown. The treatment-control difference is the coefficient from a regression of the dependent variable on an indicator variable for treatment and randomization strata (i.e. district) fixed effects. Thus, the difference shown is not exactly equal to the difference between the treatment and control means shown. Results are robust to omitting the strata fixed effects. Standard errors are clustered at the school level. The EGMA total is calculated such that numer identification, quantity discrimination, and missing number subtasks are each weighted as 20 percent of the total and addition / subtract as 40 percent of the total. Joint test excludes parent occupation and highest grade variables due to high number of missing observations (N=3964). ***p<0.01, ** p<0.05, * p<0.1

**Table A3c. Baseline covariate balance at endline (after Year 2)**

| | All students | | | Treatment | | | Control | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | n | mean | sd | n | mean | sd | n | coef | se | p-value |
| Student Characteristics | | | | | | | | | | | | |
| Identifies as female | 0.60 | (0.49) | 4606 | 0.60 | (0.49) | 2165 | 0.59 | (0.49) | 2441 | 0.01 | (0.01) | 0.329 |
| Age at baseline | 10.86 | (2.08) | 4520 | 10.90 | (2.08) | 2130 | 10.83 | (2.08) | 2390 | 0.10 | (0.12) | 0.371 |
| In grade 2 at baseline | 0.33 | (0.47) | 4606 | 0.33 | (0.47) | 2165 | 0.34 | (0.47) | 2441 | -0.02 | (0.02) | 0.324 |
| In grade 3 at baseline | 0.33 | (0.47) | 4606 | 0.34 | (0.47) | 2165 | 0.33 | (0.47) | 2441 | 0.01 | (0.01) | 0.383 |
| In grade 4 at baseline | 0.34 | (0.47) | 4606 | 0.34 | (0.47) | 2165 | 0.33 | (0.47) | 2441 | 0.00 | (0.01) | 0.775 |
| Number of school days absent in last week | 0.79 | (1.23) | 4399 | 0.77 | (1.24) | 2082 | 0.81 | (1.23) | 2317 | -0.04 | (0.05) | 0.464 |
| Travels 60 minutes or more to school | 0.13 | (0.34) | 4290 | 0.11 | (0.32) | 2027 | 0.15 | (0.35) | 2263 | -0.04 | (0.02) | 0.014 ** |
| Household Characteristics | | | | | | | | | | | | |
| Total permanent household members | 7.88 | (3.97) | 4274 | 7.99 | (4.09) | 1999 | 7.79 | (3.86) | 2275 | 0.24 | (0.18) | 0.194 |
| Single mother (father not present in) household | 0.16 | (0.37) | 4539 | 0.16 | (0.37) | 2128 | 0.16 | (0.37) | 2411 | 0.00 | (0.01) | 0.987 |
| All children 5-12 years in household attend school | 0.88 | (0.32) | 4548 | 0.88 | (0.33) | 2131 | 0.88 | (0.32) | 2417 | 0.00 | (0.02) | 0.821 |
| Household has permanent roof (e.g. iron, cement, tile) | 0.90 | (0.31) | 4511 | 0.90 | (0.31) | 2113 | 0.90 | (0.31) | 2398 | -0.01 | (0.02) | 0.812 |
| Household has wired electricity | 0.62 | (0.49) | 4528 | 0.63 | (0.48) | 2124 | 0.61 | (0.49) | 2404 | 0.03 | (0.04) | 0.524 |
| Household has indoor plumbing | 0.53 | (0.50) | 4528 | 0.54 | (0.50) | 2125 | 0.52 | (0.50) | 2403 | 0.00 | (0.03) | 0.950 |
| Household has working stove | 0.17 | (0.37) | 4448 | 0.17 | (0.38) | 2078 | 0.16 | (0.37) | 2370 | 0.01 | (0.02) | 0.702 |
| Household has working iron | 0.60 | (0.49) | 4526 | 0.61 | (0.49) | 2122 | 0.59 | (0.49) | 2404 | 0.02 | (0.02) | 0.354 |
| Household has working radio | 0.26 | (0.44) | 4527 | 0.26 | (0.44) | 2121 | 0.25 | (0.44) | 2406 | 0.00 | (0.02) | 0.873 |
| Progress-out-of-Poverty Index (PPI) score | 34.82 | (13.42) | 4163 | 35.01 | (13.51) | 1942 | 34.66 | (13.33) | 2221 | 0.07 | (1.00) | 0.943 |
| Probability household living on $1.25 USD a day | 33.76 | (27.27) | 4163 | 33.62 | (27.20) | 1942 | 33.88 | (27.34) | 2221 | 0.20 | (2.03) | 0.921 |
| Parent Schooling | | | | | | | | | | | | |
| Mother attended some school | 0.56 | (0.50) | 4532 | 0.57 | (0.50) | 2124 | 0.56 | (0.50) | 2408 | 0.01 | (0.03) | 0.705 |
| Highest primary grade attended by mother | 1.65 | (1.96) | 3551 | 1.68 | (1.97) | 1664 | 1.61 | (1.95) | 1887 | 0.05 | (0.12) | 0.663 |
| Father attended some school | 0.64 | (0.48) | 4523 | 0.63 | (0.48) | 2119 | 0.64 | (0.48) | 2404 | -0.01 | (0.03) | 0.821 |
| Highest primary grade attended by father | 1.99 | (2.12) | 3272 | 1.95 | (2.12) | 1516 | 2.02 | (2.12) | 1756 | -0.05 | (0.12) | 0.645 |
| Mother Occupation | | | | | | | | | | | | |
| Mother's occupation not reported | 0.45 | (0.50) | 4606 | 0.45 | (0.50) | 2165 | 0.46 | (0.50) | 2441 | -0.01 | (0.03) | 0.695 |
| Professional / Technical / Managerial / Clerical | 0.03 | (0.17) | 2515 | 0.02 | (0.15) | 1194 | 0.04 | (0.19) | 1321 | -0.01 | (0.01) | 0.076 * |
| Government / Public Sector | 0.02 | (0.15) | 2515 | 0.02 | (0.15) | 1194 | 0.02 | (0.15) | 1321 | 0.00 | (0.01) | 0.853 |
| Agriculture / Livestock / Processing / Fishing | 0.24 | (0.43) | 2515 | 0.25 | (0.43) | 1194 | 0.24 | (0.42) | 1321 | 0.00 | (0.03) | 0.912 |
| Production / Building / Mining / Waged Labor | 0.02 | (0.14) | 2515 | 0.02 | (0.15) | 1194 | 0.02 | (0.13) | 1321 | 0.01 | (0.01) | 0.380 |
| Sales / Food Provision | 0.48 | (0.50) | 2515 | 0.48 | (0.50) | 1194 | 0.47 | (0.50) | 1321 | 0.02 | (0.03) | 0.372 |
| Services | 0.08 | (0.27) | 2515 | 0.07 | (0.26) | 1194 | 0.08 | (0.28) | 1321 | -0.01 | (0.01) | 0.246 |
| Unemployed | 0.07 | (0.26) | 2515 | 0.07 | (0.26) | 1194 | 0.07 | (0.26) | 1321 | 0.00 | (0.01) | 0.712 |
| Other | 0.06 | (0.24) | 2515 | 0.06 | (0.24) | 1194 | 0.06 | (0.24) | 1321 | 0.00 | (0.01) | 0.853 |

**Table A3c. Baseline covariate balance at endline (after Year 2) - continued**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Father Occupation | | | | | | | | | | | | |
| Father's occupation not reported | 0.40 | (0.49) | 4606 | 0.41 | (0.49) | 2165 | 0.40 | (0.49) | 2441 | 0.01 | (0.02) | 0.696 |
| Professional / Technical / Managerial / Clerical | 0.06 | (0.23) | 2745 | 0.06 | (0.23) | 1276 | 0.05 | (0.23) | 1469 | 0.01 | (0.01) | 0.607 |
| Government / Public Sector | 0.04 | (0.20) | 2745 | 0.04 | (0.19) | 1276 | 0.04 | (0.20) | 1469 | 0.00 | (0.01) | 0.609 |
| Agriculture / Livestock / Processing / Fishing | 0.49 | (0.50) | 2745 | 0.49 | (0.50) | 1276 | 0.48 | (0.50) | 1469 | 0.02 | (0.03) | 0.571 |
| Production / Building / Mining / Waged Labor | 0.12 | (0.32) | 2745 | 0.12 | (0.32) | 1276 | 0.12 | (0.32) | 1469 | 0.00 | (0.01) | 0.786 |
| Sales / Food Provision | 0.05 | (0.21) | 2745 | 0.05 | (0.21) | 1276 | 0.05 | (0.21) | 1469 | 0.00 | (0.01) | 0.892 |
| Services | 0.12 | (0.32) | 2745 | 0.11 | (0.32) | 1276 | 0.12 | (0.33) | 1469 | -0.01 | (0.01) | 0.545 |
| Unemployed | 0.04 | (0.18) | 2745 | 0.04 | (0.20) | 1276 | 0.03 | (0.17) | 1469 | 0.01 | (0.01) | 0.160 |
| Other | 0.10 | (0.30) | 2745 | 0.09 | (0.29) | 1276 | 0.11 | (0.31) | 1469 | -0.02 | (0.02) | 0.325 |
| Language Spoken at Home | | | | | | | | | | | | |
| Twi | 0.09 | (0.29) | 4597 | 0.09 | (0.28) | 2159 | 0.09 | (0.29) | 2438 | -0.01 | (0.02) | 0.572 |
| Ewe | 0.18 | (0.39) | 4597 | 0.19 | (0.39) | 2159 | 0.18 | (0.38) | 2438 | 0.01 | (0.03) | 0.679 |
| Ga / Dangme | 0.47 | (0.50) | 4597 | 0.47 | (0.50) | 2159 | 0.48 | (0.50) | 2438 | -0.01 | (0.02) | 0.640 |
| English | 0.02 | (0.15) | 4597 | 0.02 | (0.14) | 2159 | 0.03 | (0.16) | 2438 | 0.00 | (0.01) | 0.625 |
| Other | 0.31 | (0.46) | 4597 | 0.32 | (0.47) | 2159 | 0.30 | (0.46) | 2438 | 0.01 | (0.03) | 0.698 |
| Religion | | | | | | | | | | | | |
| Catholic | 0.10 | (0.30) | 4535 | 0.11 | (0.31) | 2126 | 0.10 | (0.30) | 2409 | 0.00 | (0.02) | 0.889 |
| Presbyterian | 0.10 | (0.30) | 4535 | 0.09 | (0.28) | 2126 | 0.10 | (0.31) | 2409 | -0.02 | (0.02) | 0.354 |
| Pentacostal | 0.27 | (0.45) | 4535 | 0.26 | (0.44) | 2126 | 0.29 | (0.45) | 2409 | -0.04 | (0.02) | 0.093 * |
| Charismatic | 0.23 | (0.42) | 4535 | 0.24 | (0.43) | 2126 | 0.23 | (0.42) | 2409 | 0.02 | (0.02) | 0.265 |
| Other Christian | 0.09 | (0.28) | 4535 | 0.09 | (0.28) | 2126 | 0.09 | (0.28) | 2409 | 0.00 | (0.01) | 0.892 |
| Muslim | 0.11 | (0.31) | 4535 | 0.12 | (0.32) | 2126 | 0.10 | (0.31) | 2409 | 0.02 | (0.03) | 0.565 |
| Other | 0.10 | (0.30) | 4535 | 0.11 | (0.31) | 2126 | 0.10 | (0.29) | 2409 | 0.02 | (0.02) | 0.319 |
| EGRA Subtasks | | | | | | | | | | | | |
| Words per minute | 11.58 | (20.93) | 4606 | 11.50 | (21.59) | 2165 | 11.65 | (20.32) | 2441 | -0.07 | (1.05) | 0.948 |
| Letters per minute | 8.00 | (13.57) | 4606 | 8.04 | (14.02) | 2165 | 7.96 | (13.17) | 2441 | 0.06 | (0.74) | 0.931 |
| Invented words per minute | 2.42 | (6.98) | 4606 | 2.56 | (7.46) | 2165 | 2.30 | (6.52) | 2441 | 0.30 | (0.31) | 0.322 |
| Oral vocabulary score (out of 8) | 3.75 | (1.28) | 4606 | 3.75 | (1.25) | 2165 | 3.76 | (1.31) | 2441 | 0.01 | (0.05) | 0.857 |
| Listening comprehension score (out of 3) | 0.51 | (0.86) | 4606 | 0.53 | (0.88) | 2165 | 0.49 | (0.85) | 2441 | 0.02 | (0.05) | 0.619 |
| Reading comprehension score (out of 2) | 0.12 | (0.40) | 4606 | 0.12 | (0.40) | 2165 | 0.12 | (0.40) | 2441 | 0.00 | (0.02) | 0.863 |
| EGMA subtasks | | | | | | | | | | | | |
| Number identification score (out of 20) | 14.31 | (4.35) | 4606 | 14.31 | (4.30) | 2165 | 14.32 | (4.40) | 2441 | 0.05 | (0.22) | 0.816 |
| Quantity discrimination score (out of 10) | 6.50 | (2.70) | 4606 | 6.49 | (2.63) | 2165 | 6.51 | (2.76) | 2441 | 0.02 | (0.13) | 0.897 |
| Missing number score (out of 10) | 2.97 | (1.79) | 4606 | 2.95 | (1.77) | 2165 | 2.99 | (1.81) | 2441 | -0.03 | (0.08) | 0.698 |
| Addition / subtraction score (out of 56) | 19.86 | (12.99) | 4606 | 19.63 | (12.83) | 2165 | 20.06 | (13.13) | 2441 | -0.19 | (0.64) | 0.769 |
| EGMA total (out of 100) | 46.24 | (20.07) | 4606 | 45.93 | (19.63) | 2165 | 46.52 | (20.46) | 2441 | -0.23 | (1.04) | 0.826 |
| Joint test (p-value) - All variables | | | | | | | | | | | | 0.283 |
| Joint test (p-value) - Student & household characteristics | | | | | | | | | | | | 0.401 |
| Joint test (p-value) - EGRA & EGMA subtasks | | | | | | | | | | | | 0.574 |

*Notes:* Table shows the means and standard deviations of student baseline characteristics. For binary characteristics, the proportion of students with the characteristic is shown. The treatment-control difference is the coefficient from a regression of the dependent variable on an indicator variable for treatment and randomization strata (i.e. district) fixed effects. Thus, the difference shown is not exactly equal to the difference between the treatment and control means shown. Results are robust to omitting the strata fixed effects. Standard errors are clustered at the school level. The EGMA total is calculated such that numer identification, quantity discrimination, and missing number subtasks are each weighted as 20 percent of the total and addition / subtract as 40 percent of the total. Joint test excludes parent occupation and highest grade variables due to high number of missing observations (N=3823). ***p<0.01, ** p<0.05, * p<0.1

**Table A4. Effect of satellite-classes on average EGRA and EGMA: Alternative Specifications**

| | Simple Difference | | Controlling for baseline score | | Alternative Standardization | | Raw Scores | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel A: EGRA Words Per Minute | | | | | | | | |
| Treated | 0.03 | 0.03 | 0.03 | 0.03 | 0.00 | 0.01 | 0.09 | 0.18 |
| | (0.07) | (0.05) | (0.03) | (0.03) | (0.07) | (0.05) | (1.46) | (1.16) |
| Baseline z-score | | | 0.78*** | 0.74*** | | | | |
| | | | (0.02) | (0.02) | | | | |
| Treated * Post | | | | | 0.03 | 0.03 | 1.17 | 1.05 |
| | | | | | (0.03) | (0.03) | (1.31) | (1.31) |
| Observations | 7,822 | 7,822 | 7,822 | 7,822 | 12,849 | 12,849 | 12,849 | 12,849 |
| R-squared | 0.000 | 0.105 | 0.546 | 0.561 | 0.001 | 0.098 | 0.091 | 0.177 |
| Panel B: EGMA Math Total | | | | | | | | |
| Treated | 0.22*** | 0.21*** | 0.23*** | 0.22*** | -0.02 | -0.02 | -0.40 | -0.31 |
| | (0.08) | (0.05) | (0.03) | (0.03) | (0.08) | (0.05) | (1.62) | (1.05) |
| Baseline z-score | | | 0.79*** | 0.76*** | | | | |
| | | | (0.01) | (0.01) | | | | |
| Treated * Post | | | | | 0.24*** | 0.24*** | 4.43*** | 4.36*** |
| | | | | | (0.03) | (0.03) | (0.65) | (0.65) |
| Observations | 7,821 | 7,821 | 7,821 | 7,821 | 12,848 | 12,848 | 12,848 | 12,848 |
| R-squared | 0.012 | 0.137 | 0.591 | 0.602 | 0.008 | 0.137 | 0.124 | 0.238 |
| District fixed effects | No | Yes | No | Yes | No | Yes | No | Yes |

*Notes:* Each column represents a separate regression. Test scores in columns 1 through 4 are standardized based on the contemporaneous control distribution. Test scores in columns 5 and 6 are standardized based on the contemporaneous distribution of all students. Columns 7 and 8 use raw EGRA words-per-minute and total EGMA score. The district is the level at which the satellite classes were randomized. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1

**Table A5. Effect of satellite classes on gains in EGRA and EGMA standardized scores, by year**

| | EGRA Words Per Minute | | | EGRA Words Per Minute | | |
|---|---|---|---|---|---|---|
| | Year 1 (baseline to midline) | Year 2 (midline to endline) | Year 1 & 2 (baseline to endline) | Year 1 (baseline to midline) | Year 2 (midline to endline) | Year 1 & 2 (baseline to endline) |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated | -0.01 | 0.10*** | 0.09** | 0.20*** | 0.09*** | 0.25*** |
| | (0.03) | (0.02) | (0.04) | (0.02) | (0.02) | (0.03) |
| Baseline z-score | 0.71*** | | 0.76*** | 0.79*** | | 0.78*** |
| | (0.02) | | (0.04) | (0.01) | | (0.02) |
| Year 1 (midline) z-score | | 0.92*** | | | 0.84*** | |
| | | (0.02) | | | (0.01) | |
| Grade 2 cohort | -0.17*** | 0.01 | -0.06** | 0.07*** | 0.04* | 0.18*** |
| | (0.03) | (0.02) | (0.03) | (0.03) | (0.02) | (0.03) |
| Grade 3 cohort | -0.06** | | | -0.03 | | |
| | (0.02) | | | (0.02) | | |
| Observations | 4,761 | 2,885 | 3,061 | 4,761 | 2,884 | 3,060 |
| R-squared | 0.652 | 0.731 | 0.404 | 0.651 | 0.703 | 0.540 |
| District fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade 2 cohort | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade 3 cohort | Yes | Yes | Yes | Yes | Yes | Yes |
| Grade 4 cohort | Yes | | | Yes | | |

*Notes:* Each column represents a separate regression. The district is the level at which the satellite classes were randomized. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1

**Table A6. Teacher attributes, beliefs and pedagogical practices in treatment and control schools**

| | Treatment (Facilitator) | | Treatment (Standard Teacher) | | Control (Standard Teacher) | | Difference (Facilitator - Standard Control) | | | Difference (Standard Treatment - Control) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd | coef | se | p-value | coef | se | p-value |
| Panel A. Teacher attributes | | | | | | | | | | | | |
| Holds university degree | 0.27 | (0.45) | 0.29 | (0.46) | 0.30 | (0.46) | -0.03 | (0.05) | 0.607 | -0.01 | (0.05) | 0.918 |
| Holds education degree or certificate | 0.73 | (0.45) | 0.84 | (0.37) | 0.83 | (0.38) | -0.10 | (0.06) | 0.088 * | 0.02 | (0.05) | 0.714 |
| Speaks the Ghanaian language spoken by majority of students | 0.86 | (0.35) | 0.60 | (0.49) | 0.62 | (0.49) | 0.24 | (0.05) | 0.000 *** | -0.01 | (0.06) | 0.824 |
| Travels more than 30 minutes to school | 0.18 | (0.39) | 0.20 | (0.40) | 0.18 | (0.39) | 0.00 | (0.05) | 0.936 | 0.01 | (0.05) | 0.818 |
| Missed at least 1 lesson in the last 5 days | 0.17 | (0.38) | 0.77 | (0.42) | 0.64 | (0.48) | -0.47 | (0.06) | 0.000 *** | 0.12 | (0.05) | 0.023 ** |
| Headmaster visited classroom in last 5 days | 0.44 | (0.50) | 0.46 | (0.50) | 0.58 | (0.50) | -0.14 | (0.07) | 0.048 ** | -0.12 | (0.07) | 0.100 |
| Panel B. Teacher compensation | | | | | | | | | | | | |
| Satisified with compensation | 0.27 | (0.44) | 0.27 | (0.45) | 0.26 | (0.44) | 0.01 | (0.05) | 0.881 | 0.01 | (0.05) | 0.788 |
| Actual monthly take-home pay as a teacher (Ghana Cedi) | 670.94 | (399.23) | 697.25 | (371.29) | 727.15 | (335.99) | -58.85 | (51.70) | 0.257 | -32.55 | (44.64) | 0.467 |
| Desired monthly take-home pay as a teacher (Ghana Cedi) | 1211.80 | (691.03) | 1425.15 | (636.16) | 1532.63 | (802.46) | -323.52 | (101.10) | 0.002 *** | -110.17 | (89.51) | 0.220 |
| Difference (desired - actual) in take-home pay (Ghana Cedi) | 553.23 | (419.57) | 727.91 | (547.85) | 805.48 | (701.11) | -252.30 | (74.07) | 0.001 *** | -77.62 | (76.79) | 0.314 |
| Views teaching as a temporary job | 0.18 | (0.39) | 0.18 | (0.39) | 0.29 | (0.45) | -0.11 | (0.05) | 0.030 ** | -0.11 | (0.05) | 0.032 ** |
| Panel C: Teacher beliefs (Likert scale 1-5, with 5 indicating "strongly agree") | | | | | | | | | | | | |
| I am knowledgeable and comfortable teaching assigned content | 4.32 | (0.61) | 4.31 | (0.71) | 4.39 | (0.70) | -0.07 | (0.08) | 0.356 | -0.08 | (0.08) | 0.304 |
| The school is supportive and encouraging of teachers | 3.85 | (0.96) | 3.83 | (0.95) | 3.77 | (0.96) | 0.08 | (0.12) | 0.489 | 0.07 | (0.11) | 0.547 |
| If I need help with teaching, there is someone I can turn to | 4.16 | (0.73) | 4.10 | (0.85) | 4.15 | (0.78) | 0.01 | (0.09) | 0.941 | -0.05 | (0.10) | 0.576 |
| Necessary materials like textbooks are available as needed | 3.48 | (1.06) | 2.88 | (1.18) | 2.83 | (1.19) | 0.65 | (0.14) | 0.000 *** | 0.05 | (0.15) | 0.748 |
| Teachers are the most important determinent of student success | 3.92 | (1.09) | 3.90 | (1.11) | 4.00 | (0.96) | -0.08 | (0.13) | 0.548 | -0.09 | (0.13) | 0.472 |
| Teachers have important role in improving performance of struggling students | 4.42 | (0.51) | 4.47 | (0.51) | 4.43 | (0.53) | -0.02 | (0.06) | 0.776 | 0.03 | (0.06) | 0.586 |
| The amount a student can learn is primarily related to family background | 3.42 | (1.13) | 3.27 | (1.13) | 3.62 | (1.11) | -0.20 | (0.14) | 0.135 | -0.35 | (0.13) | 0.009 *** |
| Children from disadvantaged backgrounds cannot reach same level as advantaged | 3.02 | (1.30) | 2.88 | (1.40) | 3.08 | (1.36) | -0.05 | (0.16) | 0.767 | -0.19 | (0.15) | 0.194 |
| If finances are limited, families should prioritize boys' education over girls' | 1.68 | (0.72) | 1.62 | (0.88) | 1.65 | (0.76) | 0.04 | (0.09) | 0.624 | -0.02 | (0.09) | 0.840 |
| I am satisifed with the teaching profession | 3.79 | (1.02) | 3.64 | (1.09) | 3.74 | (0.99) | 0.05 | (0.12) | 0.678 | -0.10 | (0.12) | 0.408 |
| Panel D: Pedagogical methods used in classroom | | | | | | | | | | | | |
| Use of local language to reinforce teaching points | 0.46 | (0.50) | 0.41 | (0.49) | 0.35 | (0.48) | 0.11 | (0.08) | 0.166 | 0.05 | (0.07) | 0.482 |
| Classroom grouping and pairing for activities | 0.69 | (0.47) | 0.69 | (0.47) | 0.58 | (0.50) | 0.11 | (0.07) | 0.107 | 0.11 | (0.06) | 0.099 * |
| Class circulation during lessons | 0.34 | (0.48) | 0.34 | (0.47) | 0.32 | (0.47) | 0.02 | (0.07) | 0.824 | 0.01 | (0.07) | 0.840 |
| Pupil motivation using praise | 0.45 | (0.50) | 0.44 | (0.50) | 0.42 | (0.50) | 0.03 | (0.07) | 0.691 | 0.02 | (0.08) | 0.811 |
| Use of distance technology (microphone, chat function) | 0.08 | (0.27) | 0.07 | (0.26) | 0.01 | (0.12) | 0.06 | (0.03) | 0.041 ** | 0.06 | (0.03) | 0.038 ** |
| Activity and role play | 0.72 | (0.45) | 0.69 | (0.47) | 0.72 | (0.45) | -0.01 | (0.06) | 0.912 | -0.04 | (0.06) | 0.550 |
| Other | 0.16 | (0.37) | 0.18 | (0.39) | 0.26 | (0.44) | -0.09 | (0.06) | 0.122 | -0.06 | (0.06) | 0.255 |
| Number of observations | 132 | | 163 | | 212 | | | | | | | |

*Notes*: This table shows means and standard deviations of a teacher / facilitator survey administered in the second half of the second year of the program. Surveys were administered to grades 4-6 Math and English teachers at all 147 schools, as well as with facilitators at all 70 treatment schools. In total, 507 teachers/facilitators were interview - 132 facilitators were interviewed (only 26 of which did not teach other standard classes in their schools), 163 standard teachers (who did not serve as facilitators) were interviewed in treatment schools, and 212 standard teachers in control schools were interviewed. Differences shown are the difference in average responses between facilitators vs. control teachers and non-facilitator teachers in treatment schools vs. control teachers. Standard errors clustered at the school level are shown in parentheses. ***p<0.01, ** p<0.05, * p<0.1